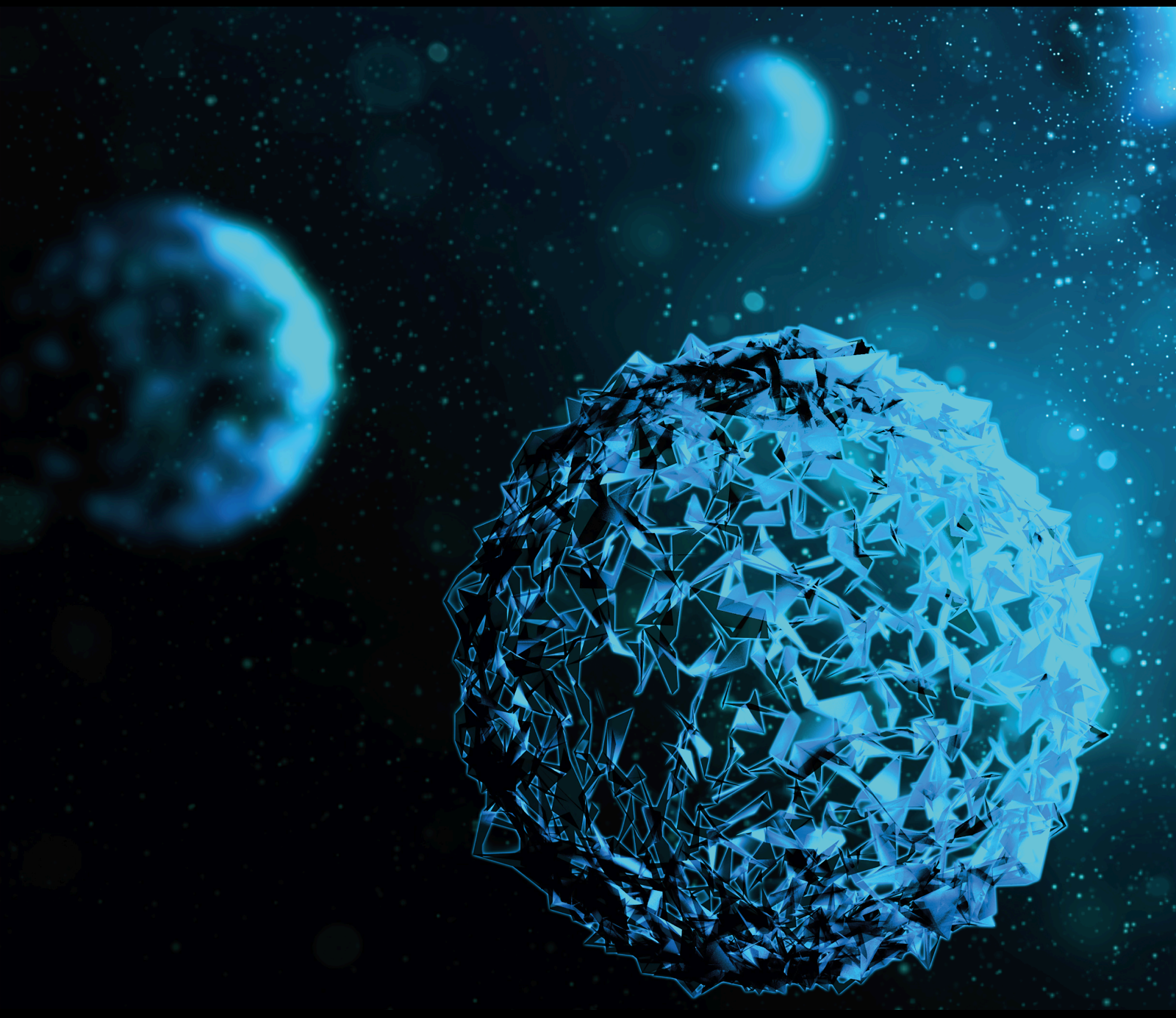
BioMed Research International

Microscopic Image Analysis in Histopathology

Lead Guest Editor: Chen Li
Guest Editors: Jun Xu, Marcin Grzegorzek, and Cheng Lu



Microscopic Image Analysis in Histopathology

BioMed Research International

Microscopic Image Analysis in Histopathology

Lead Guest Editor: Chen Li
Guest Editors: Jun Xu, Marcin Grzegorzek, and Cheng Lu

Copyright © 2024 Hindawi Limited. All rights reserved.

This is a special issue published in “BioMed Research International.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Section Editors

Penny A. Asbell, USA
David Bernardo, Spain
Gerald Brandacher, USA
Kim Bridle, Australia
Laura Chronopoulou, Italy
Gerald A. Colvin, USA
Aaron S. Dumont, USA
Pierfrancesco Franco, Italy
Raj P. Kandpal, USA
Fabrizio Montecucco, Italy
Mangesh S. Pednekar, India
Letterio S. Politi, USA
Jinsong Ren, China
William B. Rodgers, USA
Harry W. Schroeder, USA
Andrea Scribante, Italy
Germán Vicente-Rodriguez, Spain
Momiao Xiong, USA
Hui Zhang, China

Academic Editors

Pathology

Atif A. Ahmed, USA
Shinichi Aishima, Japan
Valeria Barresi, Italy
Ewa Bien, Poland
Tarek A. Bismar, Canada
Rosario Caltabiano, Italy
Jin-Jer Chen, China
Chandu De Silva, Sri Lanka
Vasiliki Galani, Greece
Wan-Ming Hu, China
Genichiro Ishii, Japan
Glen Jickling, USA
Shigeru Kobayashi, Japan
Kun Li, China
Satoshi Maruyama, Japan
Kathleen Montone, USA
Yukihiro Nakanishi, USA
Stefania Pizzimenti, Italy
Jianyu Y. Rao, USA
Takashi Saku, Japan
Claire Troakes, United Kingdom
Koichiro Wada, Japan

Contents

Retracted: Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation
BioMed Research International
Retraction (1 page), Article ID 9878053, Volume 2024 (2024)

Retracted: A Study on Nonvisual Effects of Natural Light Environment in a Maternity Ward of a Hospital in Cold Area
BioMed Research International
Retraction (1 page), Article ID 9867136, Volume 2024 (2024)

Retracted: Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma
BioMed Research International
Retraction (1 page), Article ID 9793629, Volume 2024 (2024)

Super-Resolution Swin Transformer and Attention Network for Medical CT Imaging
Jianhua Hu, Shuzhao Zheng, Bo Wang, Guixiang Luo, WoQing Huang, and Jun Zhang
Research Article (8 pages), Article ID 4431536, Volume 2022 (2022)

TSHVNet: Simultaneous Nuclear Instance Segmentation and Classification in Histopathological Images Based on Multiattention Mechanisms
Yuli Chen, Yuhang Jia, Xinxin Zhang, Jiayang Bai, Xue Li, Miao Ma, Zengguo Sun, and Zhao Pei
Research Article (17 pages), Article ID 7921922, Volume 2022 (2022)

Prediction Model of Residual Neural Network for Pathological Confirmed Lymph Node Metastasis of Ovarian Cancer
Huanchun Yao and Xinglong Zhang
Research Article (10 pages), Article ID 9646846, Volume 2022 (2022)

Segmentation of Breast Tubules in H&E Images Based on a DKS-DoubleU-Net Model
Yuli Chen, Yao Zhou, Guoping Chen, Yuchuan Guo, Yanquan Lv, Miao Ma, Zhao Pei, and Zengguo Sun
Research Article (12 pages), Article ID 2961610, Volume 2022 (2022)

Pulse Signal Analysis Based on Deep Learning Network
Quanyu E
Research Article (11 pages), Article ID 6256126, Volume 2022 (2022)

A Machine Learning Model Based on Unsupervised Clustering Multihabitat to Predict the Pathological Grading of Meningiomas
Xinghao Wang, Jia Li, Jing Sun, Wenjuan Liu, Linkun Cai, Pengfei Zhao, Zhenghan Yang, Han Lv, and Zhenchang Wang
Research Article (7 pages), Article ID 8955227, Volume 2022 (2022)

Efficient Synchronous Real-Time CADe for Multicategory Lesions in Gastroscopy by Using Multiclass Detection Model
Yiji Ku, Hui Ding, and Guangzhi Wang
Research Article (9 pages), Article ID 8504149, Volume 2022 (2022)

Application of Deep Learning on the Prognosis of Cutaneous Melanoma Based on Full Scan Pathology Images
Anhai Li, Xiaoyuan Li, Wenwen Li, Xiaoqian Yu, Mengmeng Qi, and Ding Li
Research Article (6 pages), Article ID 4864485, Volume 2022 (2022)

Computer Image Analysis Reveals C-Myc as a Potential Biomarker for Discriminating between Keratoacanthoma and Cutaneous Squamous Cell Carcinoma
Xinyun Fan, Xueli Niu, Ze Wu, Lu Yao, Shirui Chen, Wenyu Wan, Bo Huang, Rui-Qun Qi, and Tao Zhang
Research Article (17 pages), Article ID 3168503, Volume 2022 (2022)

Histopathological Tissue Segmentation of Lung Cancer with Bilinear CNN and Soft Attention
Rui Xu, Zhizhen Wang, Zhenbing Liu, Chu Han, Lixu Yan, Huan Lin, Zeyan Xu, Zhengyun Feng, Changhong Liang, Xin Chen, Xipeng Pan, and Zaiyi Liu
Research Article (10 pages), Article ID 7966553, Volume 2022 (2022)

A Pulmonary Nodule Spiculation Recognition Algorithm Based on Generative Adversarial Networks
Jing Zhang, Shi Qiu, Xiaohai Cui, and Ting Liang
Research Article (10 pages), Article ID 3341924, Volume 2022 (2022)

Application of Surface-Enhanced Raman Spectroscopy in the Screening of Pulmonary Adenocarcinoma Nodules
Bowen Peng, Huan Yan, Runrui Lin, and Gang Yin
Research Article (10 pages), Article ID 4368928, Volume 2022 (2022)

[Retracted] A Study on Nonvisual Effects of Natural Light Environment in a Maternity Ward of a Hospital in Cold Area
Haihong Song, Wei Jiang, and Peng Cui
Research Article (16 pages), Article ID 8608892, Volume 2022 (2022)

[Retracted] Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma
Shuo Xu, Dingsheng Liu, Mingming Cui, Yao Zhang, Yu Zhang, Shiqi Guo, and Hong Zhang
Research Article (21 pages), Article ID 1893351, Volume 2022 (2022)

Pooling *Operations in Deep* Learning: From "Invariable" to "Variable"
Zhou Tao, Chang XiaoYu, Lu HuiLing, Ye XinYu, Liu YunCan, and Zheng XiaoMin
Review Article (17 pages), Article ID 4067581, Volume 2022 (2022)

Contents

Computed Tomography Image Analysis of Body Fat Based on Multi-Image Information
Wei Zang, Fengrui Zhu, and Yang Yu
Research Article (11 pages), Article ID 8265211, Volume 2022 (2022)

Region-Based Convolutional Neural Network-Based Spine Model Positioning of X-Ray Images
Le Zhang, Jiabao Zhang, and Song Gao
Research Article (15 pages), Article ID 7512445, Volume 2022 (2022)

A Heart Segmentation Algorithm Based on Dynamic Ultrasound
Mingjun Tian and Minjuan Zheng
Research Article (8 pages), Article ID 1485584, Volume 2022 (2022)

Accurate Adiabatic and Diabatic Potential Energy Surfaces for the Reaction of He + H_2
Jing Cao, Nan Gao, Yuxuan Bai, Dequan Wang, Ming Wang, Shaokang Shi, Xinyu Yang, and Xuri Huang
Research Article (15 pages), Article ID 7552881, Volume 2022 (2022)

Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review
Shah Hussain, Iqra Mubeen, Niamat Ullah, Syed Shahab Ud Din Shah, Bakhtawar Abduljalil Khan, Muhammad Zahoor, Riaz Ullah, Farhat Ali Khan, and Mujeeb A. Sultan
Review Article (19 pages), Article ID 5164970, Volume 2022 (2022)

Texture Analysis of Enhanced MRI and Pathological Slides Predicts EGFR Mutation Status in Breast Cancer
Tianming Du and Haidong Zhao
Research Article (15 pages), Article ID 1376659, Volume 2022 (2022)

APU-Net: An Attention Mechanism Parallel U-Net for Lung Tumor Segmentation
Tao Zhou, YaLi Dong, HuiLing Lu, XiaoMin Zheng, Shi Qiu, and SenBao Hou
Research Article (15 pages), Article ID 5303651, Volume 2022 (2022)

The Reaction Mechanism Study for the F_3 System
Dequan Wang, Nan Gao, Hongmei Yu, Yuxuan Bai, Jing Cao, Chunmei Hu, Yanchun Li, Huiling Liu, and Xuri Huang
Research Article (7 pages), Article ID 7088063, Volume 2022 (2022)

Dense Convolutional Network and Its Application in Medical Image Analysis
Tao Zhou, XinYu Ye, HuiLing Lu, Xiaomin Zheng, Shi Qiu, and YunCan Liu
Review Article (22 pages), Article ID 2384830, Volume 2022 (2022)

[Retracted] Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation
Xinyu Lv, Na Ta, Tao Chen, Jing Zhao, and Haicheng Wei
Research Article (8 pages), Article ID 7020804, Volume 2022 (2022)

Habitat Analysis of Breast Cancer-Enhanced MRI Reflects BRCA1 Mutation Determined by Immunohistochemistry

Tianming Du and Haidong Zhao

Research Article (9 pages), Article ID 9623173, Volume 2022 (2022)

Hindawi
BioMed Research International
Volume 2024, Article ID 9878053, 1 page
https://doi.org/10.1155/2024/9878053

Retraction

Retracted: Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Lv, N. Ta, T. Chen, J. Zhao, and H. Wei, "Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation," *BioMed Research International*, vol. 2022, Article ID 7020804, 8 pages, 2022.

Hindawi
BioMed Research International
Volume 2024, Article ID 9867136, 1 page
https://doi.org/10.1155/2024/9867136

Retraction

Retracted: A Study on Nonvisual Effects of Natural Light Environment in a Maternity Ward of a Hospital in Cold Area

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Song, W. Jiang, and P. Cui, "A Study on Nonvisual Effects of Natural Light Environment in a Maternity Ward of a Hospital in Cold Area," *BioMed Research International*, vol. 2022, Article ID 8608892, 16 pages, 2022.

Hindawi
BioMed Research International
Volume 2024, Article ID 9793629, 1 page
https://doi.org/10.1155/2024/9793629

Retraction

Retracted: Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Xu, D. Liu, M. Cui et al., "Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma," *BioMed Research International*, vol. 2022, Article ID 1893351, 21 pages, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 4431536, 8 pages
https://doi.org/10.1155/2022/4431536

Research Article

Super-Resolution Swin Transformer and Attention Network for Medical CT Imaging

Jianhua Hu,[1] Shuzhao Zheng,[1] Bo Wang,[1] Guixiang Luo,[2] WoQing Huang,[1] and Jun Zhang[1]

[1]*Computer Engineering Technical College, Guangdong Polytechnic of Science and Technology, Zhuhai Guangdong, China*
[2]*School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou Guangdong, China*

Correspondence should be addressed to Jun Zhang; 546949043@qq.com

Received 19 August 2022; Revised 20 September 2022; Accepted 28 September 2022; Published 8 December 2022

Academic Editor: Chen Li

Copyright © 2022 Jianhua Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computerized tomography (CT) is widely used for clinical screening and treatment planning. In this study, we aimed to reduce X-ray radiation and achieve high-quality CT imaging by using low-intensity X-rays because CT radiation is damaging to the human body. An innovative vision transformer for medical image super-resolution (SR) is applied to establish a high-definition image target. To achieve this, we proposed a method called swin transformer and attention network (STAN) that uses the swin transformer network, which employs an attention method to overcome the long-range dependency difficulties encountered in CNNs and RNNs to enhance and restore the quality of medical CT images. We adopted the peak signal-to-noise ratio for performance comparison with other mainstream SR reconstruction models used in medical CT imaging. Experimental results revealed that the proposed STAN model yields superior medical CT imaging results than the existing SR techniques based on CNNs. The proposed STAN model employs a self-attention mechanism to more effectively extract critical features and long-range information, hence enhancing the quality of medical CT image reconstruction.

1. Introduction

Computerized tomography (CT) images are used by doctors in clinical practice to judge a patient's condition. Good image quality is crucial for effective and accurate screening and diagnosis of a patient's condition [1]. CT imaging played a vital role in the diagnosis and treatment of COVID-19 [2–4]. CT images are obtained using X-rays. However, X-ray radiation is harmful to the human body. Therefore, to reduce the auxiliary radiation, the X-ray intensity is reduced during operation, resulting in low resolution and blurring of CT images. Therefore, how to obtain high-definition medical CT images through superresolution (SR) is an important research topic. For high-quality medical CT images, plenty of deep learning-(DL-) based SR techniques have been proposed [4–6].

Convolutional neural networks (CNNs) have been used to accomplish SR tasks. Initially, the SRCNN network and CNN were used for performing high-resolution reconstruction tasks [7]. This is the earliest reconstruction from low-resolution to high-resolution images by using CNNs and point-to-point nonlinear feature mapping and reconstruction. Currently, DL is widely used in SR applications [8–10]. Shan et al. [11] improved the initial CNN-based SR method by introducing residual learning and attention mechanism.

Powerful reconstruction algorithms have been proposed to improve SR capability. In 2016, FSRCNN [12] was proposed to improve the SRCNN model, and upsampling was performed to increase the running speed. Many CNN-based SR algorithms have been proposed to improve residual learning, attention mechanism, model depth, speed, complexity reduction, and SR performance [13–22].

CNNs are used in the mainstream medical CT image SR algorithms because they provide a very high-performance advantage for the image domain. However, CNNs cannot realize long-range feature extraction. The transformer is mostly employed in the audio industry, but it has been, recently, employed for SR as a replacement for CNNs [23, 24] because the transformer can support long-range feature extraction by using a self-attention (SA) mechanism and yields very good

performance in the image domain. Transformer models in the field of medical imaging have been extensively studied [25]. The transformer network and the shared attention approach to limit feature extraction improve the SR performance. Nevertheless, few studies have used the transformer network to improve the SR of medical images. As such, in this study, we attempted to use the transformer network for SR reconstruction of medical CT images.

To reconstruct medical CT images, in this study, we developed the swin transformer and attention network (STAN) model. The main advantage of STAN is that it can learn feature information better. STAN consists of three types of blocks: the low-frequency feature extraction block, deep feature extraction block (including attention transformer blocks (ATBs)), and high-resolution image reconstruction block.

To preserve the low-frequency information, the low-frequency feature extraction block is directly connected to the reconstructed model. The deep feature extraction module mainly consists of ATBs. To extract image edge and texture information, a shift window size is used, which reduces resource consumption. Finally, in the high-resolution image reconstruction block, the features of the first two models are first obtained, multilayer feature fusion is performed, and finally, low-resolution to high-resolution reconstruction is realized.

The main contributions of this study are as follows:

(1) A swin transformer is proposed in this paper for an SR network of medical CT images. The use of the attention mechanism improves the network's ability to extract features and edge and flat area information from medical CT images and reconstruct high-quality CT images

(2) We developed a low-frequency extraction module with an attention mechanism to capture the long-range dependency feature of the image

(3) To handle long-range dependency images, we used a shift window mechanism, overcoming the traditional transformation limitation of dividing the input image into fixed-size patches

2. Related Work

The traditional SR algorithm uses a bicubic interpolation algorithm to upsample an image and has the disadvantages of losing details and blurring the image. Therefore, neural networks have been employed for SR. The transformer network can further improve the performance of a traditional CNN. With the development of SR, many scholars have applied the SR technology to improve the clarity and reliability of medical CT images by employing the following three approaches:

(1) Obtaining SR images by using CNNs: CNNs are mainly used to perform transformations between images of different resolutions (e.g., LR image to HR image). Due to the different characteristics of the image, different image scaling methods need to be used to recover different image details. Therefore, nonlinear mapping is performed to recover the lost high-frequency details. CNNs are widely used to reconstruct high-quality images and realize SR through dense connection convolution, multichannel networks, and symmetric jump connections [26, 27]

(2) The use of transformer networks in the field of image applications: transformer networks are generally used in the audio field. Because of their local attention mechanism and long-term compliance, transformer networks are highly suitable for image feature extraction. Therefore, the transformer [28, 29] is widely used in the field of image processing because of its ability to better access information and integrate the CNN and transformer. Pan et al. [23] proposed a high-quality reconstruction transformer to capture image global features for medical CT image reconstruction

(3) The use of SR in the field of medical CT imaging: DL technology is extensively employed for medical CT imaging [30–32]. Many scholars have applied SR technology to the medical field [33–35]. SR technology is used to reconstruct high-definition images for the characteristics of medical images, which can effectively improve image quality and reduce X-ray radiation to the human body

In this study, we designed the STAN model to reconstruct medical CT images. In addition, we introduced a self-information mechanism in the network model to enable updates to be performed on long-range information; moreover, the medical image quality smoothing area enables better image quality.

3. Methods

The architecture of the medical CT image performance enhancement network is presented here. For the SR reconstruction of medical CT images, a transformer and an attention network are employed. To improve the extraction of low-frequency and high-frequency medical feature information, we designed the STAN model. We used a transformer network instead of traditional CNNs to considerably increase the quality of medical CT images and edge information. The proposed system comprises three types of blocks: low-frequency feature extraction block, deep feature extraction block, and high-resolution image reconstruction block.

3.1. Network Architecture. The structure of the proposed STAN model is illustrated in Figure 1. STAN employs an efficient long-range attention transformer network for reconstructing high-resolution images from low-resolution medical CT images. STAN includes the low-frequency feature extraction block (for the extraction of flat area image information), deep feature extraction block (including six ATBs), and high-resolution CT image reconstruction block (for the feature extraction of image edge information). Low-resolution images are inputted into the STAN model.

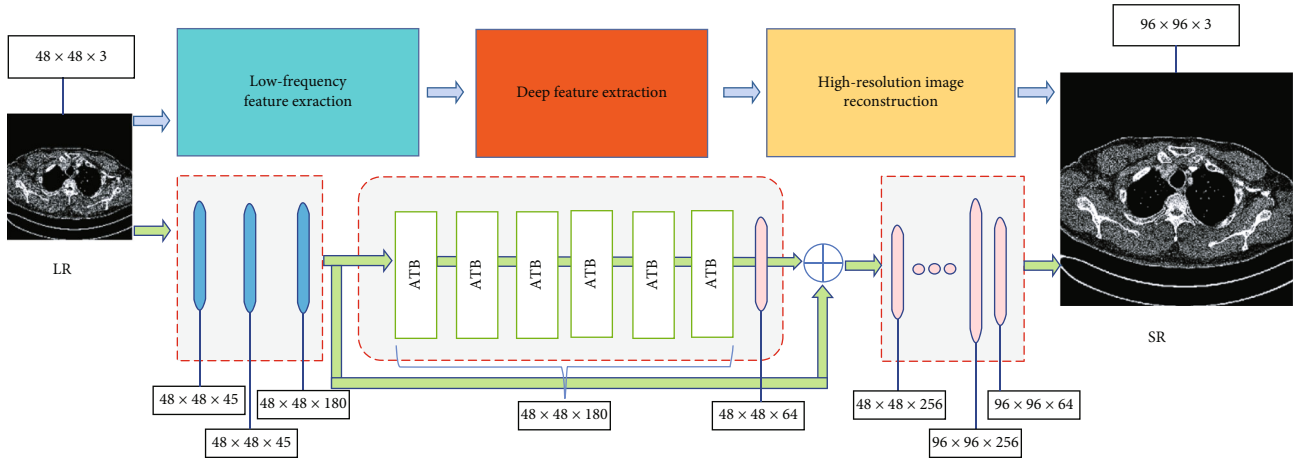


Figure 1: The architecture of the proposed STAN model. The STAN includes the low-frequency feature extraction block (for the extraction of flat area image information), deep feature extraction block (including six ATBs), and high-resolution CT image reconstruction block (for the feature extraction of image edge information).

Input: low-resolution image input I_{LQ}
Output: high-resolution image output I_{RHQ}
1: $F_0 = H_{LF} * I_{LQ}$
2: **For** $i = 1 \longrightarrow M$**do**//i denotes the i-th ATB, here M is 6
3: $//F_i = H_{ATB}(F_i)$
4: **For** $j = 1, 2, \cdots, L$ **do**//j denotes the j-th STL layer, L is 6
5: $F_{i,j} = H_{STLi,j}(F_{i,j} - 1)$
6: **End for**
7: $F_{i,\text{out}} = H_{CONVi}(F_{i,L}) + F_{i,0}$
8: **End for**
9: $F_{DF} = H_{CONV}(F_M)$
10: $I_{RHQ} = H_{REC}(F_0 + F_{DF})$
11: **Return** I_{RHQ}//high-resolution image output

Algorithm 1: The implementation steps of our STAN for medical superresolution.

The low-frequency feature extraction block extracts the low-frequency feature information from medical CT images by using multilayer CNNs. The deep feature extraction block employs the self-attention mechanism transformer network to extract the edge information of medical CT images, and multichannel image information is obtained by adding to the previous network. The high-resolution CT image reconstruction block fuses the low-frequency and high-frequency information; in addition, it extracts the characteristic data of multiple channels and upsamples the image to obtain an SR medical CT image.

The proposed STAN algorithm is shown below. The low-resolution input image is I_{LQ}. Transform H_{LF} is a low-frequency feature extraction block. In the deep feature extraction block, there are M ATBs, and each ATB has L STLs and a convolution operation. After the deep feature extraction block, the high-quality picture is reconstructed through high-resolution CT image reconstruction.

3.2. Low-Frequency Feature Extraction Block. As shown in Figure 2, the low-frequency feature extraction block realizes low-frequency information extraction and includes three layers. A low-resolution image is input in. After feature extraction by using 3×3 convolution operations, fine feature extraction is performed using 1×1 convolution. Finally, the low-frequency information extraction output of the current block is obtained using a convolution kernel size of 3×3.

A low-resolution image is input as I_{LQ}. Then, two 3×3 and one 1×1 convolution layer are used to obtain the low-frequency feature output as follows:

$$F_0 = H_{LF}(I_{LQ}). \tag{1}$$

This module uses a multilayer network to better accomplish the extraction of low-frequency information.

3.3. Deep Feature Extraction Block. Deep features $F_{DF} \in R^{H \times W \times C}$ are extracted from the low-frequency feature output F_0 as follows:

$$F_{DF} = H_{DF}(F_0), \tag{2}$$

where comprises six ATBs. The composition and principle of the ATB are described in detail here.

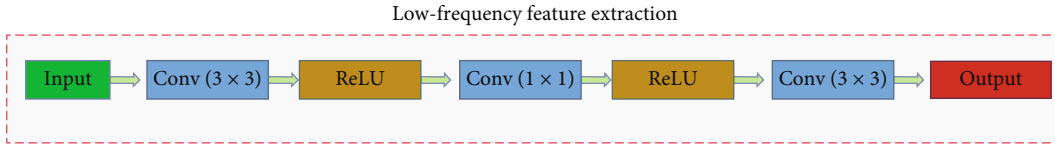


Figure 2: Low-frequency feature extraction block.

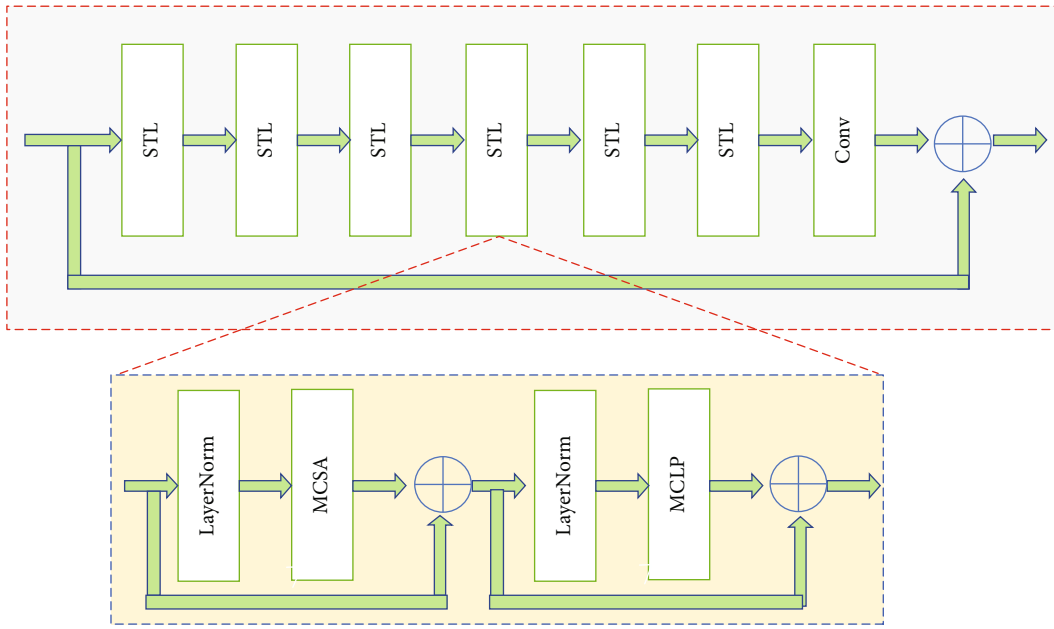


Figure 3: Attention transformer block.

As shown in Figure 3, ATBs are composed of swin transformer layers (STLs) and convolutional layers with self-awareness functions. The STL is the base component of the ATB. The base network comprises multiple STLs and ends with a convolutional layer to form the ATB. In this study, the number of STLs in an ATB was set as 6 to achieve a balance between extraction performance and model complexity.

$F_{i,0}$ indicates the i-th ATB. Information features $F_{i,1}$, $F_{i,2}$, …, $F_{i,L}$ are extracted by the ATB layers as follows:

$$F_{i,j} = H_{\mathrm{STL}_{i,j}}\left(F_{i,j-1}\right), j = 1, 2, \cdots, L, \tag{3}$$

where $H_{\mathrm{STL}_{i,j}}(\cdot)$ is the i-th ATB and J denotes the j-th STL. This design offers two advantages: spatial variation convolution and residual connection reconstruction module.

The output of ATB can be formulated as

$$F_{i,\mathrm{out}} = H_{\mathrm{CONV}_i}(F_{i,L} + F_{i,0}), \tag{4}$$

where $H_{\mathrm{CONV}_i}(\cdot)$ is the i-th ATB swin transformer.

The STL enables the self-attentive mechanism through the transformer layer. Its most important feature is the use of local attention and shift window mechanism. By the size $H \times W \times C$, the ATB splits the input into nonoverlapping $M \times M$ local windows. In this manner, the input size is reshaped into the $(HW/M^2) \times M^2 \times C$ feature, where HW/M^2 is the number of windows.

The STL consists of three components: layer specification (LN) layer (used for regularization), multicontrol head SA (MCSA) layer, and multilayer control perceptron (MLCP) layer. The MLCP layer is composed of two completely connected neural networks, and feature extraction is performed between them through nonlinear transformation. The LN layer is added before the MCSA and MLCP layers, and then, the residuals are used to connect the two modules. The process is as follows:

$$X = \mathrm{MCSA}(LN(X)) + X, \tag{5}$$

$$X = \mathrm{MCLP}(LN(X)) + X. \tag{6}$$

3.4. High-Resolution Image Reconstruction Block. The low-frequency feature extraction of high-quality images is performed from medical CT images according to the convergence:

$$I_{\mathrm{RHQ}} = H_{\mathrm{REC}}(F_0 + F_{\mathrm{DF}}), \tag{7}$$

where $H_{\mathrm{REC}}(\cdot)$ indicates the reconstruction model. The low-frequency information mainly includes low frequencies, while the deep features are used to repair the missing high frequencies. Sloshing inverter circuits are used to transmit low-frequency information to the medical CT image

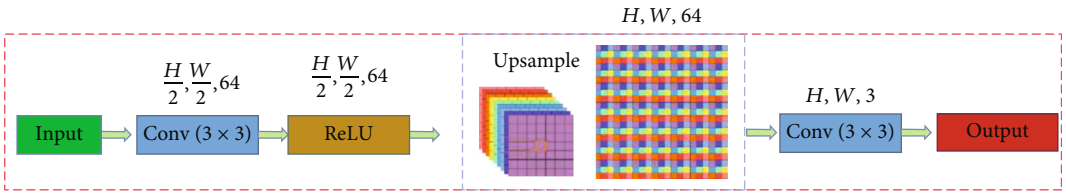


Figure 4: High-resolution image reconstruction blocks.

reconstruction module through a high skip connection and help the deep-level feature collection module to focus on high-frequency information.

The high-resolution image reconstruction block (Figure 4) comprises a 64-channel CNN with channel size $H/2, w/2, 64$. The 64-channel feature map output is obtained using the pixel shuffle upsampling method. Finally, a 3-channel CNN is used to generate the high-definition image output.

The primary function of pixel shuffle is to convert the multichannel feature map $r * r$ into size of $w * r$ and $h * r$ (e.g., the original feature map size is $4 \times 128 \times 128$, which is then adjusted to size $1 * (128 \times 2) * (128 \times 2)$), where r is the upsampling factor, which is the magnification of the image.

4. Results

We evaluated the performance of the proposed model on open-source datasets and evaluated the image quality by using the peak signal-to-noise ratio (PSNR) metric. Compared with other advanced SR methods, the proposed model offers obvious performance advantages.

4.1. Dataset. We used the largest medical CT medical image dataset, DeepLesion [36], for training and testing the model. This dataset not only includes key CT slices containing the important lesions but also provides the three-dimensional context (additional slices of 30 mm above and below the key slices). The size of the dataset is 221 GB. Because of the huge amount of data, 11,500 high-quality CT images were randomly selected and divided into three parts. The majority of the images were used for training (10,000), and the remaining were used for verification (1000) and testing (500). This dataset consists of the original image and the downsampled image through bicubic interpolation by using the function torchvision.transforms.resize() in the PyTorch library. The source HR medical CT image was reduced to a LR image as the input data, and the original HR medical CT medical image was used as the data label to be used as the input dataset of the DL neural network for training. For the sake of accuracy of model training, the training set was added through data enhancement to improve the generalization ability.

4.2. Implementation Details. The three-channel (RGB) pixels of the input image and the original data were linearly reduced to obtain the LR image, and the original data were used as label data and inputted into the network. Six ATBs were used. The sliding window size of each transform network was set as 8, and the patch size corresponding to the LR image resolution was 48.

Table 1: Ablation study on ATB design.

Number of ATBs	ATB number X4	PSNR
1	√	23.3
2	√	25.78
4	√	28.967
6	√	31.34
8	√	31.90
10	√	32.02

Adam optimizer was adopted with two improvements: gradient sliding average and bias correction. The learning rate decayed with each update factor decay set as 0.999, and the initial learning rate was 2×10^{-4}. The pixel shuffle method was used for image upsampling.

4.3. Evaluation Index. We evaluated the reconstructed SR images by using two methods: subjective evaluation and objective evaluation. Many factors influence subjective evaluation, and the reconstructed SR images are evaluated mainly based on human visual perception. In this study, the PSNR was measured as the objective evaluation metric to study the performance of high-resolution restoration networks for medical CT images. To demonstrate the superiority of the proposed model visually, we calculated the PSNR values of the SR images generated using the proposed method and other methods and compared them.

PSNR is an objective criterion for evaluating images. The calculation method is as follows:

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|J(i,j) - L(i,j)\|^2. \tag{8}$$

Then, PSNR can be obtained as follows:

$$\mathrm{PSNR} = 10\log_{10}\frac{\mathrm{MaxValue}^2}{\mathrm{MSE}} = 10\log_{10}\frac{2^{\mathrm{bits}}-1}{\mathrm{MSE}}, \tag{9}$$

where J and L are the two pixel values and the size of the image is $m \times n$. The greater the PSNR, the better the medical CT image effect, and vice versa.

4.4. Ablation Study. To better understand how STAN performs SR in medical CT images, a comprehensive ablation study on ATBs was performed to evaluate the role of key

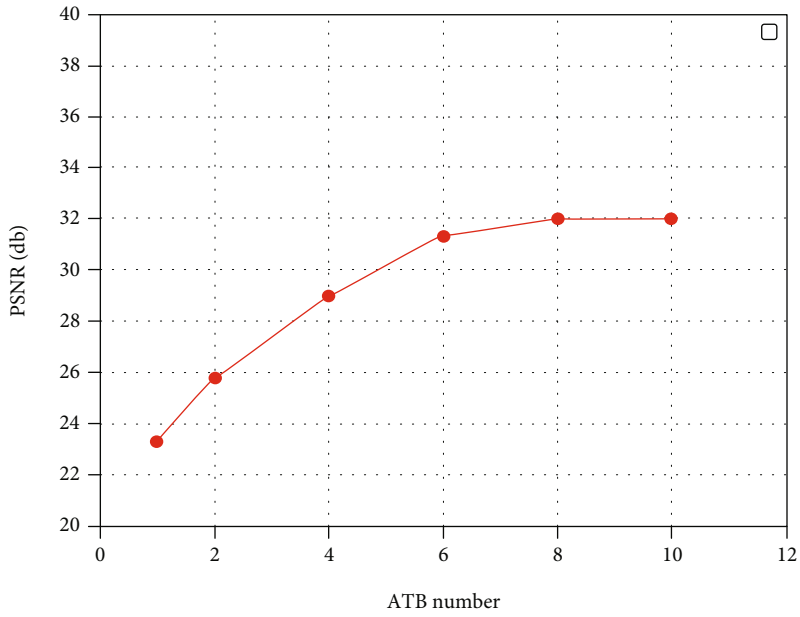


Figure 5: Relationship between PSNR and number of ATBs in STAN.

parts of the proposal STAN model, as well as the degree of depth and the choice of shared attention mechanism.

As can be seen in Table 1, we studied the effect of the removal and addition of ATB modules on the performance of the medical CT image reconstruction network. To analyze the effect of the low-frequency feature extraction block and ATB on the performance of the STAN model, we conducted ablation experiments by using different numbers of modules and studied their corresponding PSNR performance under the ×4 scaling condition. The number of ATBs affects the PSNR, i.e., the higher the number of ATBs, the higher the PSNR.

As shown in Figure 5, we studied the relationship between the number of ATBs and the PSNR performance on DeepLesion for image SR (×4). To obtain a relatively lightweight model, the number of ATBs was selected as 6, and the number of convolutional layers was 3 in the final test performance experiment.

4.5. Analysis of Experimental Results. Network optimization was performed. The performance comparison results in terms of PSNR with ×2 and ×4 scale factors are presented in Table 2. We analyzed different algorithms on the DeepLesion testing set. Compared with the bicubic method, the PSNR of STAN improved by 9.58 and 13.36 dB when the scale factor was ×2 and ×4, respectively. Compared with the method using the DL neural network, the PSNR of STAN improved by 3.81 and 3.56 dB when the scale factor was ×2 and ×4, respectively.

As can be seen in Figure 6, the bicubic reconstruction of medical CT images yielded the worst effect and the lowest PSNR. The SR algorithm based on DL performed better than the algorithm based on interpolation. The STAN model based on transform networks proposed in this paper performed relatively better than the CNN-based SR method by 0.76 and 0.23 dB when the scale factor was ×2 and ×4, respectively.

Table 2: PSNR (dB) values for the proposed STAN model.

Model	×2	×4
Bicubic	23.32	21.76
SRCNN [37]	33.17	27.78
DRRN [38]	34.56	29.65
MDSR [39]	34.85	29.90
RDN [40]	34.96	30.24
RCAN [41]	34.99	30.35
LTE [42]	36.22	31.11
STAN (our)	36.98	31.34

Thus, the proposed STAN model exhibited superior performance to the CNN-based SR method, demonstrating that the transformer network yields obvious performance advantages in medical CT imaging.

By using different algorithms, the medical CT image was reconstructed with multiple resolutions. The results of different algorithms on the DeepLesion are shown in Figure 6.

5. Conclusions

For SR of medical CT images, we proposed an improved STAN model that uses the SA mechanism for feature extraction and solves the long-range dependency problem encountered in CNNs and RNNs. In addition, it can obtain more important feature information. In STAN, nonoverlapping feature values are computed using different window sizes, and feature extraction is performed using a shared-attention mechanism.

We experimentally demonstrated the SR effectiveness of the proposed STAN model in medical CT images. We used the PSNR metric for performance comparison. The results

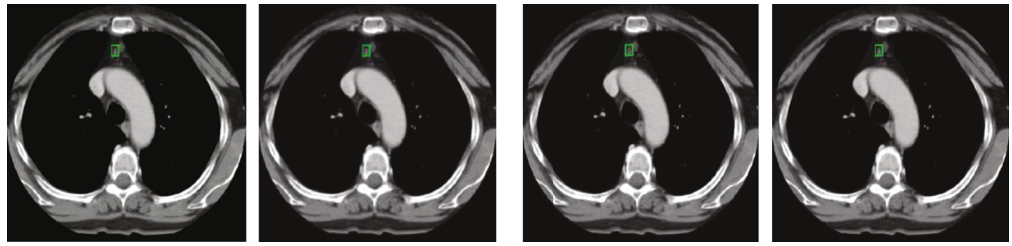


Figure 6: Medical CT image testing using different algorithms on the DeepLesion dataset.

revealed that the PSNR of the proposed STAN model is much better than that of the CNN SR method. The use of the SA mechanism in STAN yields clearer reconstruction results, and the reconstruction effect in the low-frequency regions of medical CT images is better. However, medical imaging may generate image noise due to the influence of hardware equipment and the external environment. As such, the next step is to denoise medical CT images in the SR process.

Data Availability

The medical CT medical image data used to support the findings of this study have been deposited in the https://nihcc.app.box.com/v/DeepLesion repository. This is an open source medical open source dataset, you can download and use it freely.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the Foundation for 2022 Basic and Applied Basic Research Project of Guangzhou Basic Research Plan (research on video compression algorithm based on dual neural network, Grant 202201011753), the 2020 University Industry University Research Innovation Fund New Generation Information Technology Innovation Project of China (key project, Grant 2020ITA03004), the 2021 characteristic innovation project of the Department of Education of Guangdong Province (2021KTSCX217), the Research on Classified and Accurate Training of Higher Vocational IT Talents based on Education Big Data Under the Background of Enrolment Expansion (Grant 2021GXJK714), and the Innovative Research Team in Universities of Guangdong Province of China (Grant 2021KCXTD079).

References

[1] D. J. Brenner, C. D. Elliston, E. J. Hall, and W. E. Berdon, "Estimated risks of radiation-induced fatal cancer from pediatric CT," *American Journal of Roentgenology*, vol. 176, no. 2, pp. 289–296, 2001.

[2] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: a review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.

[3] C. Long, H. Xu, Q. Shen et al., "Diagnosis of the coronavirus disease (COVID-19): rRT-PCR or CT?," *European Journal of Radiology*, vol. 126, article 108961, 2020.

[4] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[5] Z. Chen, X. Guo, P. Y. Woo, and Y. Yuan, "Super-resolution enhanced medical image diagnosis with sample affinity interaction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1377–1389, 2021.

[6] M. I. Georgescu, R. T. Ionescu, A. I. Miron et al., "Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution," 2022, https://arxiv.org/abs/2204.04218.

[7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016.

[8] L. S. Passarella, S. Mahajan, A. Pal, and M. R. Norman, "Reconstructing high resolution ESM data through a novel fast super resolution convolutional neural network (FSRCNN)," *Geophysical Research Letters*, vol. 49, no. 4, 2022.

[9] X. Wang, K. Yu, S. Wu et al., "Enhanced deep residual networks for single image super-resolution," *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018.

[10] I. Lodha, L. Kolur, K. Krishnan, K. Dheenadayalan, D. Sitaram, and S. Nandi, "Cost-optimized video transfer using real-time super resolution convolutional neural networks," in *5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD)*, Bangalore, India, 2022.

[11] L. Shan, X. Bai, C. Liu, Y. Feng, Y. Liu, and Y. Qi, "Super-resolution reconstruction of digital rock CT images based on residual attention mechanism," *Advances in Geo-Energy Research*, vol. 6, no. 2, pp. 157–168, 2022.

[12] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proceedings of the European Conference on Computer Vision (ECCV 2016)*, pp. 391–407, Amsterdam, The Netherlands, 2016.

[13] M. Ta and B. Ylmaz, "Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images," *Computers & Electrical Engineering*, vol. 90, no. 12, article 106959, 2021.

[14] A. Deshpande, V. V. Estrela, and P. Patavardhan, "The DCT-CNN-ResNet50 architecture to classify brain tumors with super-resolution, convolutional neural network, and the ResNet50," *Neuroscience Informatics*, vol. 1, no. 4, article 100013, 2021.

[15] M. O. Yaroshenko, A. Y. Varfolomieiev, and P. O. Yaganov, "Hierarchical convolutional neural network for infrared image super-resolution," *Microsystems Electronics and Acoustics*, vol. 26, no. 1, 2021.

[16] D. Qiu, L. Zheng, J. Zhu, and D. Huang, "Multiple improved residual networks for medical image super-resolution," *Future Generation Computer Systems*, vol. 116, pp. 200–208, 2021.

[17] Q. Chang, X. Jia, C. Lu, and J. Ye, "Multi-Attention Residual Network for Image Super Resolution," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 36, no. 8, article 2254009, 2022.

[18] K. G. Brown, S. C. Waggener, A. D. Redfern, and K. Hoyt, "Faster super-resolution ultrasound imaging with a deep learning model for tissue decluttering and contrast agent localization," *Biomedical Physics & Engineering Express*, vol. 7, no. 6, article 065035, 2021.

[19] X. Wang, K. Yu, S. Wu et al., "Esrgan: enhanced super-resolution generative adversarial networks," in *European Conference on Computer Vision Workshops*, pp. 701–710, Munich, Germany, 2018.

[20] J. Zhang, X. Zou, L. D. Kuang, J. Wang, R. S. Sherratt, and X. Yu, "CCTSDB 2021: a more comprehensive traffic sign detection benchmark," *Human-centric Computing and Information Sciences*, vol. 12, 2022.

[21] Y. Chen, L. Liu, V. Phonevilay et al., "Image super-resolution reconstruction based on feature map attention mechanism," *Applied Intelligence*, vol. 51, no. 7, pp. 4367–4380, 2021.

[22] J. Zhang, C. Li, S. Kosov et al., "LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, p. 107885, 2021.

[23] J. Pan, H. Zhang, W. Wu, Z. Gao, and W. Wu, "Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction," *Patterns*, vol. 3, no. 6, article 100498, 2022.

[24] B. Zhou, J. Schlemper, N. Dey et al., "DSFormer: a dual-domain self-supervised transformer for accelerated multi-contrast MRI reconstruction," 2022, https://arxiv.org/abs/2201.10776.

[25] F. Shamshad, S. Khan, S. W. Zamir et al., "Transformers in medical imaging: a survey," 2022, https://arxiv.org/abs/2201.09873.

[26] Y. Zou, F. Xiao, L. Zhang, Q. Chen, B. Wang, and Y. Hu, "Image super-resolution using convolutional neural network with symmetric skip connections," in *Fourth International Conference on Photonics and Optical Engineering*, vol. 11761, pp. 283–288, Xi'an, China, 2021 Jan 15.

[27] Y. Wang, S. Zeng, X. Wang, and J. Wang, "Super-resolution multi-focus image fusion based on convolutional neural network," *Journal of Physics: Conference Series*, vol. 1885, no. 2, article 022011, 2021.

[28] Z. Liu, Y. Lin, Y. Cao et al., "Swin transformer: hierarchical vision transformer using shifted windows," 2021, https://arxiv.org/abs/2103.14030.

[29] H. Chen, C. Li, X. Li et al., "Gashis-transformer: a multi-scale visual transformer approach for gastric histopathology image classification," 2021, https://arxiv.org/abs/2104.14528.

[30] R. Kumar, A. A. Khan, J. Kumar et al., "Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 16301–16314, 2021.

[31] K. He, X. Liu, M. Li, X. Li, H. Yang, and H. Zhang, "Noninvasive KRAS mutation estimation in colorectal cancer using a deep learning method based on CT imaging," *BMC medical imaging*, vol. 20, no. 1, p. 59, 2020.

[32] W. Wu, D. Hu, C. Niu et al., "Deep learning based spectral CT imaging," *Neural Networks*, vol. 144, pp. 342–358, 2021.

[33] R. Gupta, A. Sharma, and A. Kumar, "Super-resolution using GANs for medical imaging," *Procedia Computer Science*, vol. 173, pp. 28–35, 2020.

[34] J. Zhu, C. Tan, J. Yang, G. Yang, and P. Lio, "Arbitrary scale super-resolution for medical images," *International Journal of Neural Systems*, vol. 31, no. 10, article 2150037, 2021.

[35] J. Zhang, L. R. Gong, K. Yu, X. Qi, Z. Wen, and Q. Hua, "3D reconstruction for super-resolution CT images in the Internet of health things using deep learning," *IEEE Access*, vol. 8, pp. 121513–121525, 2020.

[36] Y. Ke, X. Wang, L. Le, and R. M. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of medical imaging*, vol. 5, no. 3, article 036501, 2018.

[37] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[38] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition.*, pp. 3147–3155, Honolulu, Hawaii, 2017.

[39] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, Honolulu, Hawaii, 2017.

[40] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2472–2481, Salt Lake City, Utah, 2018.

[41] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 286–301, Munich, Germany, 2018.

[42] J. Lee and K. H. Jin, "Local texture estimator for implicit representation function," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1929–1938, Ernest N. Moral Convention Center, New Orleans, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 7921922, 17 pages
https://doi.org/10.1155/2022/7921922

Research Article

TSHVNet: Simultaneous Nuclear Instance Segmentation and Classification in Histopathological Images Based on Multiattention Mechanisms

Yuli Chen, Yuhang Jia, Xinxin Zhang, Jiayang Bai, Xue Li, Miao Ma, Zengguo Sun, and Zhao Pei

School of Computer Science, Shaanxi Normal University, Xi'an 710119, China

Correspondence should be addressed to Yuli Chen; chenyuli@snnu.edu.cn

Received 9 August 2022; Revised 30 September 2022; Accepted 3 October 2022; Published 22 November 2022

Academic Editor: Chen Li

Copyright © 2022 Yuli Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate nuclear instance segmentation and classification in histopathologic images are the foundation of cancer diagnosis and prognosis. Several challenges are restricting the development of accurate simultaneous nuclear instance segmentation and classification. Firstly, the visual appearances of different category nuclei could be similar, making it difficult to distinguish different types of nuclei. Secondly, it is thorny to separate highly clustering nuclear instances. Thirdly, rare current studies have considered the global dependencies among diverse nuclear instances. In this article, we propose a novel deep learning framework named TSHVNet which integrates multiattention modules (i.e., Transformer and SimAM) into the state-of-the-art HoVer-Net for the sake of a more accurate nuclear instance segmentation and classification. Specifically, the Transformer attention module is employed on the trunk of the HoVer-Net to model the long-distance relationships of diverse nuclear instances. The SimAM attention modules are deployed on both the trunk and branches to apply the 3D channel and spatial attention to assign neurons with appropriate weights. Finally, we validate the proposed method on two public datasets: PanNuke and CoNSeP. The comparison results have shown the outstanding performance of the proposed TSHVNet network among the state-of-art methods. Particularly, as compared to the original HoVer-Net, the performance of nuclear instance segmentation evaluated by the PQ index has shown 1.4% and 2.8% increases on the CoNSeP and PanNuke datasets, respectively, and the performance of nuclear classification measured by $F1$_score has increased by 2.4% and 2.5% on the CoNSeP and PanNuke datasets, respectively. Therefore, the proposed multiattention-based TSHVNet is of great potential in simultaneous nuclear instance segmentation and classification.

1. Introduction

Precise nuclear instance segmentation and classification are the essential foundation for histopathological image analysis, including nuclear morphology characteristics, counts of specific nuclei, and nuclear distribution, which would significantly impact the predictive results of cancer diagnosis and prognosis [1]. However, since there are often tens of thousands of nuclei belonging to different categories and the nuclei are usually adherent to each other in a histopathological image, manual evaluation by pathologists is not only time and labor consuming but also inevitable in intraobserver and interobserver disagreement [2]. Therefore, an efficient automatic model for nuclear instance segmentation and classification in histologic images is urgently needed.

Regarding nuclear instance segmentation, varied shapes, colors, and sizes of nuclei would make it difficult to segment all the nuclear instances in the same rule [3]. Moreover, the massive adherence and overlap of nuclei in the histopathologic images would influence nuclear shape detection and nuclear feature description [4]. Concerning nuclear classification, visual appearances of interclass nuclei could be quite similar while intraclass nuclei may exhibit great differences, which results in difficulty in the nuclear classification [5]. So, it is necessary to develop a model to implement precise nuclear instance segmentation and classification.

In recent years, due to the availability of big data and the significant improvement of computing power, deep convolutional neural networks have developed rapidly and achieved amazing results in various medical image analysis tasks [6]. Regarding nuclear segmentation of histopathological images, Ronneberger et al. [7] proposed a classic U-Net network inspired by a fully convolutional neural network [8]. This network consisted of a shrinking path and an expanding path, bridging by skip connections on the same convolutional level. Jha et al. [9] proposed a ResUNet++ network that combined a Squeeze-and-Excitation Attention block and ASPP with a convolutional neural network to make the network pay more attention to the nuclear regions of interest. Concerning the classification of nuclei, a workflow of a two-stage process is usually employed: one stage for nuclear detection and the other for nuclear classification. Sharma et al. [10] firstly detected nuclei by extracting nuclear features based on color, texture, and morphology and then classified nuclei by using AdaBoost. Shuttleworth et al. [11] analyzed multiresolution textures to achieve nuclear detection and classification in medical histopathological images. It is time costly and computational consumption to separately perform the nuclear segmentation and classification. Therefore, it is necessary to construct a uniform model to simultaneously conduct multi-instance segmentation and classification of nuclei. To this end, Graham et al. [12] proposed a HoVer-Net network performing nuclear segmentation and classification by using horizontal and vertical distance map predictions to separate clustered nuclei. However, due to the inherent local computational property in convolution operations, convolutional neural network-based HoVer-Net would not capture global context information well, which might constrict the accuracy of nuclear segmentation and classification.

In this paper, given the long-distance modeling property of Transformer [13], we propose a new neural network named TSHVNet, incorporating a Transformer attention module into the HoVer-Net in conjunction with an attentional module called SimAM [14], to achieve accurate multi-instance nuclear segmentation and classification in medical histopathological images. Specifically, a convolutional neural network is incapable of extracting global feature information but able to learn local features for localization. In contrast, Transformer could perceive long-distance dependencies but requires many computational resources and is insensitive to position information. To combine their strengths, therefore, in our proposed TSHVNet network, we integrate the Transformer attention module and the HoVer-Net. In addition, SimAM, a 3D attentional mechanism based on neuroscience energy function, is also employed in our network to allocate more attention to the nuclear regions of interest in medical histopathological images.

2. Related Work

2.1. Nuclear Instance Segmentation. In the 1990s, energy-based methods, especially the Watershed algorithm based on markers [15], have been widely utilized to carry out nuclear instance segmentation. For example, Yang et al. [16] used thresholding to obtain markers and energy landscape as inputs to extract nucleus instances, but this approach relied heavily on the difference between the nucleus and the background, which is not suitable for segmenting nuclei of various intensities and may result in unreliable results [17].

Since the 2010s, deep learning-based image segmentation technologies have been rapidly developed. For instance, the U-Net model [3], widely used in medical pathological images, outperformed the fully convolutional neural network (FCN) [8] in pixel-wise segmentation. The U-Net model [3] adopted a U-shaped structure with skip connections and a weighted loss function to take full advantage of the locating information that is important for nuclear instance segmentation. Although U-Net has a good representation capability, it is still hard to separate the highly adherent nuclei. Zhou et al. proposed a UNet++ [18] which alleviated the problem of unknown network depth by integrating U-Net of different depths. The network aggregated features of different semantic scales on the decoder subnetwork to produce highly flexible feature fusion schemes. In addition, Xiao et al. proposed a model named Res-UNet [19] which integrated a weighted attention mechanism into the U-Net. This network can make the network focus only on the target regions of interest and discard irrelevant noise background. Micro-Net [20] extended U-Net with a weighted loss enhancement architecture. The network processed the input at a variety of resolutions, providing an accurate representation of nuclei of different sizes. To separate complex and diverse nuclear boundaries, a hybrid attention block in Han-Net [21] was used to explore attention information and build correlations between different pixels to further expand the U-Net. SONNET [22] was a self-guided ordinal regression neural network for nuclear segmentation, which exploits the intrinsic characteristics of nuclei and focuses on highly uncertain areas during training. Liu et al. [23] proposed an Att-MoE (Attention-based Mixture of Experts) model for nuclear and cytoplasmic segmentation in fluorescent histology images, which integrates multiple expert networks using a single gating network.

Other deep learning approaches take advantage of the information of nuclear edges and centroids to segment nuclei. For example, DCAN [24] designed a dual architecture to output two independent prediction graphs of nuclear cluster and nuclear contour and then subtracted the contour from the nuclear cluster to achieve nuclear instance segmentation. In addition, there are some centroid-based nuclear segmentation methods. For example, Graham et al. [12] proposed a HoVer-Net that performed nuclear instance segmentation by using horizontal and vertical distance map predictions to separate clustered nuclei. Moreover, Cheng et al. [25] built a GAN-based model for color variation and constructed a model based on HoVer-Net with cost-sensitive loss that guides the model to pay more attention to the minority classes.

Since there are thousands of nuclei in a histopathological image, it will be a challenge to segment a large number of nuclear instances due to the inevitable nuclear adhesion

and clustering. Although most deep convolutional neural networks perform well in extracting nuclear features from local regions, few networks take into account the global long-distance dependence of diverse nuclei. The long-distance dependencies provide a broad view to discover features that may improve the accuracy of nuclear segmentation and classification.

2.2. Nuclear Classification. To facilitate the subsequent analysis of diagnosis and prognosis of diseases, it is necessary to determine the type of each nucleus in histopathologic images. Most existing methods treat nuclear classification as a two-stage task: firstly to conduct a nuclear segmentation or detection process and then to perform the nuclear classification task. Moreover, a series of morphological characteristics of the nuclei could be extracted after carrying out the segmentation operation, which would further benefit nuclei classification. For example, Nguyen et al. [26] divided the nuclei in breast cancer images into tumors and lymphocytes based on nuclear morphological features. Sirinukunwattana et al. [27] proposed a spatially constrained CNN to detect all the nuclei in the images at first and then input the patches of nuclei into a classifier for further nuclear classification. In addition, Zhao et al. [28] proposed a semisupervised learning method for the accurate classification of cervical cells using only a small amount of labeled data by introducing manual features and a voting mechanism to achieve data expansion. Recently, some studies focus on simultaneous nuclear instance segmentation and classification, such as HoVer-Net [12] and DT-Yolact (Double-Tower Yolact) [29], but they did not consider extracting the features of nuclear instances from the perspective of global dependences and attention mechanism.

2.3. Attentional Mechanism. In recent years, increasing researchers have tried to combine the attention mechanism with CNN to improve network performance [30]. Attention mechanisms can be roughly grouped as spatial attention and channel attention. For instance, employing the spatial attention mechanism, Google DeepMind proposed a network named STN [31] to transform feature maps in space automatically, and STN [31] had invariance properties in translation, rotation, and scaling. Jie et al. [32] proposed a channel-based squeeze-and-congestion network SENet which learned the interdependency of feature channels and automatically assigned different weights to each feature channel for the promotion of the useful features and the suppression of the useless ones. Yang et al. [14] proposed an attention mechanism termed SimAM, which integrated both the channel and spatial attention mechanisms. SimAM can assign different weights to each neuron, which helps the network pay more attention to the area of interest.

3. Method

The proposed network TSHVNet is aimed at predicting instance-level nuclear category and the corresponding pixel-level classification simultaneously, to facilitate the accurate analysis of nuclear morphology in histopathological images.

3.1. Architecture of TSHVNet. TSHVNet is a deep learning neural network for simultaneous nuclear instance segmentation and classification. Figure 1 shows the whole network structure of the proposed TSHVNet. The TSHVNet consists of a trunk and three branches. The trunk of TSHVNet is responsible for feature extraction and encoding. The three branches of TSHVNet are in charge of feature decoding.

To achieve precise segmentation of nuclear regions of interest and accurate distinction among various nuclei, we introduce a Transformer attention module to the trunk of the HoVer-Net [12] to explore the potential global dependencies between the network input and outputs. Besides, several SimAM attention modules are added to both the trunk and branches of the HoVer-Net to discover the important information in the channel and space on feature maps. Specifically, the output of SimAM is served as the input of Transformer attention module. Subsequently, the Transformer module is followed by a DecoderCup module consisting a reshape function and operations of Conv2d, ReLU, and BatchNormal. Hereinto, the reshape function restores the features from the size of $H * W/P^2$ to $H/P * W/P$, where W denotes width, H represents height, and P means the length of each patch $P \times P$.

The three branches decode the features yielded frome DecoderCup module on the trunk of TSHVNet in different ways. The first branch is nuclear pixel (NP) branch, which predicts whether each pixel belongs to the nucleus or background. The second branch is the HoVer (HV) branch, which predicts the horizontal and vertical distances of the nuclear pixels to the centroid of the nucleus. The third branch is the nuclear classification (NC) branch, which predicts the nuclear type of each pixel and determines the category of each nucleus.

The parameters and output dimensions of each residual block, Transformer module, and each branch of the proposed TSHVNet network are measured as shown in Table 1.

3.2. Components of TSHVNet

3.2.1. Trunk Subnetwork. Given the excellent performance of the 50-layer preactivated residual (Preact-ResNet 50) units on visual tasks and robustness to input disturbances [33], the trunk of the proposed TSHVNet employs four blocks of Preact-ResNet 50 units to extract features. The Preact-ResNet 50 units reduce the total downsampling factor from 32 to 8 via convolutions of 1×1 kernels and removing the subsequent max-pool operations. This operation effectively avoids the immediate loss of the information important for segmentation. The feature extraction network is composed of four different residual blocks, and the number of residual units in each residual block is 3, 4, 6, and 3, which are applied downsample levels of 1, 2, 4, and 8, respectively. Moreover, a SimAM module is inserted after each residual block, resulting in a total of four SimAM modules in the trunk subnetwork. In addition, a Transformer attention

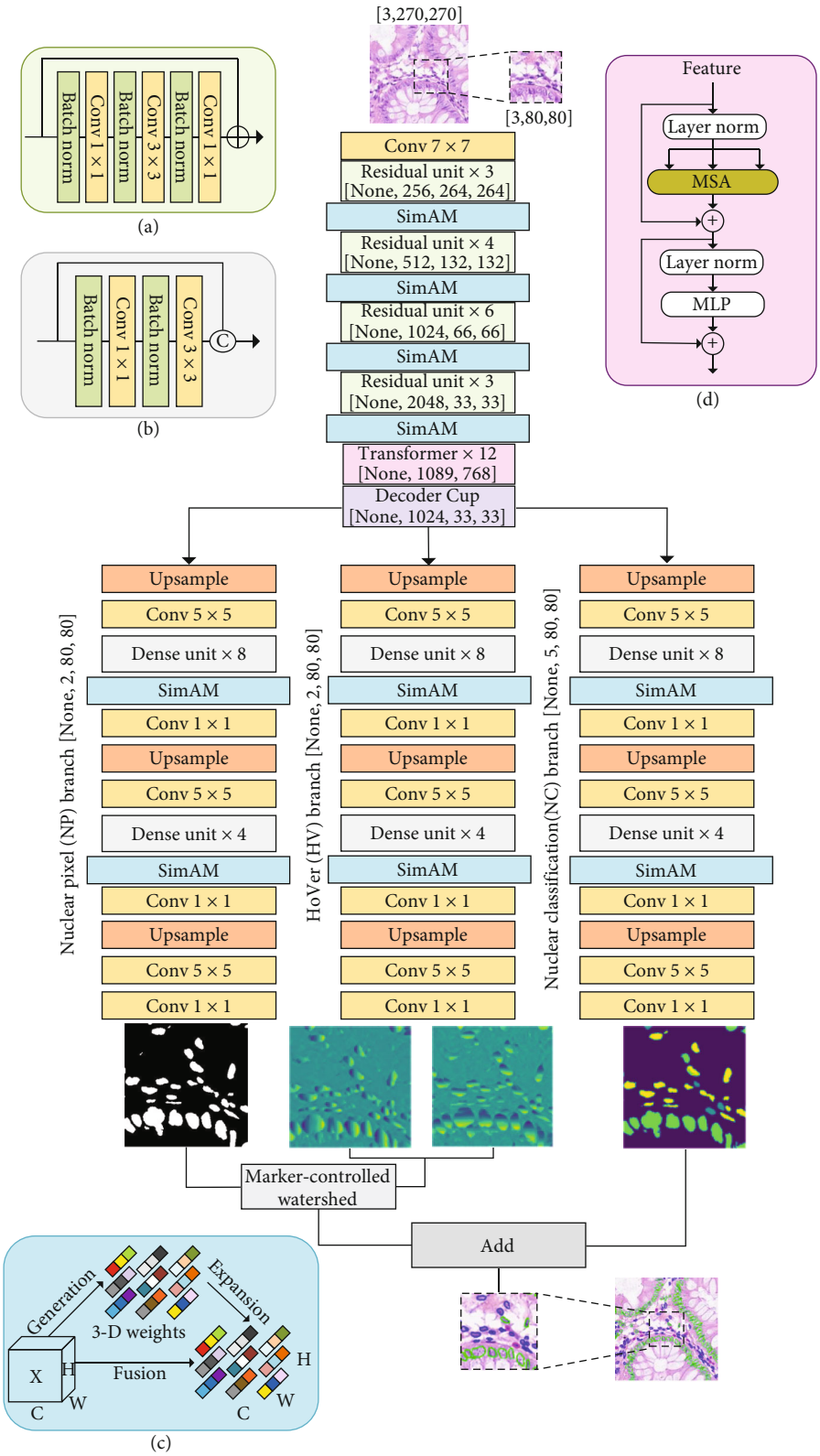


Figure 1: TSHVNet structure diagram. (a) A residual unit, (b) a dense unit, (c) a SimAM module, and (d) a Transformer attention module. The size of the original image is 270 × 270, and the center point is 80 × 80 for identification. The size of the output image is also 80 × 80. The output image and the center patch of the surrounding part are spliced to form a 270 × 270 image.

Table 1: Parameter number and the tensor dimension of each layer of TSHVNet.

Name	Parameters	Output shape
Residual block 1	145 K	[None, 256, 264, 264]
Residual block 2	938.5 K	[None, 512, 132, 132]
Residual block 3	4.86 M	[None, 1024, 66, 66]
Residual block 4	10.5 M	[None, 2048, 33, 33]
Transformer (×12)	85 M	[None, 1089, 768]
DecoderCup	7.08 M	[None, 1024, 33, 33]
Nuclear pixel (NP) branch		
Upsample, Conv(5 × 5)		[None, 256, 62, 62]
Dense unit (×8)		[None, 32, 30, 30]
SimAM, Conv(1 × 1), upsample, Conv(5 × 5)	9.72 M	[None, 128, 56, 56]
Dense unit (×4)		[None, 32, 40, 40]
SimAM, Conv(1 × 1), upsample, Conv(5 × 5), Conv(1 × 1)		[None, 2, 80, 80]
HoVer (HV) branch		
Upsample, Conv(5 × 5)		[None, 256, 62, 62]
Dense unit (×8)		[None, 32, 30, 30]
SimAM, Conv(1 × 1), upsample, Conv(5 × 5)	9.72 M	[None, 128, 56, 56]
Dense unit (×4)		[None, 32, 40, 40]
SimAM, Conv(1 × 1), upsample, Conv(5 × 5), Conv(1 × 1)		[None, 2, 80, 80]
Nuclear classification (NC) branch		
Upsample, Conv(5 × 5)		[None, 256, 62, 62]
Dense unit (×8)		[None, 32, 30, 30]
SimAM, Conv(1 × 1), upsample, conv (5 × 5)	9.72 M	[None, 128, 56, 56]
Dense unit (×4)		[None, 32, 40, 40]
SimAM, Conv(1 × 1), upsample, Conv (5 × 5), Conv(1 × 1)		[None, 5, 80, 80]

module is attached to the last SimAM module to implement global dependencies from image input to network outputs.

3.2.2. Branch Subnetworks. After feature extraction by the trunk encoder network, the obtained feature maps are processed, respectively, by three branches of the same structure but with different objectives. The three branches consist of a series of nearest neighbor upsampling and dense connection units [34] with skip connection to incorporate the features of the encoder. The three branches share one encoder, which can take advantage of the information sharing among multiple tasks, and improve the performance of the model in multitask.

Specifically, the first and second branches jointly achieve nuclear segmentation through a Marker-controlled thresholding watershed algorithm. The NC branch predicts the type of each nucleus via a majority vote on the predicted pixel-level classification results within the nucleus.

3.2.3. Transformer Module. The Transformer [13] performs multihead self-attention to extract nuclear contexts and relationships within an image and handle the long-range dependences among diverse nuclear instances, but the Transformer is incapable of catching nuclear position information. On the other hand, as a CNN-based model, HoVer-Net holds rich position information but lacks the ability of long-distance modeling. Thus, to take the advantage of both CNN and Transformer models, we integrate the Transformer attention mechanism into the HoVer-Net in the proposed TSHVNet.

To implement the transformer attention module in our model, we first change the input image $X\{X \in R^{H\times W\times C}\}$ into a flat 2D patch $\{X_p^i \in R^{P\times P\times C}\}$. The size of each patch is $P \times P$, and the number of patches is $N = HW/P^2$. And then, we use a trainable linear projection to map the vectorized patch X_P to the potential D-dimensional embedding space. In addition, to encode the spatial information of the patch, the specific position embedding matrix that will be added to the patch to maintain the position information can be calculated by the following equation:

$$Z_0 = \left[x_p^1 E; x_p^2 E; .. \cdots ; X_p^N E\right] + E_{\text{pos}}, \tag{1}$$

where $E \in R^{(P\times P\times C)\times D}$ is the embedding projection of the patch and $E_{\text{pos}} \in R^{N\times D}$ represents the position of this embedding.

Finally, Z_0 is the input to the Transformer encoder, which is composed of an L-layer multihead self-attention

(MSA) and a multilayer perceptron (MLP) module. The output of each layer can be described as follows:

$$\begin{aligned} Z'_{\sigma} &= \mathrm{MSA}(\mathrm{LN}(Z_{\sigma-1})) + Z_{\sigma-1}, \\ Z_{\sigma} &= \mathrm{MLP}\left(\mathrm{LN}\left(Z'_{\sigma}\right)\right) + Z'_{\sigma}, \end{aligned} \tag{2}$$

where LN represents the normalization operation and Z_σ is the encoded image representation. This encoder structure is represented by (d) in Figure 1.

Later, for maintaining the consistency of the image, we introduce an upsampling module to map the 2D sequence from Transformer back to the 3D space.

3.2.4. SimAM Module. To assign appropriate weights to neurons, we employed several SimAM modules that integrate 3D channel attention and spatial attention [14] in both the trunk and branches of the proposed TSHVNet model, as shown in Figure 1. In neuroscience, a neuron with more information exhibits distinctive firing patterns compared with its neighbors. So, SimAM came up with a simple energy function to give higher weight to these important neurons. This module can assign different weights to each neuron to improve the prediction of ROI, as shown in Figure 1(c).

3.3. Loss Function. In our proposed network, we use the original loss function defined in the HoVer-Net which employs three different loss functions at each branch output for better overall performance. The loss function can be formulated as

$$L = \underbrace{\lambda_a l_a + \lambda_b l_b}_{\mathrm{HV}} + \underbrace{\lambda_c l_c + \lambda_d l_d}_{\mathrm{NP}} + \underbrace{\lambda_e l_e + \lambda_f l_f}_{\mathrm{NC}}, \tag{3}$$

where l_a and l_b are the regression losses of HV branch output, l_c and l_d are the losses of NP branch output, and l_e and l_f are the losses of NC branch output. $\lambda_a, \cdots, \lambda_f$ are the weight scalars for each of the associated loss functions, specifically, making λ_b equal to 2 and the others equal to 1.

3.4. Postprocessing. Since there are significant differences in pixels between individual instances in the output of the HV branch, calculating gradients horizontally and vertically can tell where the difference in pixel intensity between different nuclei is greatest and from where the nuclei should be separated. The process can be described as

$$S_m = \max\left(H_x(p_x), H_y\left(p_y\right)\right), \tag{4}$$

where p_x and p_y refer to the horizontal and vertical predictions of the output of the HV branch; H_x and H_y denote the horizontal and vertical spatial gradient operations of the Sobel operator, respectively.

Moreover, a labeled watershed algorithm is further employed to segment different nuclei, under the condition of the given energy landscape E:

$$E = [1 - \tau(S_m, k)] * \tau(q, h), \tag{5}$$

with a controlling marker M:

$$M = \sigma(\tau(q, h) - \tau(S_m, k)), \tag{6}$$

where $\tau(a, b)$ is a threshold function that operates on a matrix a and will output "1" for the elements greater than b and output "0" otherwise. Specifically, h and k are set empirically, so that they give the optimal nuclear segmentation results. σ is a rectifier that sets all negative values to 0, and q is the output of the probability map in the NP branch.

Finally, to determine the class of a nuclear instance, a majority vote on the classes of the nucleus pixels is conducted in the output of the NC branch. Specifically, the type of a nuclear instance is assigned to the class with the largest number of pixels within the nucleus.

4. Experiment

4.1. Datasets

4.1.1. CoNSeP Dataset. The public CoNSeP dataset [35] is one of our experimental datasets. The dataset consists of 41 H&E stained images, each with pixels of 1000×1000. These 41 images are tagged with six different cell types: normal epithelial, tumor epithelial, inflammatory, necrotic, muscle, and fibroblasts. In our experiments, we group normal and malignant dysplastic epithelial nuclei into one class; fibroblast, muscle, and endothelial air conditioners are combined into a class of spindle-shaped nuclei. Therefore, in the end, there are four label types for the 41 images: inflammatory, epithelial, spindle-shaped, and other. Since the size of these 41 images is too large, we first cut these images into 256×256 images, resulting in 4563, 2366, and 2366 images in our training set, validation set, and test set, respectively.

4.1.2. PanNuke Dataset. At the same time, we also adopt the public PanNuke dataset [36] as our experimental dataset. The PanNuke dataset contains 7,901 images from 19 different organizations, and each image is 256×256 in size. The nuclei are divided into six categories: neoplastic cells, inflammatory cells, connective cells, dead cells, epithelial cells, and others. Since the public PanNuke dataset has not annotated the centroid of each nucleus, we extract the centroid of each nucleus in advance to make the format of the dataset conform to the requirements of the network structure. Finally, we obtain 2656, 2523, and 2722 images on the training set, validation set, and test set, respectively.

4.2. The Evaluation Index

4.2.1. Evaluation Index of Nuclear Segmentation. In the nuclear segmentation task of medical histopathology, the following two indicators are usually used to evaluate the results of nuclear segmentation: (1) DICE [37] is used to calculate the similarity between the real sample and the predicted sample. (2) Aggregated Jaccard Index (AJI) [38] computes the ratio of an aggregated intersection cardinality

to an aggregated union cardinality between ground truth and prediction. DICE and AJI are calculated as below:

$$\begin{aligned} \text{DICE} &= \frac{2 \times (X \cap Y)}{(|X| + |Y|)}, \\ \text{AJI} &= \frac{\sum_{i=1}^{N} \left| G_i \cap P_M^i \right|}{\sum_{i=1}^{N} \left| G_i \cup P_M^i \right| + \sum_{F \in U} |P_F|}, \end{aligned} \tag{7}$$

where X represents ground truth and Y represents the predicted nucleus. N represents the number of real nuclei, P_M^i refers to the connected region that generates the nuclear mask, G_i refers to the connected domain that P_M^i has the largest intersection with the ground truth, U refers to the set of connected regions that have no intersection with the ground truth, and P_F denotes the element F inside the set U.

The previous two evaluation indexes assess only the performance of instance segmentation but not the quality of predicted results associated with the true segmentation. To this end, we use the evaluation index PQ originally proposed by Kirillov et al. [39] in our experiment, and the index PQ is computed as below:

$$\text{PQ} = \frac{|\text{TP}|}{|\text{TP}| + (1/2)|\text{TP}| + (1/2)|\text{FN}|} \times \frac{\sum_{(x,y) \in TP} \text{IOU}(x, y)}{|\text{TP}|}, \tag{8}$$

where TP, FP, and FN denote true positive, false positive, and false negative, respectively. Besides, x represents ground truth, and y denotes the predicted segmentation region. IoU is the ratio of the intersection to the union of x and y. The left-hand side of the multiplier indicates the quality of detection. The right-hand side can be interpreted as the distance between each correctly detected instance and the ground truth they match. In this way, we can evaluate the quality of the nuclear instance segmentation and classification objectively and comprehensively.

4.2.2. Evaluation Index of Nuclear Classification. To assess the classification performance of nuclei, we adopt the evaluation index $F1$_score which is the result of a harmonic average between the precision and recall scores. $F1$_score can comprehensively evaluate the quality of nuclear classification and is calculated as follows:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F1_\text{score} &= \frac{2^* \text{precision}^* \text{recall}}{\text{precison} + \text{recall}}, \end{aligned} \tag{9}$$

where TP, FP, and FN are shorted for true positive, false positive, and false negative, respectively. Particularly, Fd is used to indicate the overall $F1$_score, Fi is used to indicate the $F1$_score for inflammatory nuclei, Fc is used to indicate the $F1$_score for connective cells, Fdead denotes the $F1$_score for dead cells, Fe represents the $F1$_score for epithelial, Fn represents the $F1$_score for neoplastic cells, FS represents the $F1$_score for spindle, and Fm indicates the $F1$_score for miscellaneous.

4.3. Implementation Details

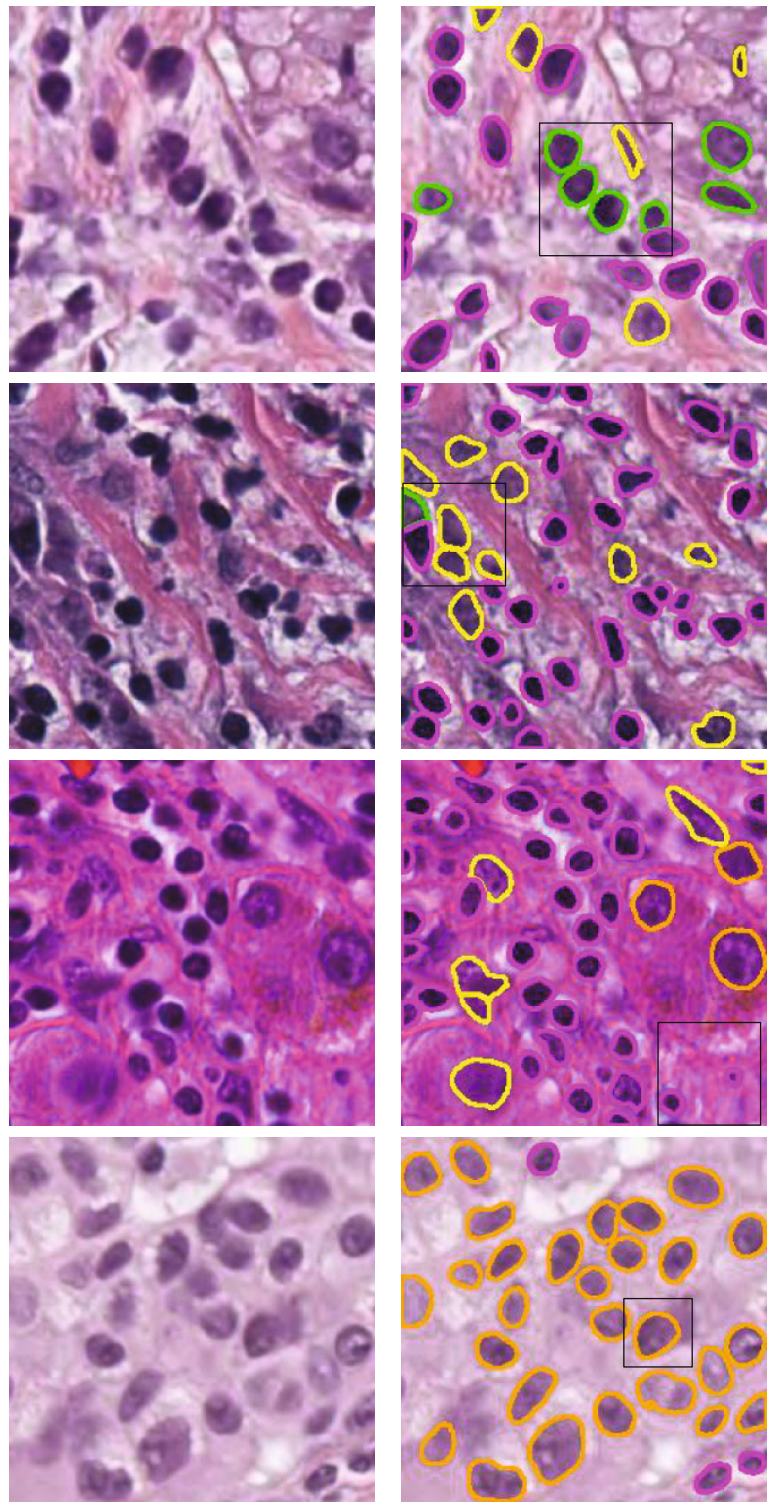
4.3.1. Experimental Environment. We carried out the experiments in this paper by using two NVIDIA 1080Ti graphics processor (GPU) accelerators on the Ubuntu 16.04 operating system and PyTorch 1.0.1, a deep learning framework.

4.3.2. Training Details. In our experiment, we use flip, rotation, Gaussian blur, and median blur data augmentation methods. The network input size of the CoNSeP dataset is 270×270, and that of the PanNuke dataset is 256×256. For the Transformer attention module, we set the size of the patch to 16 and the layer number of MSA and MLP to 12. The TSHVNet model was initialized by the parameters pretrained on the ImageNet dataset.

The TSHVNet was trained using the following two-stage training strategy where the epoch numbers were set empirically for the best accuracy. Firstly, we trained only the decoder at the first 50 epochs. Secondly, during the next 100 epochs, we fine-tuned the parameters of the entire network. In addition, batch sizes of 8 and 4 are used in the first stage and the second stage, respectively. Adam optimizer was used, and the initial learning rate was set at 10^{-4} and then reduced to 10^{-5} after 25 epochs.

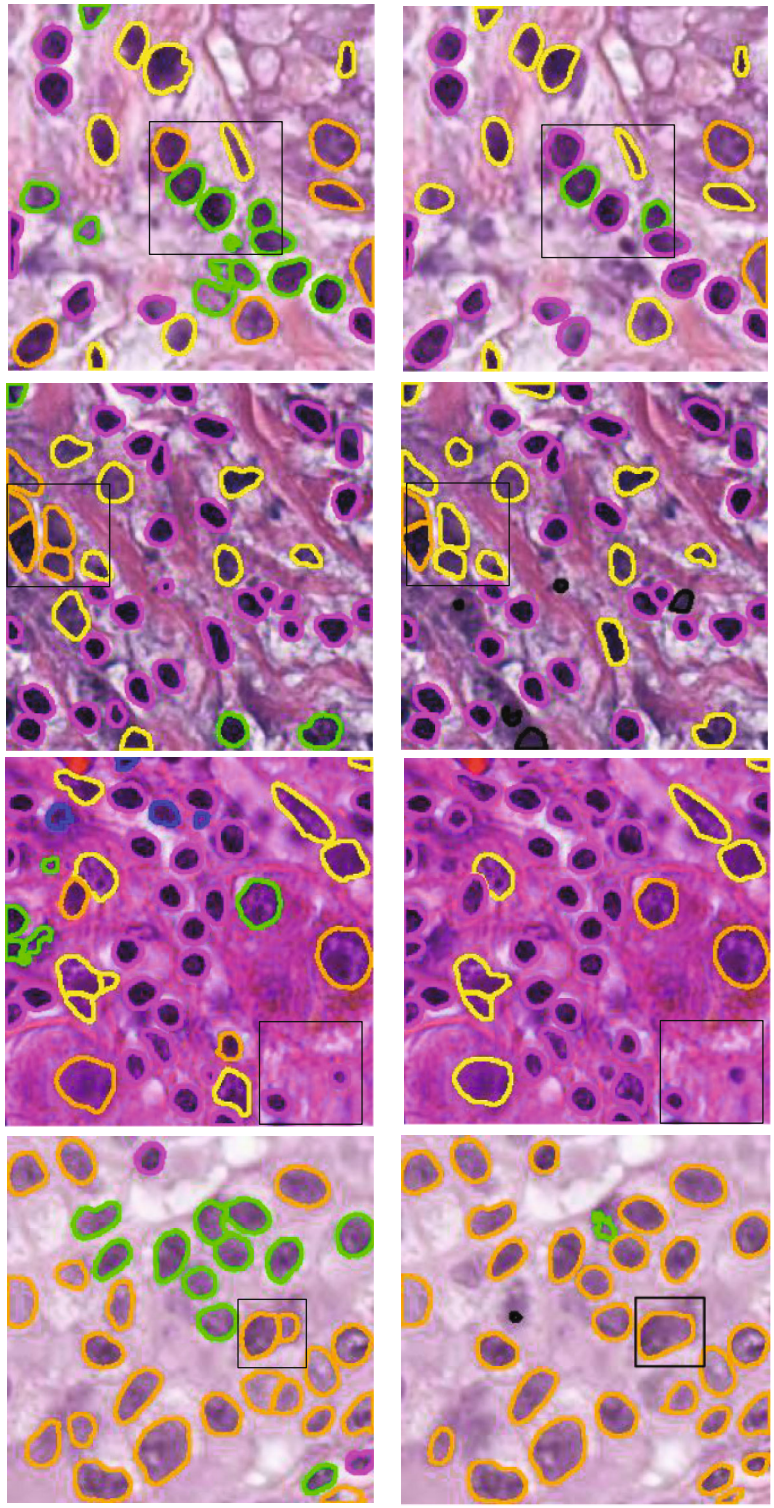
4.4. Demonstration of Visual Results. In this part, we visualize the results of simultaneous nuclear instance segmentation and classification obtained by different methods in different datasets, as shown in Figures 2 and 3. Figure 2 illustrates the visualization results on the PanNuke dataset, and Figure 3 demonstrates the results on the CoNSeP dataset. For both Figures 2 and 3, the first column is the original image, and the second column is the ground truth, where different colors represent different categories. The third column represents the predicted results of the Micro-Net structure. The fourth column represents the predicted results of the Dist structure. The fifth column represents the HoVer-Net prediction result. The sixth represents the prediction result of our algorithm TSHVNet. Red is miscell, dark red is inflamed, green is epithe, blue is spin, yellow is Connec, cyan is dead, and orange is neoplastic.

As can be seen from Figure 2, the proposed TSHVNet model has better performance than those of the current state-of-art networks on both classification and segmentation on the PanNuke dataset. According to the parts marked by black boxes in the first two rows of Figure 2, the proposed TSHVNet model can correctly determine the types of the nucleus. As can be seen from the part marked by black boxes in the last two rows of Figure 2, the proposed TSHVNet can more accurately locate and segment the nuclei, as well as better classify the nuclear types of the nuclei. Moreover, in Figure 3, the proposed TSHVNet model also exhibits better performance than that of the HoVer-Net network in classification and segmentation on CoNSeP datasets. As compared



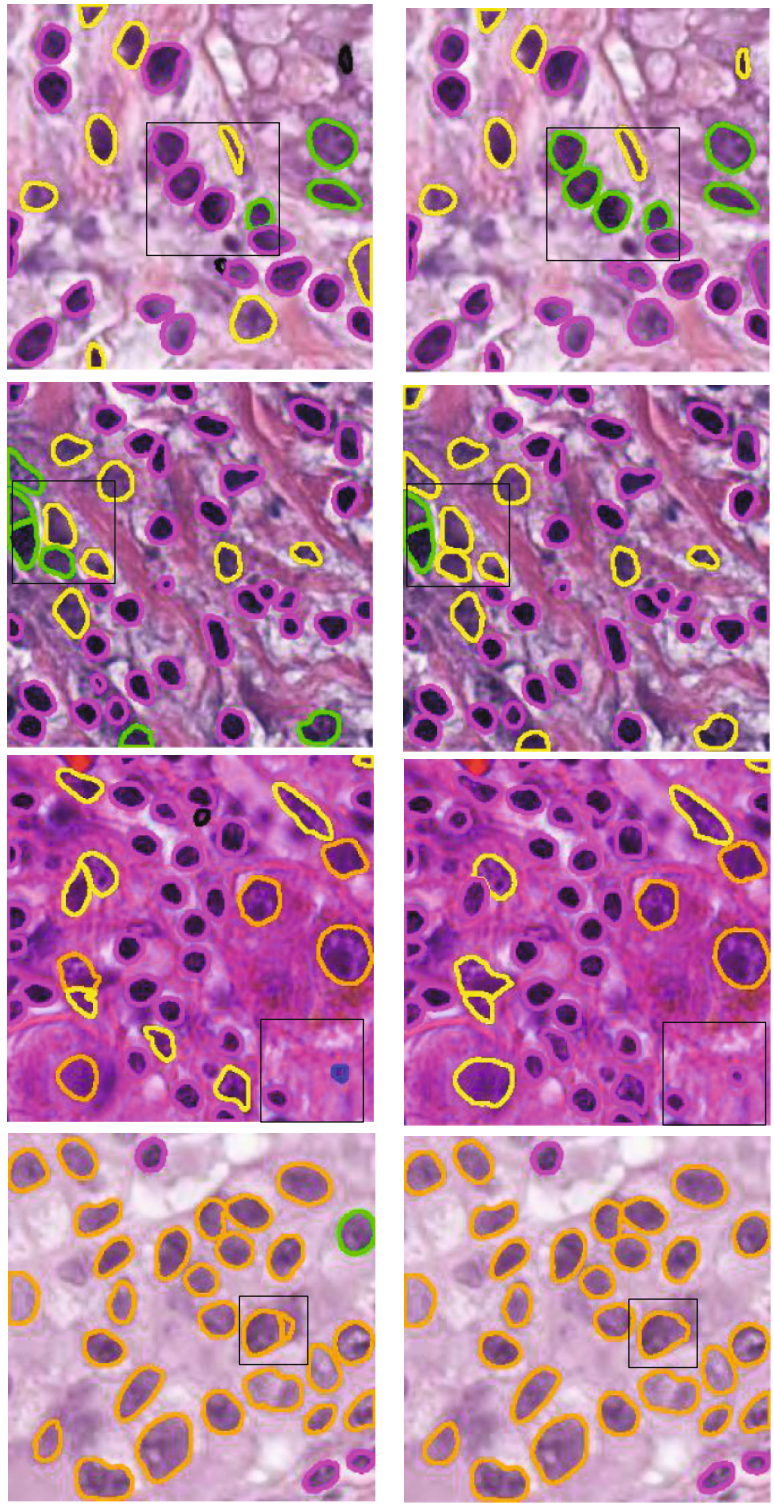
(a) Original image (b) GT

Figure 2: Continued.



(c) Micro-Net (d) Dist

Figure 2: Continued.

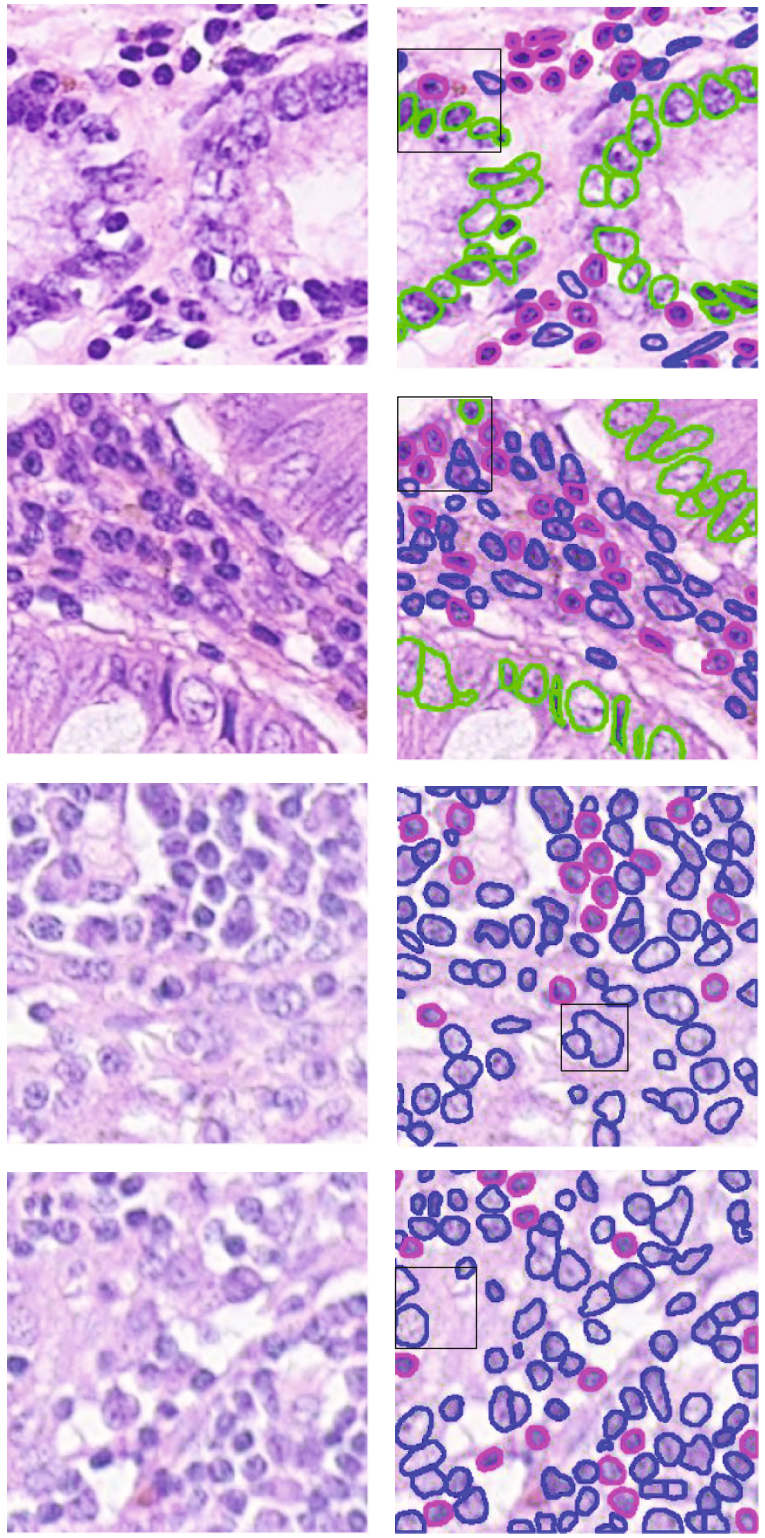


(e) HoVer-Net (f) TSHVNet

Figure 2: Visual demonstration of the comparative performance of different models on the PanNuke dataset.

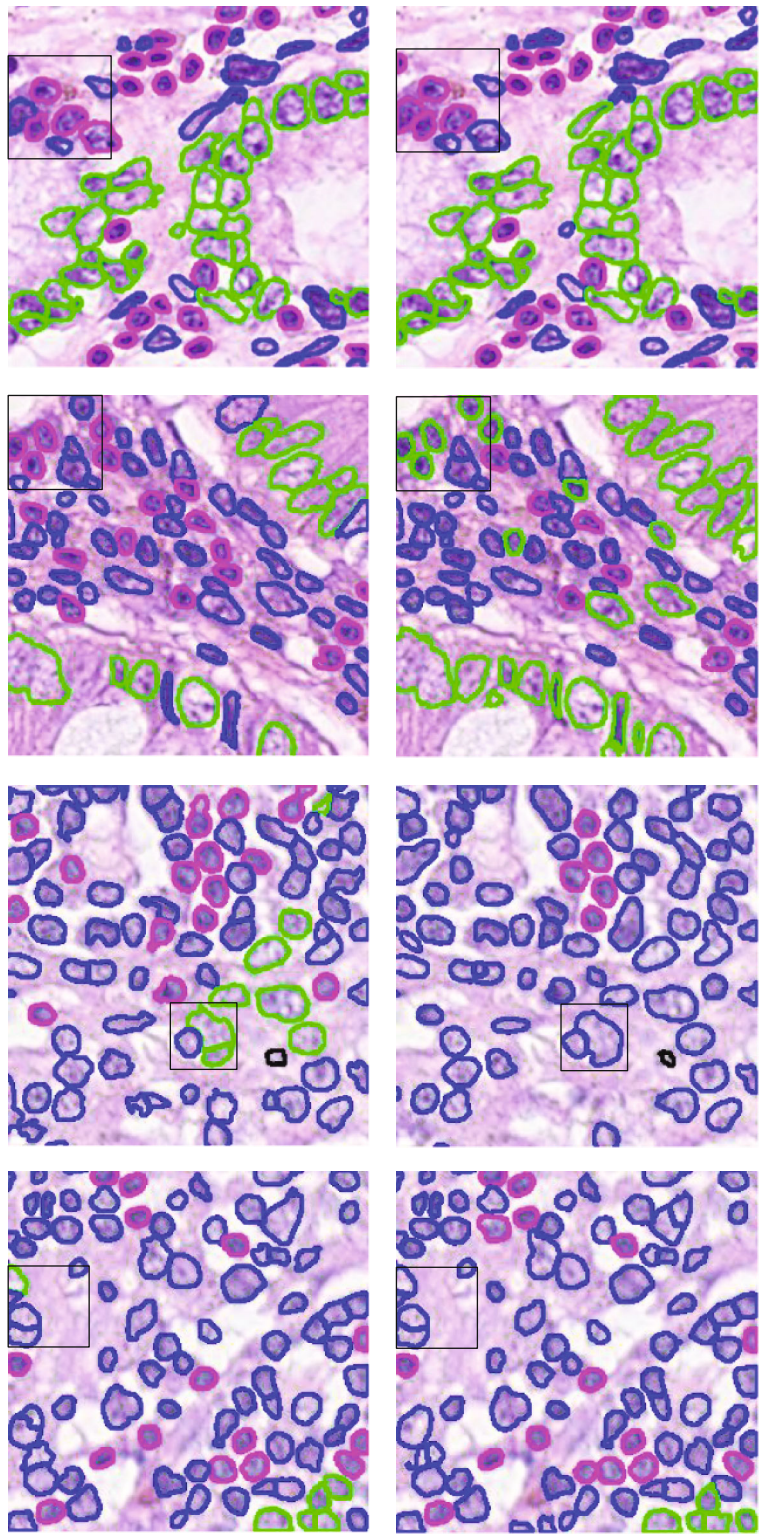
to the instance segmentation networks Micro-Net and Dist, the superiority of the proposed TSHVNet can also be observed. For instance, the black boxes in the first two rows of Figure 3 show that the nuclear instance segmentation and classification results performed by the TSHVNet are more consistent with the ground truth images.

The superior performance of the proposed TSHVNet model can be attributed to the integration of the Transformer



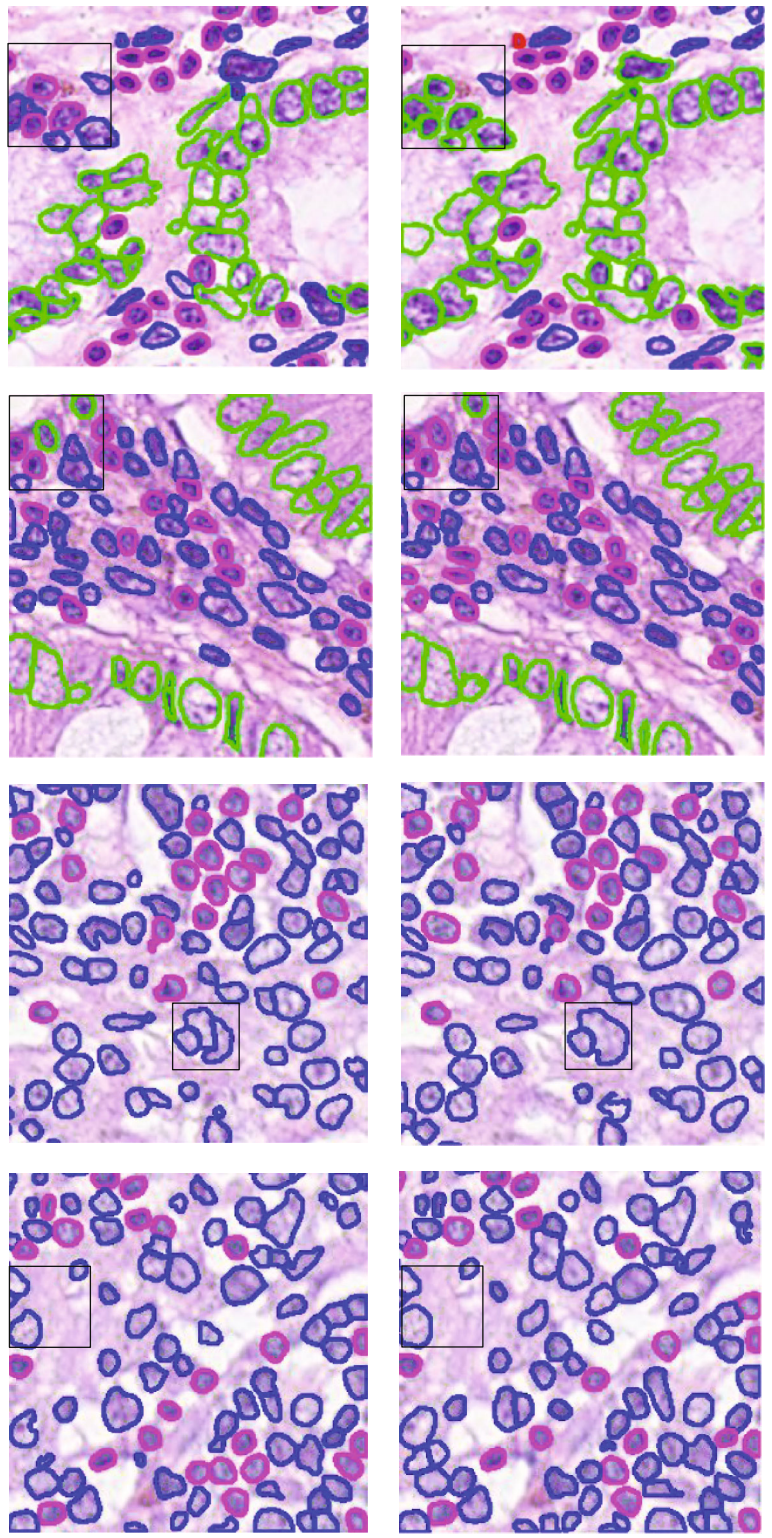
(a) Original image (b) GT

Figure 3: Continued.



(c) Micro-Net (d) Dist

Figure 3: Continued.



(e) HoVer-Net (f) TSHVNet

Figure 3: Visual demonstration of the comparative performance of different models on the CoNSeP dataset.

module into the traditional CNN, which enables the network to achieve long-distance modeling and retrieve precise location information. In addition, the SimAM modules with 3D channel attention and space attention inserted into the trunk and branches of the proposed TSHVNet model can also help to better identify the ROI of nuclear instances.

Table 2: Quantitative comparison of instance segmentation results performed by different models on the CoNSeP and PanNuke datasets.

Model	Dataset metrics									
	CoNSeP					PanNuke				
	DICE	AJI	DQ	SQ	PQ	DICE	AJI	DQ	SQ	PQ
Dist	0.794	0.505	0.544	0.725	0.397	0.782	0.598	0.636	0.764	0.499
Micro-Net	0.780	0.513	0.560	0.709	0.415	0.810	0.654	0.740	0.794	0.599
HoVer-Net	0.849	0.556	0.687	0.772	0.532	0.818	0.651	0.757	0.783	0.609
TSHVNet	0.856	0.558	0.690	0.777	0.546	0.835	0.662	0.779	0.813	0.637

Table 3: Quantitative comparison of nuclear classification results performed by different models on the CoNSeP and PanNuke datasets.

Model	Dataset metrics									
	CoNSeP					PanNuke				
	*F*d	*F*e	*F*i	*F*s	*F*m	*F*d	*F*i	*F*c	*F*dead	*F*e	*F*n
Dist	0.732	0.626	0.554	0.509	0.025	0.741	0.402	0.391	0.000	0.137	0.518
Micro-Net	0.721	0.601	0.550	0.495	0.105	0.787	0.464	0.401	0.104	0.448	0.580
HoVer-Net	0.738	0.618	0.564	0.532	0.348	0.790	0.465	0.413	0.153	0.463	0.591
TSHVNet	0.763	0.669	0.621	0.583	0.358	0.820	0.531	0.460	0.179	0.536	0.623

Table 4: Evaluation results of nuclear segmentation on different combinations of SimAM modules.

Method	Dataset metrics									
	CoNSeP					PanNuke				
	DICE	AJI	DQ	SQ	PQ	DICE	AJI	DQ	SQ	PQ
HoVer-Net	0.849	0.556	0.687	0.772	0.532	0.818	0.651	0.757	0.783	0.609
HoVer-Net+Res	0.849	0.558	0.685	0.774	0.531	0.823	0.650	0.765	0.813	0.631
HoVer-Net+Des	0.848	0.556	0.689	0.773	0.532	0.823	0.652	0.767	0.812	0.662
HoVer-Net+Res+Des	0.849	0.557	0.690	0.772	0.533	0.824	0.653	0.768	0.811	0.633
TSHVNet (Res)	0.847	0.554	0.680	0.776	0.528	0.825	0.660	0.776	0.813	0.633
TSHVNet (Des)	0.848	0.556	0.683	0.776	0.530	0.830	0.661	0.777	0.810	0.629
TSHVNet	0.856	0.558	0.690	0.777	0.546	0.835	0.662	0.779	0.813	0.637

Table 5: $F1$ _scores of various nuclear categories on different combinations of SimAM modules.

Method	Dataset metrics										
	CoNSeP					PanNuke					
	*F*d	*F*e	*F*i	*F*s	*F*m	*F*d	*F*i	*F*c	*F*d	*F*e	*F*n
HoVer-Net	0.738	0.618	0.564	0.532	0.348	0.790	0.465	0.413	0.153	0.463	0.591
HoVer-Net+Res	0.761	0.655	0.622	0.579	0.393	0.809	0.495	0.440	0.169	0.470	0.599
HoVer-Net+Des	0.760	0.666	0.615	0.570	0.360	0.806	0.480	0.445	0.168	0.471	0.563
HoVer-Net+Res+Des	0.763	0.659	0.601	0.562	0.356	0.809	0.502	0.445	0.168	0.470	0.591
TSHVNet (Res)	0.757	0.647	0.598	0.553	0.362	0.815	0.528	0.455	0.176	0.530	0.620
TSHVNet (Des)	0.759	0.650	0.600	0.560	0.360	0.816	0.527	0.454	0.177	0.532	0.621
TSHVNet	0.763	0.669	0.621	0.583	0.358	0.820	0.531	0.460	0.179	0.536	0.623

4.5. Experiment and Analysis. To objectively measure the ability of simultaneous nuclear segmentation and classification of our proposed method, we conducted different experiments on the two public datasets, and the results are shown in Tables 2 and 3. Table 2 shows the segmentation quantization indexes of the two datasets in different methods, and Table 3 shows the classification quantization indexes of the two datasets in different methods.

From Table 2, we can see that the proposed TSHVNet outperforms all the listed current state-of-the-art nuclear instance segmentation methods in all the evaluation indicators. In terms of nuclear instance segmentation, the metrics values of DICE, AJI, DQ, SQ, and PQ of TSHVNet are 0.7%, 0.2%, 0.3%, 0.5%, and 1.4% higher than those of HoVer-Net on the CoNSeP dataset, respectively. From the comparison of TSHVNet and HoVer-Net on the PanNuke

dataset, increments of 1.7%, 1.1%, 2.2%, 3.0%, and 2.8% on DICE, AJI, DQ, SQ, and PQ can be observed, respectively.

Concerning nuclear classification, it can be seen from Table 3 that on the CoNSeP dataset, the evaluation scores of *F*d, *F*e, *F*i, *F*s, and *F*m for the proposed TSHVNet are 2.5%, 4.1%, 5.7%, 3.1%, and 1.0% higher than those of the HoVer-Net, respectively. The increments have become 3.0%, 6.7%, 4.7%, 2.6%, 7.3%, and 3.2% on the PanNuke dataset, respectively.

The improvement could be attributed to the global dependencies modeling by the Transformer attention module and the simultaneous spatial and channel attention brought by the SimAM modules, which makes the network pay more attention to the nuclear regions of interest and focus more on the distinctive features to classify the nuclei in the histopathology images.

4.6. Ablation Experiments of SimAM Modules. In this part, we perform a series of ablation experiments to verify the effect of SimAM at different locations on the entire network. Because there are four residual units in our feature extraction, we abbreviate SimAM behind the four residual units as "Res." The three branches contain two dense units, and we also indicate the SimAM module behind the two dense units as "Des" for short. The experimental results are shown in Tables 4 and 5.

Table 4 shows the comparison of nuclear segmentation performances by the HoVer-Net and the TSHVNet when TSHVNet is combined with SimAM modules at different positions of the network. Table 5 shows the comparison results of nuclear classification by the HoVer-Net and the TSHVNet when the TSHVNet is equipped with the different combinations of SimAM modules. The "(Res)" and "(Des)" in both Tables 4 and 5 represent the individual combination of the SimAM attention modules behind the residual units in the trunk and the dense units in the three branches of TSHVNet, respectively.

It can be inferred from Tables 4 and 5 that the SimAM attention modules added behind the residual units and dense units in the network can effectively improve the segmentation and classification ability of the network.

For nuclear segmentation, we combined all the SimAM modules into the TSHVNet model other than using the "Res" SimAM modules or the "Des" SimAM modules alone, which would gain 1.8% and 1.6% promotions in the PQ score on the CoNSeP dataset and 0.4% and 0.8% increments on the PanNuke dataset, respectively.

Regarding the nuclear classification, as compared to the TSHVNet model with only the "Res" SimAM modules or the "Des" SimAM modules, the proposed TSHVNet model integrating all the SimAM modules simultaneously gains 0.6% and 0.4% increments in the *F*d score (i.e., the overall *F*1_score) on the CoNSeP dataset and 0.5% and 0.4% promotions on the PanNuke dataset, respectively. Therefore, the SimAM modules, which integrates 3D channel attention and spatial attention mechanism according to the energy function of neuroscience, are able to perceive contextual spatial information between diverse nuclear instances and show competitive instance segmentation performance.

5. Conclusion

In this paper, we proposed a deep network named TSHVNet to simultaneously conduct nuclear segmentation and nuclear classification on medical histopathological images, which integrated Transformer and SimAM modules into the HoVer-Net. The Transformer attention module could automatically learn the global and local dependencies between the image input and network outputs but is insensitive to the location information. We made up for the inaccurate positioning of the Transformer by using a hybrid model design integrating the characteristics of CNN and SimAM attention modules, which enabled the network to pay more attention to the contextual spatial information among diverse nuclear instances for a more accurate simultaneous nuclear segmentation and classification. In addition, we also discussed the effect of the coexistence channel and spatial attention advantages of the SimAM modules in our entire network. The SimAM modules help assign appropriate weights to neurons for distinguishing the nuclear instances with similar interclass but discrepant intraclass appearances. Experiments on the public CoNSeP and PanNuke datasets have proved the superior performance of the proposed TSHVNet model in both tasks of nuclear segmentation and nuclear classification as compared to the state-of-art competing methods.

Data Availability

Datasets in this paper are publicly available in the following: CoNSeP dataset link: https://warwick.ac.uk/fac/cross_fac/tia/data/hovernet/ and PanNuke dataset link: https://warwick.ac.uk/fac/cross_fac/tia/data/pannuk.

Conflicts of Interest

The authors declare that no conflict of interest exists in the submission of this manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61501287, No. 61971273, No. U2001205, and No. 61877038), the Key Research and Development Program in Shaanxi Province of China (No. 2021GY-032), the State Key Laboratory of Geo-Information Engineering (SKLGIE2019-M-3-5), the Fundamental Research Funds for the Central Universities (No. GK202003077), and the Research and Development Project funded by Central University/Public Health and Health Research Project (GK202007033).

References

[1] K. Sirinukunwattana, D. Snead, D. Epstein et al., "Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer," *Scientific Reports*, vol. 8, no. 1, pp. 13690–13692, 2018.

[2] J. G. Elmore, G. M. Longton, P. A. Carney et al., "Diagnostic concordance among pathologists interpreting breast

biopsy specimens," *JAMA*, vol. 313, no. 11, pp. 1122–1132, 2015.

[3] X. Liu, P. Yuan, R. Li et al., "Predicting breast cancer recurrence and metastasis risk by integrating color and texture features of histopathological images and machine learning technologies," *Computers in Biology and Medicine*, vol. 146, article 105569, 2022.

[4] S. Yang, J. Xiang, and X. Wang, "A deep learning framework for nuclear segmentation and classification in histopathological images," 2022, https://arxiv.org/abs/2203.03420.

[5] L. Balkenende, J. Teuwen, and R. M. Mann, "Application of deep learning in breast cancer imaging," *Seminars in Nuclear Medicine*, vol. 52, no. 5, pp. 584–596, 2022.

[6] A. Ravikumar, H. Sriraman, P. M. Saketh, S. Lokesh, and A. Karanam, "Effect of neural network structure in accelerating performance and accuracy of a convolutional neural network with GPU/TPU for image analytics," *PeerJ Computer Science*, vol. 8, article e909, 2022.

[7] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Springer International Publishing, 2015.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Boston, MA, USA, 2015.

[9] D. Jha, P. H. Smedsrud, M. A. Riegler et al., "ResUNet++: an advanced architecture for medical image segmentation," in *2019 IEEE International Symposium on Multimedia (ISM)*, pp. 225–2255, San Diego, CA, USA, 2019.

[10] H. Sharma, N. Zerbe, D. Heim et al., "A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images," in *International Conference on Computer Vision Theory & Applications,*, pp. 37–46, Berlin, Germany, 2015.

[11] J. K. Shuttleworth, A. G. Todman, R. N. Naguib, B. M. Newman, and M. K. Bennett, "Multiresolution colour texture analysis for classifying colon cancer images," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*, pp. 1118-1119, Houston, TX, USA, 2002.

[12] S. Graham, Q. D. Vu, S. E. Raza et al., "Hover-Net: simultaneous segmentation and classification of nuclei in multi- tissue histology images," *Medical Image Analysis*, vol. 58, article 101563, 2019.

[13] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[14] L. Yang, R. Y. Zhang, L. Li, and X. Xie, "SimAM: a simple, parameter-free attention module for convolutional neural networks," in *International Conference on Machine Learning. PMLR*, pp. 11863–11874, 2021.

[15] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 13, no. 6, pp. 583–598, 1991.

[16] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 11, pp. 2405–2414, 2006.

[17] M. Skobel, M. Kowal, J. Korbicz, and A. Obuchowicz, "Cell nuclei segmentation using marker-controlled watershed and Bayesian object recognition," in *International Conference on Information Technologies in Biomedicine*, pp. 407–418, Springer, Cham, 2019.

[18] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[19] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted Res-UNet for high-quality retina vessel segmentation," in *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 327–331, Hangzhou, China, 2018.

[20] S. E. Raza, L. Cheung, M. Shaban et al., "Micro-Net: a unified model for segmentation of various objects in microscopy images," *Medical Image Analysis*, vol. 52, pp. 160–173, 2018.

[21] H. He, C. Zhang, J. Chen et al., "A hybrid-attention nested UNet for nuclear segmentation in histopathological images," *Frontiers in Molecular Biosciences*, vol. 8, no. 6, pp. 3–5, 2017.

[22] T. N. Doan, B. Song, T. T. Vuong, K. Kim, and J. T. Kwak, "SONNET: a self-guided ordinal regression neural network for segmentation and classification of nuclei in large-scale multi-tissue histology images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 3218–3228, 2022.

[23] J. Liu, C. Desrosiers, and Y. Zhou, "Att-MoE: attention-based mixture of experts for nuclear and cytoplasmic segmentation," *Neurocomputing*, vol. 411, pp. 139–148, 2020.

[24] H. Chen, X. Qi, L. Yu, and P. A. Heng, "DCAN: deep contour-aware networks for accurate gland segmentation," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2487–2496, Las Vegas, NV, USA, 2016.

[25] J. Cheng, X. Pan, F. Hou et al., "A Standardized Pipeline for Colon Nuclei Identification and Counting Challenge," 2022, https://arxiv.org/abs/2203.00171.

[26] K. Nguyen, B. Sabata, and A. Jain, "Prostate cancer detection: fusion of cytological and textural features," *Journal of Pathology Informatics*, vol. 2, p. S3, 2011.

[27] K. Sirinukunwattana, S. E. Raza, Y. W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[28] S. Zhao, Y. He, J. Qin, and Z. Wang, "A semi-supervised deep learning method for cervical cell classification," *Analytical Cellular Pathology*, vol. 2022, Article ID 4376178, 12 pages, 2022.

[29] Y. Jia, C. Lu, X. Li et al., "Nuclei instance segmentation and classification in histopathological images using a DT-Yolact," in *2021 20th International Conference 2021*, pp. 414–420, London, United Kingdom, 2021.

[30] A. de Santana Correia and E. L. Colombini, "Attention, please! A survey of neural attention models in deep learning," *Artificial Intelligence Review*, vol. 55, pp. 6037–6124, 2022.

[31] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 7132–7141, 2019.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on*

computer vision and pattern recognition, pp. 770–778, Las Vegas, NV, USA, 2016.

[34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, Honolulu, HI, USA, 2017.

[35] A. Okka, U. Dogrusoz, and H. Balci, "CoSEP: a compound spring embedder layout algorithm with support for ports," *Information Visualization*, vol. 20, no. 2-3, pp. 151–169, 2021.

[36] J. Gamper, N. A. Koohbanani, K. Benes et al., "PanNuke dataset extension, insights and baselines," 2020, https://arxiv.org/abs/2003.10778.

[37] Q. D. Vu, S. Graham, T. Kurc et al., "Methods for segmentation and classification of digital microscopy tissue images," *Frontiers in Bioengineering and Biotechnology*, vol. 53, 2019.

[38] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[39] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413, Long Beach, CA, 2019.

Hindawi
BioMed Research International
Volume 2022, Article ID 9646846, 10 pages
https://doi.org/10.1155/2022/9646846

Research Article

Prediction Model of Residual Neural Network for Pathological Confirmed Lymph Node Metastasis of Ovarian Cancer

Huanchun Yao[1] **and Xinglong Zhang**[2]

[1]*Department of Cancer, Shengjing Hospital of China Medical University, No. 36, Sanhao Street, Heping District, Shenyang, 110004 Liaoning Province, China*
[2]*Department of Hematology, The Fourth Affiliated Hospital of China Medical University, Shenyang, 110032 Liaoning Province, China*

Correspondence should be addressed to Xinglong Zhang; habali414@163.com

Received 10 August 2022; Revised 31 August 2022; Accepted 12 September 2022; Published 11 October 2022

Academic Editor: Chen Li

Copyright © 2022 Huanchun Yao and Xinglong Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. We want to develop a model for predicting lymph node status based on positron emission computed tomography (PET) images of untreated ovarian cancer patients. We use the feature map formed by wavelet transform and the parameters obtained by image segmentation to build the model. The model is expected to help clinicians and provide additional information about what to do with first-visit patients. *Materials and Methods*. Our study included 224 patients with ovarian cancer. We have chosen two main methods to extract information from images. On the one hand, we segmented the image to extract the parameters to evaluate the clustering effect. On the other hand, we used wavelet transform to extract the image's texture information to form the image's feature map. Based on the above two kinds of information, we used residual neural network and support vector machine for modeling. *Results*. We established a model to predict lymph node metastasis in patients with primary ovarian cancer using PET images. On the training set, our accuracy was 0.8854, AUC: 0.9472, CI: 0.9098-0.9752, sensitivity was 0.9865, and specificity was 0.7952. On the test set, our accuracy was 0.9104, AUC: 0.9259, CI: 0.8417-0.9889, sensitivity was 0.8125, and specificity was 1.0000. *Conclusions*. We used wavelet transform to process the preoperative medical images of ovarian cancer patients, and the residual neural network can effectively predict the lymph node metastasis of ovarian cancer patients, which is undoubted of great significance for patients' staging and treatment options.

1. Introduction

Ovarian cancer is one of the most deadly gynecologic malignancies, with the highest incidence in North America and Central and Eastern Europe. According to the recent epidemiological and related cohort studies, the incidence rate of ovarian cancer shows a gradually increasing trend [1]. However, the vast majority of serous carcinomas are not diagnosed until stage III (51%) or IV (29%) because of the lack of symptoms prior to entire abdominal metastases [2], which is very detrimental to the survival of patients. The global overall 5-year survival rate is as low as 30% [3]. Epithelial ovarian cancer is the most crucial pathological type, accounting for more than 90% of all ovarian cancer [4]. Its subtypes include serous, mucinous, endometrioid, and clear cell carcinoma. According to the current research point of view, ovarian cancer, especially epithelial type, often has a high degree of pathological and molecular heterogeneity in the middle and late stages of the tumor [5], which is also an important reason for the different responses to treatment schemes [6]. Although nearly 75% of ovarian cancer patients achieve pathological remission after primary tumor reduction surgery and chemotherapy, 40-60% of ovarian cancer patients will eventually relapse [7], which seriously affects the quality of life and life safety of patients. According to FIGO guidelines, lymph node metastasis directly affects the surgical pathological stage of the tumor [8]. On this basis, lymph node and other organ involvement affects the extent of surgical resection, especially in patients who wish to preserve fertility. After operation, tumor stage will also affect

the choice of chemotherapy regimen. Lymphatic invasion is known as a predictor of tumor invasion and affects the survival of patients with ovarian cancer [9]. The lymph node invasion status of patients with ovarian cancer determines the scope of lymph node dissection in tumor reduction surgery and may be related to surgical complications [10]. Researchers have made many efforts to integrate molecular features for accurate prognosis which helps classify patients into risk groups and may provide more personalized treatment. Information on predictors of lymphatic invasion status in patients with ovarian cancer is still lacking, and few molecular prognostic classifiers are available. A study has shown that PET/CT images can partly reflect preoperative lymph node metastasis in patients with ovarian cancer using traditional artificial imaging diagnostic methods. But this approach is not entirely satisfactory [11]. Lymph node metastasis is thought to be associated with the prognosis of ovarian cancer [12].

Traditionally, radiologists subjectively evaluate medical images based on their training and experience to provide an assessment of the diagnostic disease or clinical status. This method produces irresistible instability in image interpretation. Using more automated imaging analysis tools in research or clinical trials can reduce this instability, provide more objective clinical-related information, and seek more beneficial treatment solutions for patients. Radiomics has been introduced as an emerging tool for post-processing images in medical fields such as computed tomography (CT) or magnetic resonance imaging (MR) and developing new quantitative indicators to link qualitative and quantitative imaging data with clinical endpoints to form an interpretable mathematical model [13]. This method does not rely too much on the will of the physician's supervisor, and through computer program, reliability and repeatability are more guaranteed. Texture analysis is a common research method in medical imaging. It requires researchers to extract specific texture features and then explore the relationship between texture features and clinical features. This correlation tends to be relatively simple and restrictive. However, deep learning can form different features for different tasks more flexibly and can also better explain the complex relationship between image features and clinical features. Ovarian cancer is a disease with complex genetic changes, and the heterogeneity of this genome and tumor microenvironment is related to the platinum resistance of patients [14]. For ovarian cancer, a tumor with complex lesions, noninvasive medical image analysis can better distinguish the heterogeneity of related tumors and play a positive role in tumor patients [15].

In the field of medical image, deep learning is one of the most effective methods. Neural network is a kind of model that simulates human brain in order to realize machine learning technology like artificial intelligence. This technique simulates the structure of human neurons to deal with complex problems. In medical laboratory science, researchers use neural networks to count microorganisms [16]. This method can be applied in the environment without human body [17]. In radiology, researchers use deep learning to identify and segment images. A study on acute intracranial haemorrhage showed that using deep learning methods, researchers were able to classify acute intracranial haemorrhage to the level of manual classification [18]. In gastroenterology, researchers use endoscopy to distinguish benign and malignant tumors, which has great significance in the screening of digestive tract tumors [19]. In assisted reproduction, researchers use neural networks to identify sperm [20]. Deep learning has proved its potential in many medical fields.

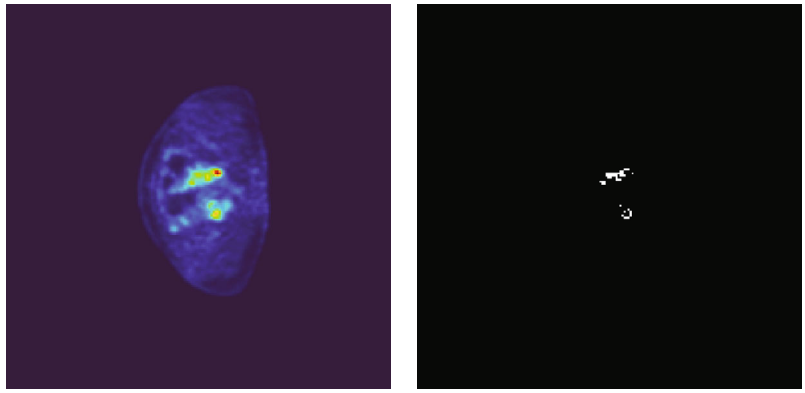
Researchers have widely used residual neural networks in various feature extraction applications. When the number of layers of deep learning networks is deeper, the target learning ability will theoretically be stronger. However, when the depth of the convolutional neural network reaches a certain level, the classification performance will not be improved, but the network will converge more slowly, and the accuracy rate will decrease. Even if we increase the sample size of the training set to solve the problem of overfitting, the classification performance and accuracy will not be improved. Residual neural network is a solution to this problem. In medicine, residual neural networks are widely used to deal with pathological and medical images.

2. Materials and Methods

2.1. Study Participants. We reviewed the radiology database of Shengjing Hospital of China Medical University. We identified 270 patients who underwent PET/CT from April 2013 to January 2019. 23 of these patients were excluded due to missing or poor image quality, and 23 were treated before PET or were diagnosed with nonprimary ovarian tumors on subsequent tests. Therefore, a total of 224 patients were included in this study. The inclusion criteria were that the patients did not have endocrine diseases or other primary tumors that could affect ovarian imaging, underwent PET/CT examination, surgery and pathological examination within two weeks at Shengjing Hospital of China Medical University, and were diagnosed with primary ovarian cancer. Previously, we had excluded patients with advanced ovarian cancer with severe invasion of other organs, large necrotic tumors, patients with severe ascites, and patients with the secondary ovarian cancer.

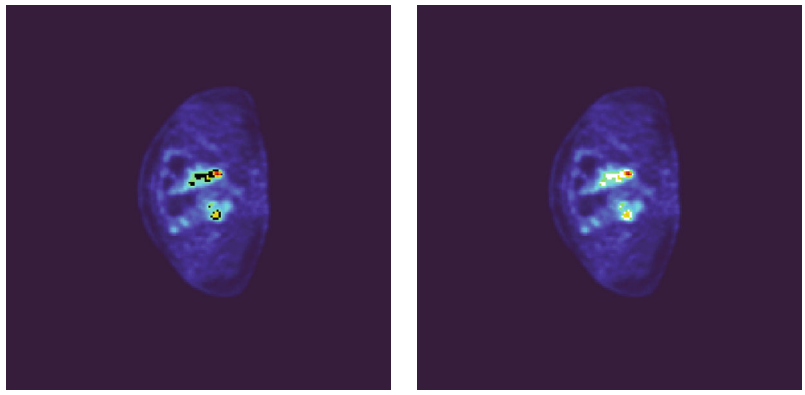
2.2. PET/CT. Patients fasted from food and water for more than 6 hours, and their blood glucose was controlled below 7 mmol/L. Patients accepted PET/CT 1 h after injection of 18F-FDG (GE MINItrace II; GE Healthcare, Milwaukee, WI) 0.08-0.16 MCI/kg. PET/CT was performed from the head to the middle of the femur (GE Discovery PET/CT Elite; GE Healthcare, Milwaukee, WI). A 3D PET model was used with a $192 * 192$ matrix and 2 min/bed position exposure time. Low-dose spiral CT was performed at 120~140 kV and 80 mA. After attenuation correction of CT, PET images were reconstructed using the algorithm of time-of-flight and point-spread-function, including two iterations and 24 subsets.

We used an advantage workstation 4.6 with PET volume computation-assisted reading software (PET VCAR; GE Healthcare, Milwaukee, WI) to measure PET metabolic



(a) (b)

Figure 1: An example of ROI. (a) is from a 56-year-old patient with primary ovarian cancer who was positive for lymph node metastases. (b) shows the ROI determined according to the above method.



(a) (b)

Figure 2: An example of segmentation. The same patient is shown in Figure 1. The black area in (a) is the ROI area, and (b) is the ROI area segmented into three different subareas according to our segmentation method and marked with different colors.

parameters. Two radiologists with 15 years of experience reviewed the film independently. If no agreement can be reached, we submit the results to a senior physician for a final decision. The software uses an iterative adaptive algorithm to calculate PET parameters, automatically determines the threshold to delineate the tumor edge, and delineates the region of interest (ROI) based on the above information. The radiologist manually corrected the obtained ROI again and determined the final ROI. Ideally, the ROIs should not contain necrotic or cystic components, as shown in Figure 1. In the case of unsatisfactory images, the radiologist delineated the ROIs manually.

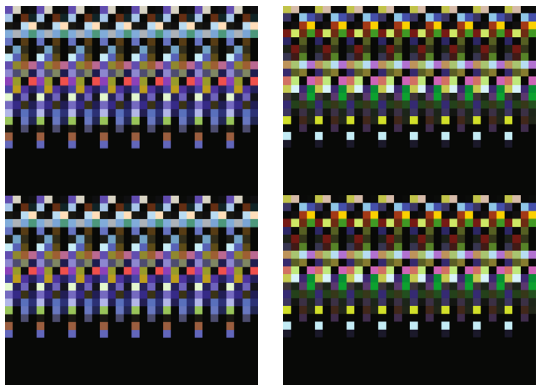
2.3. Lymph Node Status. We obtained pathological slides from the patients enrolled above. No data were excluded because the diagnosis was not precise. The Pathology Department of Shengjing Hospital of China Medical University diagnosed whether lymph nodes were benign or malignant. Each sample was fixed in 10% buffered formalin solution. Paraffin-embedded and serial sections were made at a thickness of 0.004 mm. Fixed tissues were removed by dehydration in an automatic tissue processor and stained with HE. The samples were first dewaxed in xylene and alcohol. The samples were stained with hematoxylin for 5 minutes and eosin for 3 minutes [21]. Finally, the samples were immersed in alcohol and xylene for dehydration and transparency. The slides were secured with synthetic resin. A senior pathologist diagnosed the pathological sections to determine whether the lymph nodes had metastases from ovarian malignancies.

2.4. Image Segmentation. The main processing of images is based on Python. Load the images using the nibabel library [22]. After the image is processed to remove noise, according to the corresponding gray value of each voxel in the 3D ROI, the Gaussian mixture model is used to cluster each voxel. We segmented the PET images according to the clustering results, as shown in Figure 2.

Finally, we calculated the evaluation cluster parameters as modeling parameters. Three features are extracted, which are

(1) Calinski-Harabaz index (CHI)

The score S (Formula (1)) is defined as the ratio of between-cluster dispersion to within-cluster dispersion and



(a) (b)

Figure 3: An example of characteristic map. (a) is a characteristic map of a 62-year-old patient with ovarian cancer who was negative for lymph node metastasis and immunohistochemical results: positive for ER, positive for PR, and negative for P53. (b) is a characteristic map of a 63-year-old patient with ovarian cancer who was negative for lymph node metastases and had immunohistochemical results: positive for ER, positive for PR, and positive for P53. The difference in their feature maps is noticeable.

is calculated by evaluating between-class and within-class variance. The smaller the covariance of the data within categories, the better, and the larger the covariance between categories, the better, in which case CHI will be significant [23]. To sum up, the higher the score, the better the clustering effect.

$$S = \frac{\mathrm{SS}_B}{\mathrm{SS}_W} \times \frac{N-k}{k-1}. \tag{1}$$

Formula (1) is the calculation formula of CHI.

SSB represents the intracluster distance, SSW represents the intercluster distance, the intracluster distance is represented by the distance between the sample point in the cluster and the center point of the cluster, and the intercluster distance is represented by the distance between the sample point and the center point of other clusters. The specific calculation formula is the following formula.

$$\mathrm{CH} = \frac{\left[\sum_{k=1}^{K} n_k c_k - c^2/K - 1\right]}{\sum_{k=1}^{K} \sum_{i=1}^{n_k} d_i - {c_k}^2/N - K}. \tag{2}$$

Formula (2) is the specific calculation formula of CHI.

(2) Silhouette coefficient

Silhouette coefficient is a way to evaluate the clustering effect [24]. Peter J. Rousseeuw first proposed it in 1986. It combines two factors: cohesion and separation. Based on the same original data, it can be used to evaluate the impact of different algorithms or different operation modes of algorithms on the clustering results (Formula (3)). The average distance between the sample and other sample points in the cluster is defined as the cluster cohesion degree a, and the average distance between the sample and all sample points in the nearest cluster is defined as the cluster separation degree b.

$$s = \frac{b-a}{\max(\mathrm{a}, \mathrm{b})}. \tag{3}$$

Formula (3) is the silhouette coefficient calculation formula for a single sample.

For all samples, the contour coefficient is the average of each sample contour coefficient. The value of this index ranges from -1 to 1. When the degree of separation between clusters B is much larger than the degree of cohesion A, the value of the contour coefficient is approximately 1. In this case, the heterogeneity between clusters is high, but the similarity is poor. Therefore, the closer to 1 the value of this index is, the better the clustering effect.

(3) Davies-Bouldin index

As shown in Formula (4), DBI is an index proposed by David L. Davis and Donald Bouldin to evaluate the advantages and disadvantages of clustering algorithms [25].

$$\mathrm{DB} = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} \left(\frac{\mathrm{avg}(C_i) + \mathrm{avg}(C_j)}{d_{\mathrm{cen}}(C_i, C_j)}\right). \tag{4}$$

Formula (4) is the DB calculation formula.

The avg (c) represents the closeness of clustering clusters, and the formula is as the following formula.

$$\mathrm{avg}(C) = \frac{2}{|C|(|C|-1)} \sum_{1<i<j\leq|C|} \mathrm{dist}(x_i, x_j). \tag{5}$$

Formula (5) is the avg(c) calculation formula.

Calculate the distance of sample points in the cluster, d (Formula (6)) represents the distance between the center points of different clusters.

$$d_{\mathrm{cen}}(C_i, C_j) = \mathrm{dist}(u_i, u_j). \tag{6}$$

Formula (6) is the d calculation formula.

2.5. Discrete Wavelet Transform. We used 2D discrete wavelet transform to process PET images [26]. First, 1D-DWT is carried out on each image row to obtain the low-frequency component L and high-frequency component H of the original image in the horizontal direction, and then, 1D-DWT is carried out on each column of the transformed data. The low-frequency component LL in the horizontal and vertical directions, the low-frequency component LH in the horizontal and vertical directions, the low-frequency component HL in the horizontal and vertical directions, and the high-frequency component HH in the horizontal and vertical directions of the original image are obtained. We select three components, LL, LH, and HL, to extract texture information and form the feature map (Figure 3).

ResNet50 is established based on the library TensorFlow. This model is a specific neural network introduced in residual learning for image recognition by He et al., published in 2015 [27]. Due to the use of the global clustering method instead of the connection layer, the model's size is much smaller, which reduces the size of the ResNet model. ResNet's unique feature is looping learning blocks. This means that each layer must be connected to the next, and the distance jumps directly into the layer about two to three hops away. As shown in Figure 4, the model consists of five maximally pooled convolutional layers, followed by a flat layer, a dropout layer, and finally, a single FC layer. 70% of the feature maps are used as training data, and 30% of the feature maps are used as validation data. The training and validation sets' accuracy are the main parameters of supervised model training.

2.6. Data Fusion. We obtained features from two different sources, image texture of image segmentation and wavelet transform, and different forms of data, analyzed the predictive ability of each feature for lymph node condition, selected features with strong correlation, and then performed data fusion. On this basis, we established a model again to predict the lymph node condition.

3. Result

3.1. Clinical Features. The clinical information of the patients is shown in Table 1.

3.2. Prediction Model. Our model shows good prediction ability on both training and testing sets. On the training set, our accuracy was 0.8854, AUC: 0.9472, CI: 0.9098-0.9752, sensitivity was 0.9865, and specificity was 0.7952 (Figure 5).

On the test set, our accuracy was 0.9104, AUC: 0.9259, CI: 0.8417-0.9889, sensitivity was 0.8125, and specificity was 1.0000. As shown in Figure 6, there was no overfitting or underfitting of the model.

For the training and test sets, we plotted calibration curves (Figure 7). Our model shows stable and considerable predictive ability in training and test sets.

4. Discussion

In this study, we obtained the image data of patients with primary ovarian cancer from the texture feature map formed by wavelet transform and the parameters extracted after tumor segmentation. We use the feature map to train the model, which simplifies the training process of the model and eliminates the interference of nontumor region information to the model. The above two kinds of information were modeled to predict lymph node metastasis in patients with primary ovarian cancer. On the one hand, it provides a reference for radiologists to make a more detailed diagnosis of lymph node metastases in patients with imaging findings that are difficult to detect, and on the other hand, it provides clinicians with additional information at the time of surgery.

Studies have shown that multiple imaging methods have a good effect on detecting lymph node metastasis in primary ovarian cancer [28]. A study has shown that CT or PET/CT can effectively detect pelvic and para-aortic lymph node metastasis in ovarian cancer [11]. However, this approach relies heavily on the personal experience of the radiologist [29].

A study conducted on MR has demonstrated that ovarian cancer is a biologically heterogeneous tumor [30]. In ovarian cancer, this heterogeneity is reflected not only in response to neoadjuvant chemotherapy but also in imaging. From the pathological point of view, uncontrolled tumor growth is accompanied the by uneven distribution of tumor parenchyma and stroma, including blood vessels. This uneven distribution of blood vessels will lead to the formation of tumor subareas with different blood supply and support and the formation of different tumor microenvironments. Under such survival pressure, mutations in different directions were formed in each subregion of the tumor, which further aggravated the uneven distribution of parenchyma and interstitium. In conclusion, the subregion composition of tumors themselves can reflect the heterogeneity of the biological behavior of tumors. In this case, the accuracy of the texture information extracted from the tumor imaging data in reflecting the heterogeneity of the tumor is inferior to the characteristics describing the relationship between the various tumor subregions. Therefore, different from traditional radiomics analysis, we did not choose to directly extract texture information from tumor imaging data or extract texture features of each subregion after segmentation for modeling. After image segmentation, we did not simply use the percentage of the total number of voxels in each subregion to the total number of voxels in ROI as a parameter. However, we extracted three parameters describing the relationship between subregions usually used to evaluate the clustering effect of clustering methods. These three parameters not only consider the differences between subregions but also describe the similarities between subregions, which enables us to have a more detailed description of the subregions within the tumor, that is, the imaging heterogeneity of the tumor.

Texture representation is an important problem in medical image analysis. Image texture can be defined as the spatial relationship of pixel values in an image region. In medical images, we consider texture as a local feature pattern that identifies the image intensity of tissue. Texture also determines the local spectral or frequency content in the image. A local texture change should cause a local spatial frequency change. Texture analysis is of interest in medical imaging because as biological tissues become abnormal during the course of the disease, their underlying texture may also change. In the medical field, wavelet transform is widely used to process signals and medical images, and noise removal is undoubtedly the most common field of its use [31]. At the same time, discrete wavelet transform also shows its unique advantages in the nuclear magnetic examination of the human brain [32]. In this experiment, we selected three wavelet transform subbands containing more information to form the feature map and then used the deep learning method to reflect the lymph node metastasis of

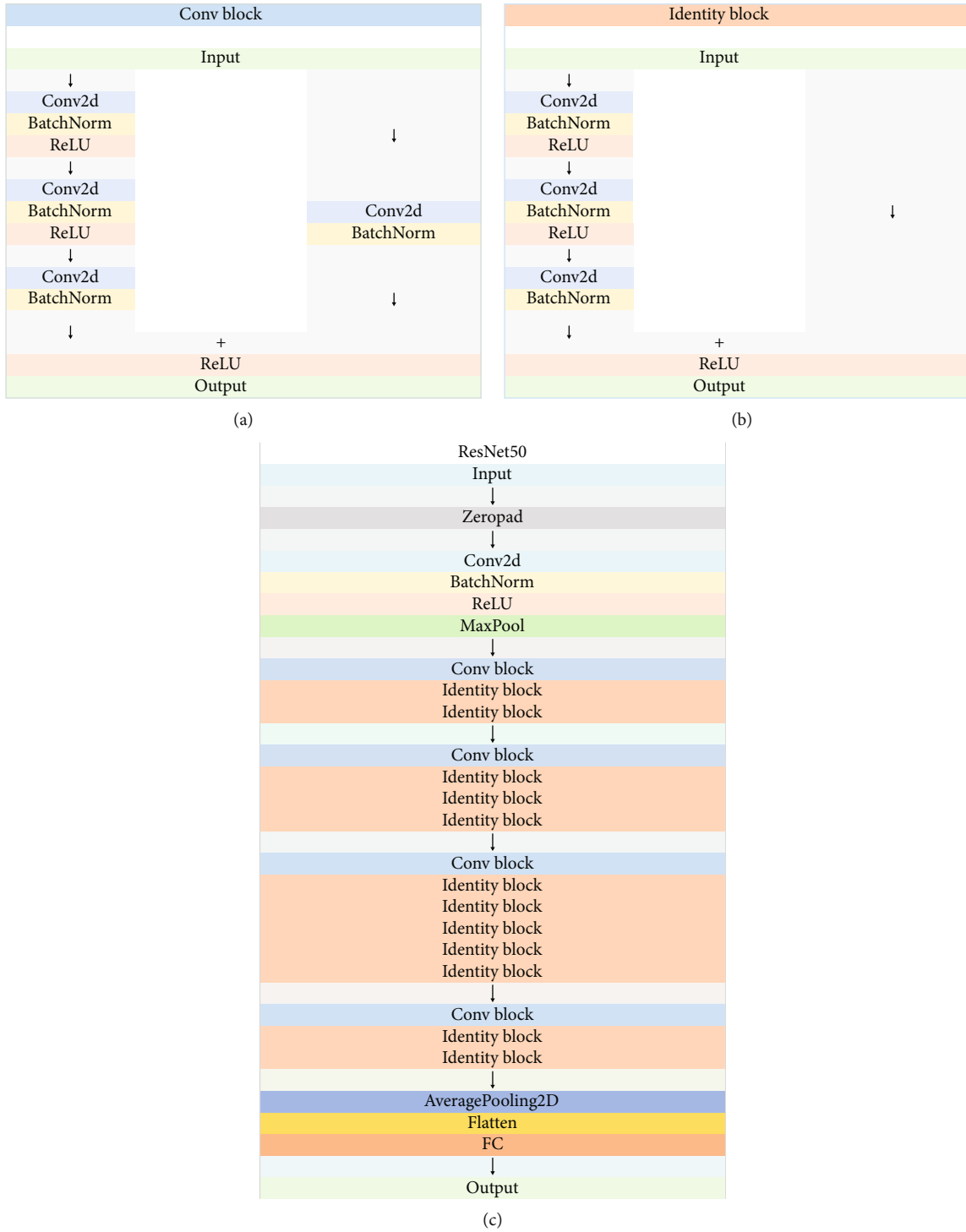


Figure 4: Structure of ResNet50. (a) shows the structure of the Conv block, (b) shows the structure of the Identity block, and (c) shows the structure of ResNet50. In this experiment, we used the classic architecture of ResNet50 without any modification.

ovarian cancer. In fact, with the development of artificial intelligence, computer-aided diagnosis is increasingly used in the field of medicine. According to the global epidemiological situation, there is no doubt that deep learning studies related to COVID-19 are being widely carried out [33]. In addition, the computer-aided diagnosis of breast cancer has been fairly accurate [34]. Due to the number of open source databases, gastrointestinal cancer research is undoubtedly the most widely carried out [35]. Ovarian cancer is the three most common gynecological diseases in the world, especially in the field of medical imaging, including tomography, ultrasonography (US), magnetic resonance imaging (MRI), and other imaging methods. We have made great strides in artificial intelligence technology [36].

Table 1: Clinical information of patients in test and training sets.

Age (year)	Training set	Test set
<65	17	7
>65	140	60
Lymph nodes		
Positive	74	32
Negative	83	35
HER-2		
Positive	93	41
Negative	64	26
PR		
Positive	59	23
Negative	98	44
p53		
Positive	101	47
Negative	56	20

HER-2: human epidermal growth factor receptor 2; PR: progesterone receptor. The PR, HER-2, and p53 status of patients was determined by immunohistochemistry.

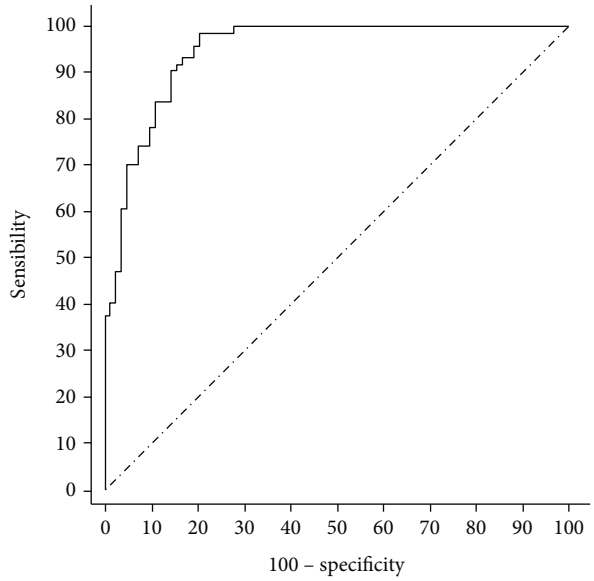


Figure 5: The sensitivity and specificity curve of the training set.

Therefore, ResNet50 was also used to investigate the lymph node metastasis of ovarian cancer. Previously, ResNet50 was used in the medical field to detect COVID-19 and was considered good results [37]. In addition, multiple image segmentation methods have been proposed for deep learning of pathology [38]. On the inspection side, a neural network is more effective at identifying microbes [39]. In addition to the simple recognition and segmentation of medical images, deep learning can combine the medical information contained in images with the clinical information of patients to predict the prognosis of patients. A study has shown that combining patient pathology, CT, and genetic information can predict the prognosis of ovarian cancer patients [40]. Of course, manual diagnosis is currently the gold standard

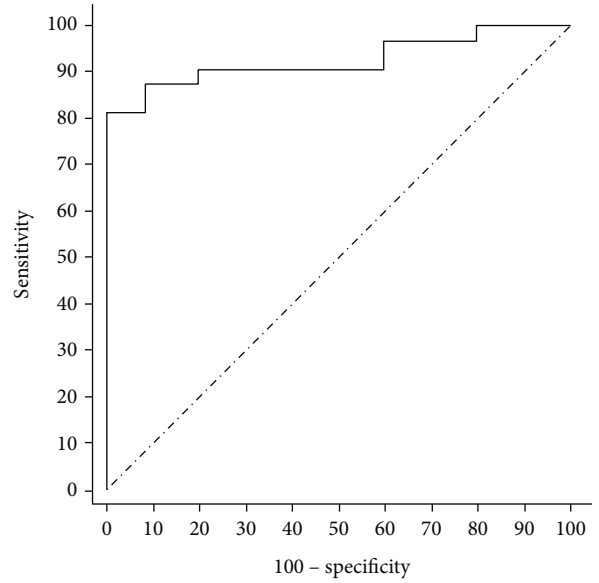


Figure 6: The sensitivity and specificity curve of the test set.

of imaging diagnosis. However, computer-aided diagnosis is undoubtedly more sensitive and efficient in finding hidden lesions and exploring the correlation between different medical information of patients. On the one hand, this will effectively reduce the workload of radiologists; on the other hand, it will also quantify the diagnosis.

In previous studies, researcher mainly have adopted two strategies for predicting of lymph node metastasis. One strategy adopts the deep learning method, which divides the tumor's imaging data into different samples according to the scanning level, and uses the whole CT or MRI for learning [41]. Another strategy is to delineate the ROI of the tumor and perform machine learning on the texture features extracted from the ROI [42]. In a variety of different tumors, ROIs covering tumors were used to predict lymph node metastasis. Both texture analysis and deep learning have made good progress. A study has shown that *T*2 -weighted magnetic resonance imaging (T2WI) texture features have a high value in predicting preoperative lymph node invasion in rectal cancer [43]. In papillary thyroid carcinoma, MRI can effectively identify lymph node metastasis [44]. An MRI texture analysis based on machine learning has shown that occult lymph node metastases in early oral tongue squamous cell carcinoma can also be effectively identified [45]. And a study using deep learning confirmed that MRI can be used to predict cervical cancer metastasis [46]. The problem with the first strategy is that the tumor is not a normal physiological structure. If the researchers segment the tumor image according to the scan level, they undoubtedly destroys the integrity of the tumor space. Cancer is a heterogeneous disease, both within the tumor and between different patients. So this method of generating datasets is also difficult to standardise. In addition, this method often directly learns the features of the metastatic lymph nodes, ignoring the characteristics of the tumor itself. The problem with the second strategy is that as deep learning techniques mature, traditional machine learning algorithms are no

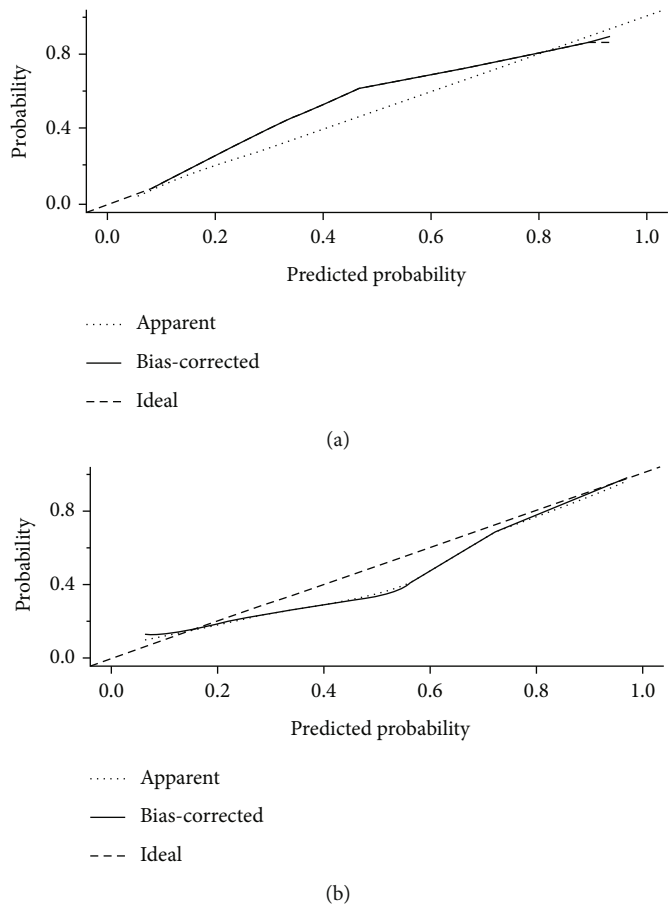


Figure 7: Calibration curve. (a) is the calibration curve of the training set, and (b) is the calibration curve of the test set.

longer the most advanced. In this experiment, after obtaining the 3D ROI, we conducted deep learning on the extracted texture information and the segmented parameters. In addition to obtaining more image information, we pay more attention to the characteristics of the tumor itself, so as to predict the lymph node metastasis. This approach is more clinical.

However, despite the consideration of various problems, there are still several shortcomings in this study. First of all, the data of this study are all from Shengjing Hospital of China Medical University, which has few cases and is not a multicenter study. To ensure the quality of images, it is undoubtedly most important to expand the sample size for radiological model testing. Second, in image segmentation, we use the Gaussian mixture model to segment the image into three subregions. Technically, this segmentation method is beyond criticism. However, whether it is the most accurate segmentation method to reflect the internal subregion of ovarian cancer tumors still needs further experiments to verify the clinical significance of this segmentation method. Typically, imaging subregional changes in patients who develop resistance during chemoradiotherapy, or specific binding probes, are potent answers to this question [47]. A study shows that clustering after hypersegmentation of images can more accurately describe the features of the inner subregions of images [48]. In addition, our modeling only considers the imaging characteristics of patients. For an early warning model, there is no doubt that the inclusion of patients' clinical information can improve the model's accuracy on the premise of ensuring clinical significance. Age, endocrine status, and genetic mutations, common risk factors for ovarian cancer, were not included because data sources limited them. Finally, this study was limited to patients with untreated early-stage ovarian cancer because it was actually to investigate the relationship between tumor heterogeneity and its biological behavior. For patients with treated or advanced ovarian cancer, the heterogeneity on medical imaging has changed and therefore the conclusions of this study are not applicable. In the future, we may consider further prospective experiments to include clinical information to improve the accuracy of the model, follow up on the prognosis of patients, and complete the monitoring of the whole course of ovarian cancer patients.

5. Conclusion

In this study, we used PET images of ovarian cancer to segment the images on the one hand and extract feature parameters according to the segmentation results; on the other hand, we used discrete wavelet transform to process the images and extract feature maps. We established a model to predict the presence of lymph node metastasis by the two aspects of information. This method can provide a

reference for the treatment of patients, such as surgery and medication, and the prognosis of patients under the noninvasive condition.

Data Availability

The image data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was funded by China Medical University, Postgraduate Research Scholarship, 2019120823.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[2] G. C. Jayson, E. C. Kohn, H. C. Kitchener, and J. A. Ledermann, "Ovarian cancer," *The Lancet*, vol. 384, no. 9951, pp. 1376–1388, 2014.

[3] R. J. Kurman and I.-M. Shih, "The dualistic model of ovarian carcinogenesis: revisited, revised, and expanded," *The American Journal of Pathology*, vol. 186, no. 4, pp. 733–747, 2016.

[4] S. Lheureux, C. Gourley, I. Vergote, and A. M. Oza, "Epithelial ovarian cancer," *The Lancet*, vol. 393, no. 10177, pp. 1240–1253, 2019.

[5] S. Zhang, C. Cheng, Z. Lin et al., "The global burden and associated factors of ovarian cancer in 1990–2019: findings from the global burden of disease study 2019," *BMC Public Health*, vol. 22, no. 1, pp. 1–41, 2022.

[6] K. Moore, N. Colombo, G. Scambia et al., "Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer," *New England Journal of Medicine*, vol. 379, no. 26, pp. 2495–2505, 2018.

[7] S. Banerjee, K. N. Moore, N. Colombo et al., "Maintenance olaparib for patients with newly diagnosed advanced ovarian cancer and a _BRCA_ mutation (SOLO1/GOG 3004): 5-year follow-up of a randomised, double-blind, placebo-controlled, phase 3 trial," *The Lancet Oncology*, vol. 22, no. 12, pp. 1721–1731, 2021.

[8] J. Prat and F. C. O. G. Oncology, "Staging classification for cancer of the ovary, fallopian tube, and peritoneum," *International Journal of Gynaecology and Obstetrics: The Official Organ of the International Federation of Gynaecology and Obstetrics*, vol. 124, no. 1, pp. 1–5, 2014.

[9] K. Matsuo, T. B. Sheridan, K. Yoshino et al., "Significance of lymphovascular space invasion in epithelial ovarian cancer," *Cancer Medicine*, vol. 1, no. 2, pp. 156–164, 2012.

[10] S. I. Kim, M. C. Lim, J. S. Lee et al., "Impact of lower limb lymphedema on quality of life in gynecologic cancer survivors after pelvic lymph node dissection," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 192, pp. 31–36, 2015.

[11] C. Mimoun, R. Rouzier, J. L. Benifla, A. Fauconnier, and C. Huchon, "Preoperative ct or pet/ct to assess pelvic and para-aortic lymph node status in epithelial ovarian cancer? A systematic review and meta-analysis," *Diagnostics*, vol. 11, no. 10, p. 1748, 2021.

[12] I. O. Lee, J.-Y. Lee, H. J. Kim et al., "Prognostic significance of supradiaphragmatic lymph node metastasis detected by 18F-FDG PET/CT in advanced epithelial ovarian cancer," *BMC Cancer*, vol. 18, no. 1, pp. 1–8, 2018.

[13] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, no. 1, pp. 1–9, 2014.

[14] S. Nougaret, M. Tardieu, H. Vargas et al., "Ovarian cancer: an update on imaging in the era of radiomics," *Diagnostic and Interventional Imaging*, vol. 100, no. 10, pp. 647–655, 2019.

[15] C. Balana, S. Castañer, C. Carrato et al., "Preoperative diagnosis and molecular characterization of gliomas with liquid biopsy and radiogenomics," *Frontiers in Neurology*, vol. 13, p. 1097, 2022.

[16] J. Zhang, C. Li, M. M. Rahaman et al., "A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approaches," *Artificial Intelligence Review*, vol. 55, pp. 2875–2944, 2021.

[17] J. Zhang, X. Zhao, T. Jiang et al., "An application of pixel interval down-sampling (PID) for dense tiny microorganism counting on environmental microorganism images," *Applied Sciences*, vol. 12, no. 14, p. 7314, 2022.

[18] H. Lee, S. Yune, M. Mansouri et al., "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets," *Nature Biomedical Engineering*, vol. 3, no. 3, pp. 173–182, 2019.

[19] D. Zhou, F. Tian, X. Tian et al., "Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer," *Nature Communications*, vol. 11, no. 1, pp. 1–9, 2020.

[20] A. Chen, C. Li, S. Zou et al., "SVIA dataset: a new dataset of microscopic videos and images for computer- aided sperm analysis," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 204–214, 2022.

[21] G. Swamy and N. Satyanarayana, "Clinicopathological analysis of ovarian tumors–a study on five years samples," *Nepal Medical College Journal*, vol. 12, no. 4, pp. 221–223, 2010.

[22] M. Brett, C. J. Markiewicz, M. Hanke et al., *nipy/nibabel: 3.0. 0*, Zenodo, Geneve, Switzerland, 2019.

[23] J. Wang, W. Zhang, T. Hua, and T. C. Wei, "Unsupervised learning of topological phase transitions using the Calinski-Harabaz index," *Physical Review Research*, vol. 3, no. 1, article 013074, 2021.

[24] F. Krasnov and A. Sen, "The number of topics optimization: clustering approach," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, p. 25, 2019.

[25] J. Xiao, J. Lu, and X. Li, "Davies Bouldin index based hierarchical initialization K-means," *Intelligent Data Analysis*, vol. 21, no. 6, pp. 1327–1338, 2017.

[26] K. H. Ghazali, M. F. Mansor, M. M. Mustafa, and A. Hussain, "Feature Extraction Technique Using Discrete Wavelet Transform for Image Classification," in *2007 5th Student Conference on Research and Development*, pp. 1–4, Selangor, Malaysia, 2007.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.

[28] Y. Yuan, Z.-X. Gu, X.-F. Tao, and S. Y. Liu, "Computer tomography, magnetic resonance imaging, and positron emission tomography or positron emission tomography/computer tomography for detection of metastatic lymph nodes in patients with ovarian cancer: a meta-analysis," *European Journal of Radiology*, vol. 81, no. 5, pp. 1002–1006, 2012.

[29] K. Bagul, D. Vijaykumar, A. Rajanbabu, M. A. Antony, and V. Ranganathan, "Advanced primary epithelial ovarian and peritoneal carcinoma—does diagnostic accuracy of preoperative CT scan for detection of peritoneal metastatic sites reflect into prediction of suboptimal debulking? A prospective study," *Indian Journal of Surgical Oncology*, vol. 8, no. 2, pp. 98–104, 2017.

[30] E. Sala, M. Y. Kataoka, A. N. Priest et al., "Advanced ovarian cancer: multiparametric MR imaging demonstrates response- and metastasis-specific effects," *Radiology*, vol. 263, no. 1, pp. 149–159, 2012.

[31] A. F. Laine, "Wavelets in temporal and spatial processing of biomedical images," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 511–550, 2000.

[32] E. Bullmore, J. Fadili, M. Breakspear, R. Salvador, J. Suckling, and M. Brammer, "Wavelets and statistical analysis of functional magnetic resonance images of the human brain," *Statistical Methods in Medical Research*, vol. 12, no. 5, pp. 375–399, 2003.

[33] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[34] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," 2016, https://arxiv.org/abs/1606.05718.

[35] H. Chen, C. Li, X. Li et al., "Gashis-transformer: a multi-scale visual transformer approach for gastric histopathology image classification," https://arxiv.org/abs/2104.14528.

[36] X. Chen, X. Huo, Z. Wu, and J. Lu, "Advances of artificial intelligence application in medical imaging of ovarian cancers," *Chinese Medical Sciences Journal*, vol. 36, no. 3, pp. 196–203, 2021.

[37] W. Zouch, D. Sagga, A. Echtioui et al., "Detection of COVID-19 from CT and chest X-ray images using deep learning models," *Annals of Biomedical Engineering*, vol. 50, pp. 825–835, 2022.

[38] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-Unet: an improved neural network based on Unet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, pp. 21420–21428, 2019.

[39] J. Zhang, C. Li, S. Kosov et al., "LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, article 107885, 2021.

[40] K. M. Boehm, E. A. Aherne, L. Ellenson et al., "Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer," *Nature Cancer*, vol. 3, no. 6, pp. 723–733, 2022.

[41] S. Bedrikovetski, N. N. Dudi-Venkata, G. Maicas et al., "Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: A systematic review and meta-analysis," *Artificial Intelligence in Medicine*, vol. 113, article 102022, 2021.

[42] Y. Y. Yu, R. Zhang, R. T. Dong et al., "Feasibility of an ADC-based radiomics model for predicting pelvic lymph node metastases in patients with stage IB–IIA cervical squamous cell carcinoma," *The British Journal of Radiology*, vol. 92, no. 1097, article 20180986, 2019.

[43] L. Song and J. Yin, "Application of texture analysis based on sagittal fat-suppression and oblique axial T2-weighted magnetic resonance imaging to identify lymph node invasion status of rectal cancer," *Frontiers in Oncology*, vol. 10, p. 1364, 2020.

[44] H. Qin, Q. Que, P. Lin et al., "Magnetic resonance imaging (MRI) radiomics of papillary thyroid cancer (PTC): a comparison of predictive performance of multiple classifiers modeling to identify cervical lymph node metastases before surgery," *La Radiologia Medica*, vol. 126, no. 10, pp. 1312–1327, 2021.

[45] Y. Yuan, J. Ren, and X. Tao, "Machine learning–based MRI texture analysis to predict occult lymph node metastasis in early-stage oral tongue squamous cell carcinoma," *European Radiology*, vol. 31, no. 9, pp. 6429–6437, 2021.

[46] Q. Wu, S. Wang, S. Zhang et al., "Development of a deep learning model to identify lymph node metastasis on magnetic resonance imaging in patients with cervical cancer," *JAMA Network Open*, vol. 3, no. 7, pp. e2011625–e2011625, 2020.

[47] J. R. Teruel, M. G. Heldahl, P. E. Goa et al., "Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer," *NMR in Biomedicine*, vol. 27, no. 8, pp. 887–896, 2014.

[48] C.-M. Pun, X.-C. Yuan, and X.-L. Bi, "Image forgery detection using adaptive oversegmentation and feature point matching," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 8, pp. 1705–1716, 2015.

Hindawi
BioMed Research International
Volume 2022, Article ID 2961610, 12 pages
https://doi.org/10.1155/2022/2961610

Research Article

Segmentation of Breast Tubules in H&E Images Based on a DKS-DoubleU-Net Model

Yuli Chen,[1] **Yao Zhou**,[1] **Guoping Chen**,[1] **Yuchuan Guo**,[2] **Yanquan Lv**,[2] **Miao Ma**,[1] **Zhao Pei**,[1] **and Zengguo Sun**[1]

[1]*School of Computer Science, Shaanxi Normal University, Xi'an, China*
[2]*Shandong Junteng Medical Technology Co., LTD, China*

Correspondence should be addressed to Zengguo Sun; sunzg@snnu.edu.cn

Received 5 August 2022; Accepted 10 September 2022; Published 30 September 2022

Academic Editor: Cheng Lu

Copyright © 2022 Yuli Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The formation of breast tubules plays an important role in the pathological grading of breast cancer. Breast tubules surrounded by a large number of epithelial cells are located in the subcutaneous tissue of the chest. The shapes of breast tubules are various, including tubular, round, and oval, which makes the process of breast tubule segmentation a difficult task. Deep learning technology, capable of learning complex data structures via efficient representation, could help pathologists accurately detect breast tubules in hematoxylin and eosin (H&E) stained images. In this paper, we propose a deep learning model named DKS-DoubleU-Net to accurately segment breast tubules with complex appearances in H&E images. The proposed DKS-DoubleU-Net model suggests using a DenseNet module as the encoder of the second subnetwork of DoubleU-Net, which utilizes dense features between layers and strengthens the propagation of features extracted in all previous layers, in order to better discover the intrinsic characteristics of breast tubules with complex structures and diverse shapes. Moreover, a feature fusing module called Kernel Selecting Module (KSM) is inserted before each output layer of the two U-Net branches of the DoubleU-Net, to implement a multiscale feature fusion via a self-adaptive kernel selecting for the sake of accurate segmentation of breast tubules in different sizes. The experiments on the public BRACS dataset and a private clinical dataset have shown that our model achieves better segmentation performance, compared to the state-of-art models of U-Net, DoubleU-Net, ResUnet++, HRNet, and DeepLabV3+. Specifically, on the public BRACS dataset, our method produced an F1-Score of 92.98%, which outperforms the F1-Score of U-Net, DoubleU-Net, and HRNet by 4.24%, 0.37%, and 1.68%, respectively, and is much better than performances of DeepLabV3+ and ResUnet++ by 7.83% and 23.84%, respectively. On the private clinic dataset, the proposed model achieved an F1-Score of 73.13%, which has shown an improvement of 10.31%, 1.89%, 4.88%, 15.47%, and 31.1% to the performances of the U-Net, DoubleU-Net, HRNet, DeepLabV3+, and ResUnet++, respectively. Superior performance could also be observed when comparing the proposed DKS-DoubleU-Net with the others using the metrics of Dice and mIou.

1. Introduction

Breast cancer is one of the most common cancers for women in the world [1]. With the increase of breast cancer patients worldwide, the precise diagnosis and prognosis of breast cancer have become urgent tasks. The Nottingham grading system [2] has become the most adopted clinic routine criteria in the diagnosis and prognosis of breast cancer. As one of the three essential factors (tubule formation, nuclear pleomorphism, and mitotic counts) in the Nottingham grading system, the assessment of breast tubule formation plays an important role [2].

The arrangement and morphology of tumor cells within breast tubules are vital criterions for the diagnosis of breast ductal carcinoma. Breast ductal carcinoma will destroy the function of breast ducts, resulting in the disappearance of tubular structure of breast tubules. Therefore, it is particularly important to accurately identify breast tubules in pathological images for a precise grading and diagnosis of breast cancer patients.

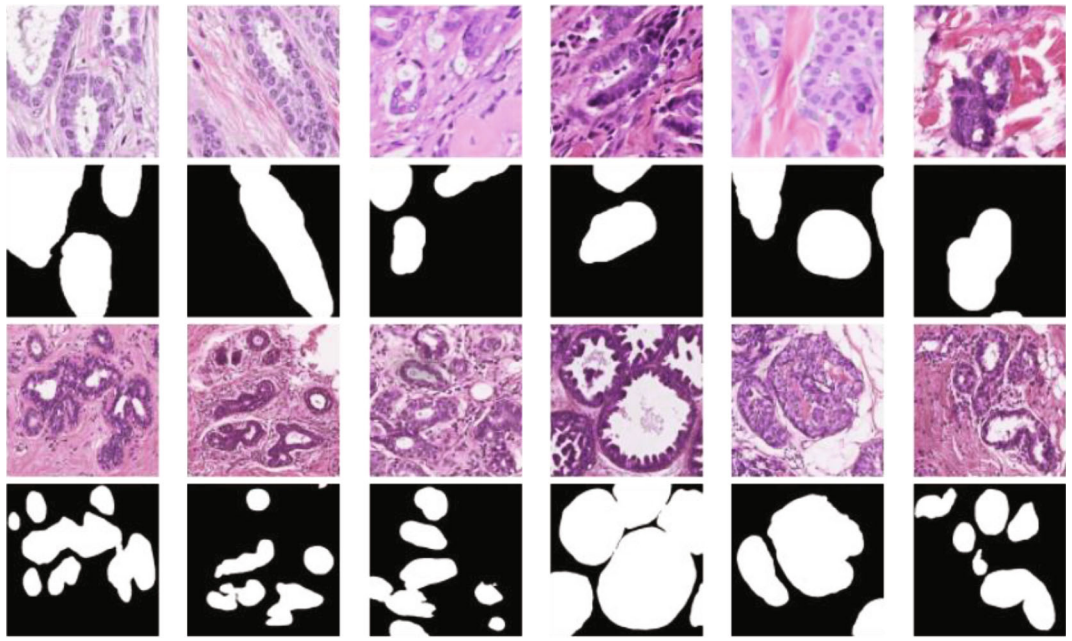


Figure 1: Demostration of the complexity of breast tubules. Breast tubules are usually of diverse appearances, complex structures, and highly sophisticated morphology. The first and third row show several breast tubular images from public BRACS dataset and private clinical dataset. The second and fourth row represent their corresponding ground truths.

The identification of breast tubules is usually done by the manual observation of pathologists through microscopes. This process is not only time- and labor-consuming but also prone to errors due to interobserver and intraobserver discrepancies. Thus, it is necessary to develop a model to automatically recognize the breast tubule in histopathological images for a more objective and accurate computational pathology analysis.

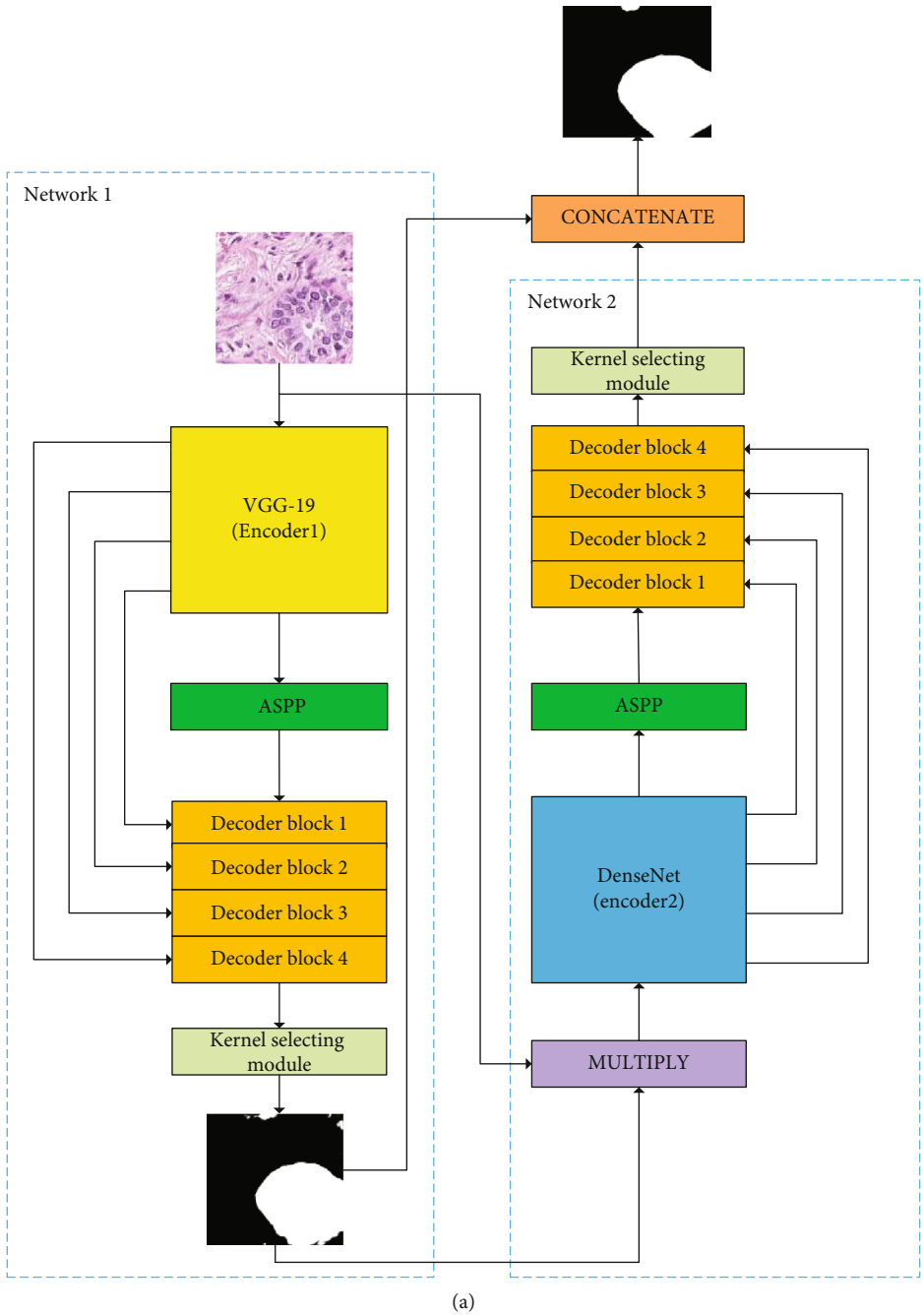
Breast tubules are located in the subcutaneous tissue of the chest. Although breast tubules are characterized by clear lumina surrounded by epithelial cells, they are easily confused with clefts induced by shrinkage artifacts [2]. Moreover, breast tubules are typical of diverse appearances, complex structures, and highly sophisticated morphology, such as various shapes (tubular, round, or oval), colors, and textures, as shown in Figure 1. Therefore, it is a challenging task for traditional machine learning methods to identify breast tubules effectively. A promising way is to train end-to-end deep learning models [3–7], which could automatically discover the essential characteristics of breast tubules through the nonlinear representation of multiple layers of neurons.

Nowadays, with the rapid development of deep learning technology, increasing image segmentation models, such as U-Net [8], ResUNet++ [9], DeepLabV3+ [10], DoubleU-Net [11], and HRNet [12], have shown good performance in the field of medical image segmentation. For example, the U-Net model was applied to the segmentation of neuronal structures in electron microscopic recordings and the cell segmentation task in light microscopic images. Moreover, many U-Net-based Convolutional Neural Networks (CNNs) have also demonstrated superior performance in the task of medical image segmentation [13–17]. Particularly, among the outstanding U-Net derived models, DoubleU-Net has shown promising results on various medical image datasets, including polyp detection dataset, lesion boundary segmentation dataset, and nuclei image dataset.

There have been some approaches dedicated to semantically segmenting breast tubules in hematoxylin and eosin (H&E) stained images. For example, Wang et al. [18] proposed a two-step clustering and random forest method for automatic recognition and segmentation of breast tubules. Basavanhally et al. [19] applied the O'Callaghan neighborhood in modeling and imposed spatial distance with directional constraints on object attributes for tubular identification on histopathological images stained with hematoxylin and eosin (H&E). Maqlin et al. [20] used K-means clustering to identify objects including glandular tubules and fatty regions. Zhang et al. [21] proposed a method for detecting tubules in testicular images based on boundary weighting and circular shortest path. Janowczyk and Madabhushi [22] detected breast tubules, lymphocytes, and epithelial cells by using the deep learning framework of the Caffe AlexNet network model.

However, the current deep learning models are still incapable of accurate segmentation of breast tubules, because breast tubules are usually of diverse appearance, complex structures, and different sizes for patients in different grading stages and various tissue transections.

Although deep-learning-based methods have been applied to the segmentation of breast tubules [8–12], most of them have not taken into account the reuse and fusion of multilevel features to better describe the tubule ROI regions of complex structures, diverse shapes, and different sizes for a more accurate segmentation performance.



(a)

Figure 2: Continued.

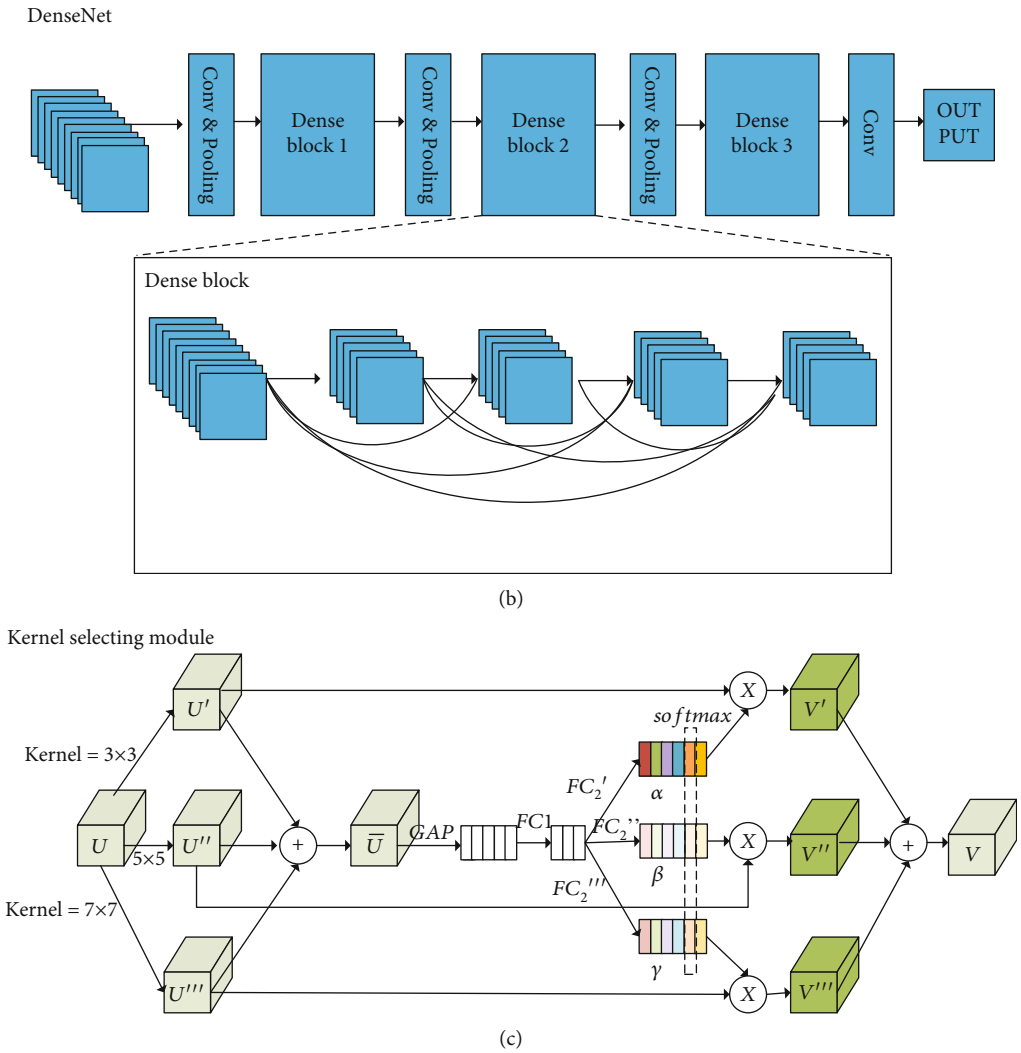


Figure 2: The overall architecture of the proposed DKS-DoubleU-Net model and the details of the introduced DenseNet and Kernel Selecting Module (KSM) for semantic breast tubule segmentation. (a) The overall architecture of the proposed DKS-DoubleU-Net model. (b) The detailed architecture of the DenseNet module [23]. (c) The detailed architecture of the Kernel Selecting Module (KSM) module [24].

Table 1: The quantitative comparison of the tubule segmentation results obtained by different models on the test set of the private clinical dataset.

Model	Dice	mIou	Recall	Precision	F1-score
DKS-DoubleU-Net	0.7002	0.5547	0.7636	0.7016	0.7313
DoubleU-Net+DenseNet	0.6897	0.5412	0.7442	0.7066	0.7249
DoubleU-Net+KSM	0.6820	0.5364	0.7165	0.7243	0.7204
DoubleU-Net	0.6798	0.5308	0.7450	0.6825	0.7124
U-Net	0.6115	0.4519	0.6568	0.6020	0.6282
ResNet++	0.3922	0.2672	0.6610	0.3081	0.4203
DeepLabV3+	0.5513	0.3955	0.5885	0.5651	0.5766
HRNet	0.6608	0.5085	0.6496	0.7188	0.6825

In this paper, we propose a novel semantic segmentation model named DKS-DoubleU-Net by integrating a DenseNet [23] module and a Kernel Selecting Module (KSM) [24] into the DoubleU-Net model for accurate segmentation of breast tubules in H&E images. The adopted DenseNet encoder utilizes dense features between layers to discover intrinsic characteristics and strengthen the feature propagation of breast tubules. Besides, the KSM module conducts a self-adaptive kernel selecting for the sake of a multilevel feature fusion, facilitating the detection of breast tubules of different sizes. Experiments on

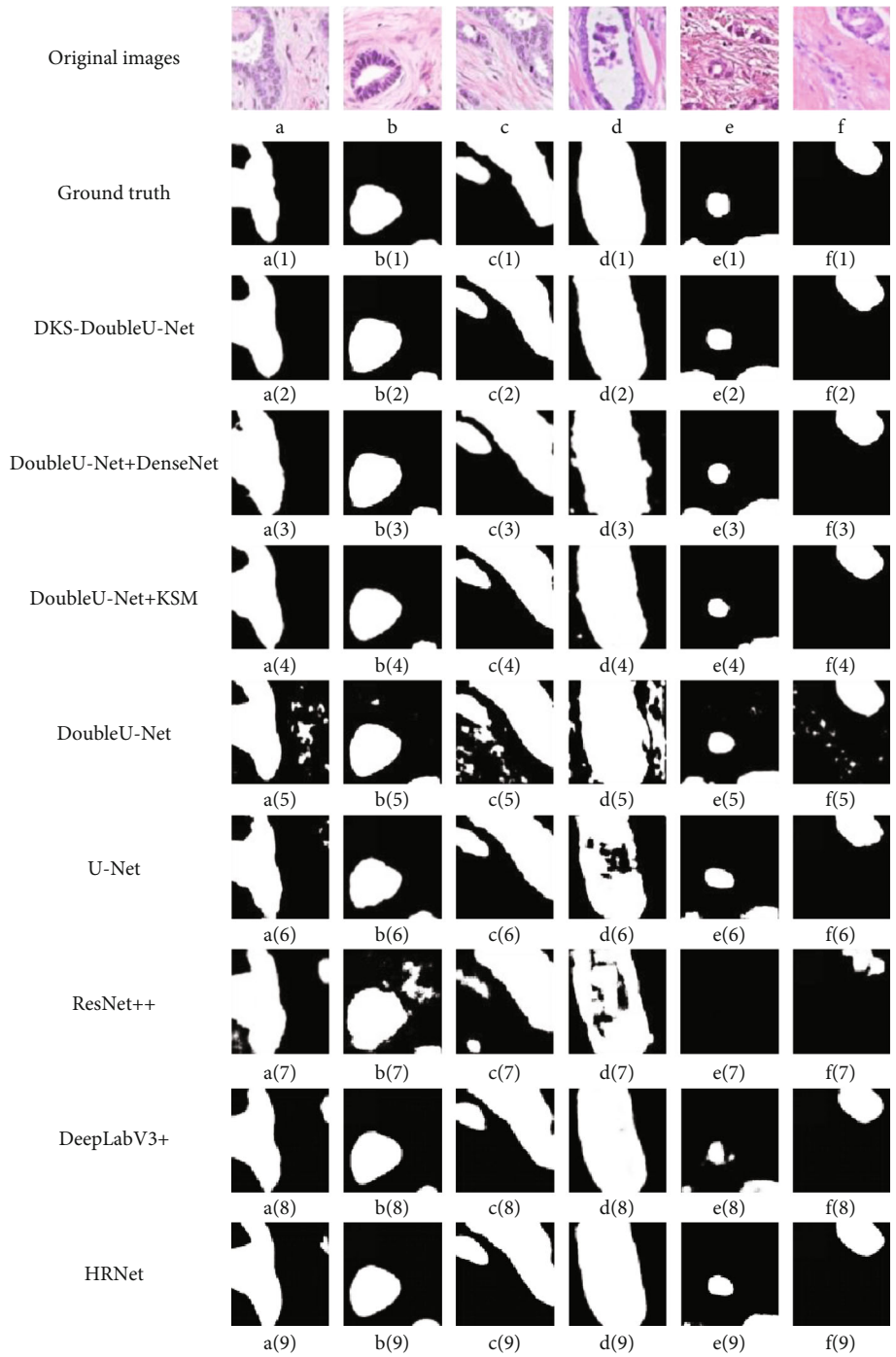


Figure 3: The illustrative tubule segmentation results of different models on the test set of the private clinical dataset. The first row lists the original images; images a(1)-f(1) in the second row are the corresponding tubule annotation masks; images a(2)-f(2) in the third row are the tubule segmentation results performed by DKS-DoubleU-Net; images a(3)-f(3) in the fourth row are the results output by the DoubleU-Net which added DenseNet; images a(4)-f(4) in the fifth row are the results obtained by the DoubleU-Net which added Kernel Selecting Module (KSM); images a(5)-f(5) in the sixth row are the results output by the DoubleU-Net; images a(6)-f(6) in the seventh row are the results output by Unet; images a(7)-f(7) correspond to the results of ResUnet++; images a(8)-f(8) are the results obtained by DeepLabV3+, and images a(9)-f(9) show the tubule segmentation results of HRNet.

the public BRACS dataset [25] and a private clinical dataset have demonstrated the outperformance of the proposed DKS-DoubleU-Net in the semantic segmentation of the breast tubules, compared to the original DoubleU-Net as well as the state-of-the-art segmentation models including U-Net [8], HRNet [12], DeepLabV3+ [10], and ResUNet++ [9].

2. Methods

2.1. The Proposed DKS-DoubleU-Net Architecture. For accurate semantic segmentation of breast tubules with complex structures and shapes, the proposed DKS-DoubleU-Net model is constructed through two main processes: (1) substituting the encoder of the second U-Net subnetwork branch of DoubleU-Net with a DenseNet [23] module; (2) inserting a Kernel Selecting Module (KSM) before each output layer of the two U-Net branches of the DoubleU-Net. The overall architecture of the proposed DKS-DoubleU-Net model is shown in Figure 2.

As shown in Figure 2, the proposed DKS-DoubleU-Net is composed of two U-Net subnetwork branches which are connected end to end, and each network branch is composed of an encoder and a decoder. Specifically, the VGG19 [26] is employed as the encoder module in the first network branch. The decoder modules of the two subnetwork branches are identical, both including four decoder blocks, and each of the decoder blocks consists of a squeeze and excitement block [27], 3×3 convolutional layers, and an upsampling layer. For each branch, the encoder and the decoder are connected with an Atrous Spatial Pyramid Pooling (ASPP) [10], which expands the receptive field of the convolutional layer. When the input image is fed to the first network, a preliminary binary segmentation mask will be generated to indicate the coarse detected regions of the tubules. Subsequently, the binary preliminary output tubule segmentation mask is multiplied by the original input image to remain only the coarse tubule ROI region in the original image. And the coarse tubule ROI image is further served as the input of the second network branch. The KSM module is inserted before each output layer of the two U-Net subnetwork branches. Finally, after the DenseNet encoder module and the decoder block of the second subnetwork branch, the final result of breast tubule segmentation on the input H&E image is output.

We will describe each component of the proposed DKS-DoubleU-Net model in detail.

2.2. Encoder. Feature extraction is the major function of the encoders in both branches [28]. The encoder of the first network branch adopts a VGG19 module, and all the convolutional layers of the module employ the same small size of 3×3 convolution kernels. Thus, VGG19 with a deep enough network structure could extract plenty of nonlinear features.

As the encoder of the second network branch, the DenseNet with a dense connection mechanism is introduced to extract the complex tubule structure features for a better tubule segmentation in H&E images. Specifically, each Dense Block (as shown in Figure 2(b)) of the introduced DenseNet module densely collects the output feature maps of all previous Dense Blocks as its input. Such dense connection of Dense Blocks makes the propagation and utilization of multilevel features more effective, resulting in a more efficient training process. That is, by establishing dense connections between the front and back layers, the reuse and propagation of features between different layers could alleviate the problem of gradient disappearance caused by the deepening of the network to a certain extent. Therefore, the introduced DenseNet encoder has the potential to effectively extract, fuse, and propagate multilevel features of breast tubules on the ROI regions detected by the first U-Net branch.

2.3. Decoder. Decoders are used to restore the feature maps to the original resolution. The model has two decoders, each of which consists of four decoder blocks, as shown in Figure 2(a). For each layer of the decoder block, an upsampling operation is applied to enlarge the image through interpolation [29]. And then the upsampling feature map is concatenated with the feature map output from the corresponding encoder through a skip connection. Different from the skip connection from the encoder to the decoder in the first subnetwork branch, there are two skip connections from both encoder1 and encoder2 to the decoder in the second subnetwork branch. The purpose of this process is to reduce information loss in the process of feature extraction. After the upsampling, there are two 3×3 convolution operations, each of which is followed by the operations of normalization and ReLU [30]. The subsequent layers are composed of a squeeze and excitation block, a 1×1 convolution operation, and a sigmoid activation function [31]. The final output of the model is a binary image [32] having the same size as the input image to indicate the segmentation regions of breast tubules.

2.4. Kernel Selecting Module (KSM). The Kernel Selecting Module (KSM) fuses multiple convolutional branches with different kernel sizes to implement adaptive receptive field selection and multiscale feature fusion, as shown in Figure 2(c). Firstly, the input $U \in \mathfrak{R}^{H \times W \times C}$ performs 3×3, 5×5, and 7×7 convolution operations, respectively, to obtain $U' \in \mathfrak{R}^{H \times W \times C}$, $U'' \in \mathfrak{R}^{H \times W \times C}$, and $U''' \in \mathfrak{R}^{H \times W \times C}$. Secondly, U', U'', and U''' are added to obtain $\bar{U}$. The calculation formula of $\bar{U}$ is

$$\bar{U} = U' + U'' + U'''. \tag{1}$$

Then, the global average pooling operation (GAP) and the full connection operation (FC_1) are performed on $\bar{U}$, and after that, we perform the full connection operation (FC_2) on the obtained result again. The specific calculation formula is as follows:

$$\begin{aligned} \alpha' &= FC_2{}'\left(FC_1\left(GAP\left(\bar{U}\right)\right)\right), \\ \beta' &= FC_2{}''\left(FC_1\left(GAP\left(\bar{U}\right)\right)\right), \\ \gamma' &= FC_2{}'''\left(FC_1\left(GAP\left(\bar{U}\right)\right)\right), \end{aligned} \tag{2}$$

where $\alpha' \in \mathfrak{R}^{1 \times 1 \times C}$, $\beta' \in \mathfrak{R}^{1 \times 1 \times C}$, and $\gamma' \in \mathfrak{R}^{1 \times 1 \times C}$. Three obtained outputs are calculated by softmax, respectively.

Table 2: The quantitative comparison of tubule segmentation results obtained by different models on the test set of the public BRACS dataset.

Model	Dice	mIou	Recall	Precision	F1-Score
DKS-DoubleU-Net	0.9272	0.8655	0.9292	0.9305	0.9298
DoubleU-Net+DenseNet	0.9247	0.8617	0.9288	0.9259	0.9273
DoubleU-Net+KSM	0.9269	0.8655	0.9328	0.9264	0.9296
DoubleU-Net	0.9219	0.8570	0.9204	0.9319	0.9261
U-Net	0.8793	0.7887	0.9239	0.8537	0.8874
ResNet++	0.6547	0.5020	0.7450	0.6450	0.6914
DeepLabV3+	0.8413	0.7301	0.8547	0.8483	0.8515
HRNet	0.9089	0.8348	0.9368	0.8903	0.9130

The specific calculation method is as follows:

$$\alpha_c = \frac{e^{\alpha_c'}}{e^{\alpha_c'} + e^{\beta_c'} + e^{\gamma_c'}},$$
$$\beta_c = \frac{e^{\beta_c'}}{e^{\alpha_c'} + e^{\beta_c'} + e^{\gamma_c'}}, \quad (3)$$
$$\gamma_c = \frac{e^{\gamma_c'}}{e^{\alpha_c'} + e^{\beta_c'} + e^{\gamma_c'}},$$

where α, β, and γ denote the soft attention vectors, respectively. Note that α_c is the c-th element of α, and is the same for β_c and γ_c.

Finally, the feature maps obtained by different convolution kernels and the obtained attention vectors are multiplied and added to obtain the final output feature map V.

$$V_c = \alpha_c \cdot U' + \beta_c \cdot U'' + \gamma_c \cdot U''', \quad (4)$$

where α, β, and γ need to satisfy $\alpha_c + \beta_c + \lambda_c = 1$, and $V = [V_1, V_2, \cdots, V_c]$, $V_c \in \mathfrak{R}^{H \times W}$.

3. Experiments

3.1. Datasets. In the task of breast tubule segmentation, we conducted experiments on the public BRACS dataset [25] and a private clinical dataset. For each dataset, we randomly selected 10% of images for testing, 10% of images for verification, and the rest 80% of the images for model training.

The first experimental dataset is a private clinical dataset with a total of 398 images, each of which is 2000 × 2000 in size. The tubule masks on each of the images in this dataset were annotated by a pathologist, with white areas representing the breast tubules and black areas indicating irrelevant areas.

The second dataset is an openly accessible dataset named BRACS, which has a total of 4539 H&E images of various sizes. Since the breast tubules on the images of this dataset were not labeled, we did manual annotation of the breast tubules with the help of a pathologist before our experiments.

3.2. Implementation Details. Our experiments ran on the Ubuntu operation system with GPUs of GeForce GTX 1080 [33]. Our model was built using the Keras framework with TensorFlow [34, 35] as the back end. We used the Dice function [36] as the loss function of model training, the optimizer was Adam [37], the learning rate was set to 1e-5, the batch size of the DKS-DoubleU-Net model and those of the other comparable models were set to 4, and the epoch was set to 200.

Our private dataset includes 398 original images with a size of 2000 × 2000 cropped from WSI images and their corresponding ground truths of tubule masks. Since the size of each of the private clinical pictures is 2000 × 2000, to reduce the computational pressure of the graphics processing system [38] and improve the efficiency of code running, we cut each 2000 × 2000 image into 16 random overlapping 512 × 512 tiles, resulting in 6368 tiles in total. Since the cropped images do not change the structures of tubules, it will not affect the actual segmentation effects.

The public BRACS dataset is composed of 4539 H&E images of different image sizes, we resized the images to 512 × 512 for consistency with the 512 × 512 tile size of the private dataset. Since there are no tubule annotations in the public dataset, we generated the labels of the breast tubules beforehand. To speed up the annotation process, we adopted a coarse-to-fine annotation procedure on this dataset.

Specifically, as the first step, we randomly selected 200 H&E images and asked a pathologist to manually annotate the breast tubules using HistoView [39]. Subsequently, we trained the proposed model by using the 200 annotated images and obtained a preliminary model. The preliminary model in our experiment was a DoubleU-Net adopting a deep Residual module as the encoder of the second subnetwork. With the help of this preliminarily trained model, we tested the remaining unlabeled images and treat their outputs as coarse annotations of tubules. Then, the coarse annotations were denoised by applying the "bwareaopen" function in MATLAB to remove small areas [40]. Finally, the refined breast tubule annotations were generated by manually eliminating the false positives and false negatives on the coarse annotation images with the help of the pathologist [41].

3.3. The Objective Function and Evaluation Metrics. To train the proposed DKS-DoubleU-Net model, Dice function is used as the loss function to emphasize the accurate semantic segmentation of the tubule regions. The Dice loss function is

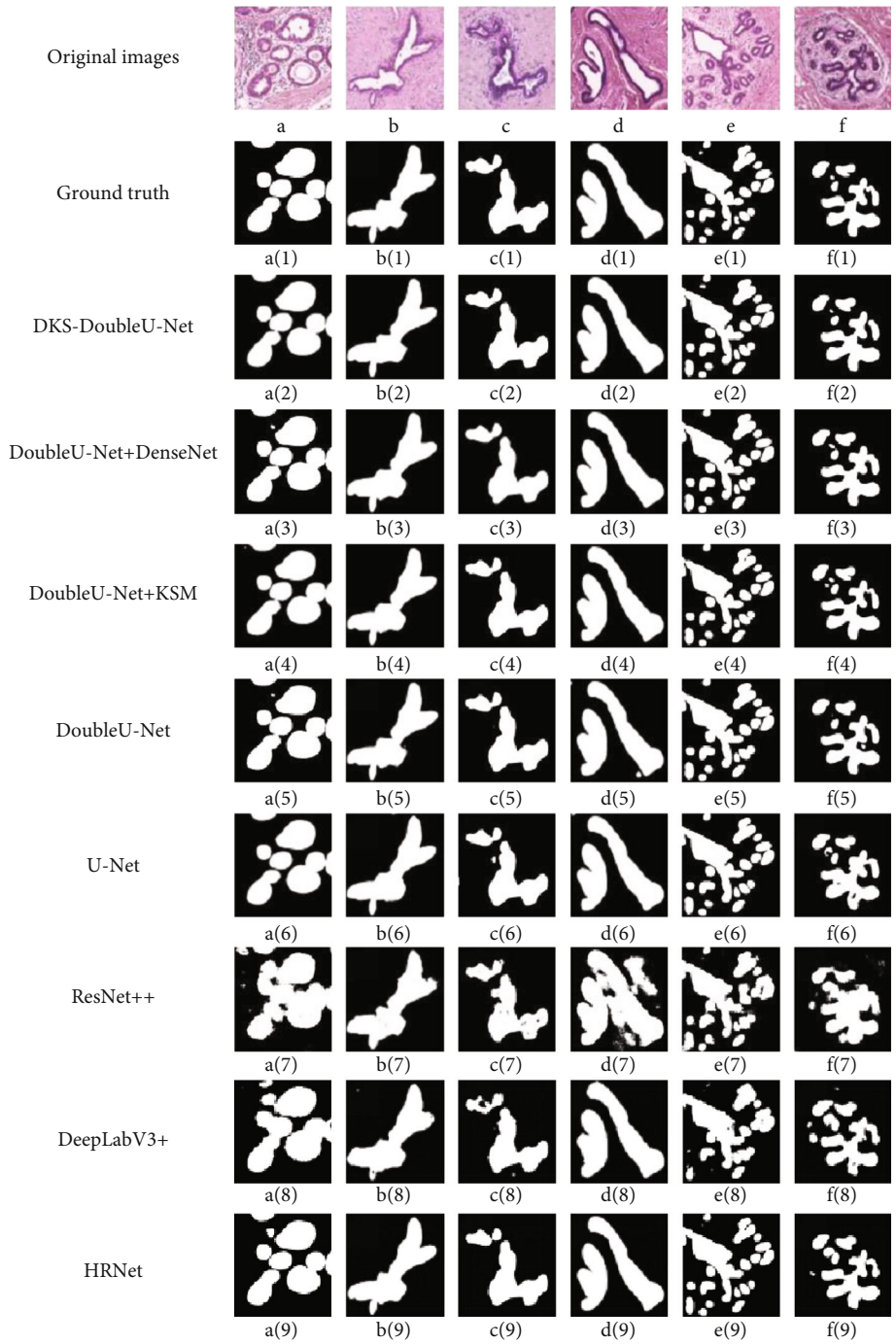


Figure 4: The illustrative tubule segmentation results of different models on the test set of the public BRACS dataset. The first row lists the original images; images a(1)-f(1) in the second row are the corresponding tubule annotation masks; images a(2)-f(2) in the third row are the tubule segmentation results performed by DKS-DoubleU-Net; images a(3)-f(3) in the fourth row are the results output by the DoubleU-Net which added DenseNet; images a(4)-f(4) in the fifth row are the results output by the DoubleU-Net which added Kernel Selecting Module (KSM); images a(5)-f(5) in the sixth row are the results output by the DoubleU-Net; images a(6)-f(6) in the seventh row are the results output by Unet; images a(7)-f(7) correspond to the results of ResUnet++; images a(8)-f(8) are the results obtained by DeepLabV3+, and images a(9)-f(9) show the tubule segmentation results of HRNet.

Table 3: Comparison of the computational complexity of the models.

Model	Trainable params	FLOPs
DKS-DoubleU-Net	127.38 M	254.71 M
DoubleU-Net+DenseNet	127.24 M	254.42 M
DoubleU-Net+KSM	111.88 M	223.69 M
DoubleU-Net	111.73 M	223.40 M
U-Net	98.68 M	197.31 M
ResNet++	15.50 M	30.97 M
DeepLabV3+	155.88 M	311.76 M
HRNet	108.93 M	217.85 M

widely used in medical image segmentation tasks. The definition of Dice Loss is as follows:

$$\text{DiceLoss} = 1 - \frac{2|X \cap Y|}{|X| + |Y|}, \tag{5}$$

where $X \cap Y$ denotes the intersection between X and Y, $|X|$ and $|Y|$ represent the number of elements, respectively, and the numerator is multiplied by 2 to ensure that the value range is between 0 and 1.

To measure the superiority of the proposed model more quantitatively and comprehensively, Dice, mIou, Precision, Recall, and F1-Score, which are the most important and commonly used metrics in semantic segmentation tasks [42–45], served as evaluation metrics in our experiments:

$$\begin{aligned} \text{Dice} &= \frac{2|X \cap Y|}{|X| + |Y|}, \\ \text{mIou} &= \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F1 - \text{Score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \end{aligned} \tag{6}$$

where TP represents true positive, FP denotes false positive, and FN means false negative.

4. Results and Analysis

To evaluate the performance of the proposed DKS-DoubleU-Net model, we conducted experiments on both the private clinical dataset and the public BRACS dataset independently. The experiment results of the proposed DKS-DoubleU-Net model are compared with those of the state-of-the-art segmentation models, including U-Net [8], DoubleU-Net [11], ResUnet++ [9], DeepLabV3+ [10], and HRNet [12] models. To assess the performances of different models objectively, we employ the most commonly used evaluation metrics in semantic segmentation tasks, such as Dice, mIou, Precision, Recall, and F1-Score in our experiments.

4.1. Experiments on the Private Clinical Dataset. The quantitative comparison of the tubule segmentation results obtained by different models on the private clinical dataset is shown in Table 1. The proposed DKS-DoubleU-Net model achieves a Dice of 70.02%, a mIou of 55.47%, a Recall of 76.36%, a Precision of 70.16%, and an F1-Score of 73.13%, which is superior to the original DoubleU-Net model by increasing 2.04%, 2.39%, 1.86%, 1.91%, and 1.89%, respectively. The individual combination of the DenseNet module or KSM module with the original DoubleU-Net also exhibited an improvement in performance as shown in Table 1. This may be attributed to the combined action of the powerful feature extraction ability of DenseNet and the multiscale feature fusion mechanism of KSM. The former could accurately identify the various shapes of tubules and the latter improves the detection accuracy of tubules in different sizes. Moreover, compared to the U-Net model, the proposed DKS-DoubleU-Net has improved the scores of Dice, mIou, Recall, Precision, and F1-Score obtained by 8.87%, 10.28%, 10.68%, 9.96%, and 10.31%, respectively. In summary, the proposed DKS-DoubleU-Net model has achieved the highest scores in Dice, mIou, and F1-Score, except for precision, among all listed state-of-the-art models. Although the model of "DoubleU-Net+KSM" has the highest precision score, the decrease in its recall score leads to a decrease in its F1-Score (i.e., the harmonic mean of Recall and Precision).

To visualize the performances of different models, we demonstrate several tubule semantic segmentation results on the test set of the private clinical dataset, as shown in Figure 3. As can be seen from Figure 3, DKS-DoubleU-Net has the most excellent performance even in segmenting the challenging targets. Specifically, its segmentation results are the closest to the ground truth images and with the fewest noise spots.

4.2. Experiments on the Public BRACS Dataset. To further analyze the performance of breast tubule segmentation by the proposed DKS-DoubleU-Net, we conducted experiments on the public BRACS dataset [25]. The comparative evaluation results of different models on the public BRACS dataset are shown in Table 2.

Table 2 shows that the proposed DKS-DoubleU-Net achieves scores of 92.72% in Dice, 86.55% in mIou, 92.92% in Recall, 93.05% in Precision, and 92.98% in F1-Score. Specifically, after introducing the modules of DenseNet and KSM into the original DoubleU-Net model, the segmentation results have increased by 0.53%, 0.85%, 0.88%, and 0.37% in Dice, mIou, Recall, and F1-Score, respectively. This may be put down to the flexible ability of the introduced DenseNet and KSM modules in segmenting the breast tubules in different sizes and diverse shapes. Compared to the U-Net, the proposed DKS-DoubleU-Net has gained the biggest increase in Dice, mIou, Recall, Precision, and F1-Score by 4.79%, 7.68%, 0.53%, 7.68%, and 4.24%, respectively. On the public BRACS data set, our model fails to reach the highest recall and precision scores, but its F1-Score, resulting from the harmonic mean of recall and precision, is the most superior among all state-of-the-art methods as shown in Table 2.

The illustrative tubule segmentation results obtained by different models on the test set of the public BRACS dataset are shown in Figure 4. It can be seen from Figure 4 that, compared to the state-of-the-art networks, the tubule segmentation results performed by the proposed DKS-DoubleU-Net are more effective in suppressing noisy spots and more accurate in detecting the tubules with various sizes and diverse shapes.

In summary, the proposed DKS-DoubleU-Net model has shown outstanding performance in detecting breast tubules of diverse shapes and different sizes on both the private clinical dataset and the public BRACS dataset.

Specifically, from the quantitative comparison results in Table 1 and Table 2, it can be seen that, among the state-of-the-art models, the proposed DKS-DoubleU-Net model achieves the highest evaluation scores of Dice, mIou, and F1-Score in the tubule segmentation on both datasets. Besides, the compared illustrative tubule segmentation results, as shown in Figures 3 and 4, also demonstrate that the proposed DKS-DoubleU-Net performs best in the accurate breast tubule segmentation, with the results closest to the ground truth among all stat-of-the-art models including DoubleU-Net, U-Net, ResUnet++, DeepLabV3+, and HRNet. Therefore, the introduced modules of DenseNet and Kernel Selecting Module (KSM) in the proposed DKS-DoubleU-Net model could effectively improve the model's ability in complex feature extraction and multiscale feature fusion, which are important in the performance promotion of breast tubule segmentation.

4.3. Computational Complexity Analysis. In order to compare the computational complexity of the models, we measure the computational complexity from the amount of parameters (Paras) and the amount of calculation (FLOPs) as shown in Table 3. Each parameter is a float, that is, a parameter is 4 bytes. And the units of these two metrics are MB. The amount of parameters corresponds to the space complexity of the model, that is, the size of the model. The amount of computation corresponds to the time complexity of the model, that is, the length of the network execution time. It can be seen from Table 3 that the amounts of parameters and calculation of the DKS-DoubleU-Net network are a little bit larger than those of the original DoubleU-Net model, due to the addition of the DenseNet and KSM modules. Although the space complexity and time complexity of the DKS-DoubleU-Net network is a little bit higher, it results in a high performance.

5. Discussion

Breast tubule formation is one of the three essential factors in the clinical routine Nottingham grading system. The assessment of breast tubule formation is critical to the accurate diagnosis and prognosis of breast cancer [2]. In practice, it is a challenging task to accurately segment the breast tubules in H&E images due to their diverse appearance, complex structures, and highly sophisticated morphology. Although deep-learning-based methods have been applied to the segmentation of breast tubules, most of them have not taken into account the reuse of dense features among layers and the fusion of multilevel features for a more accurate semantic segmentation performance of breast tubules. To solve this problem, the proposed DKS-DoubleU-Net model introduced a DenseNet [23] as the encoder module of the second network branch in DoubleU-Net. Besides, the Kernel Selecting Module (KSM) was inserted in front of the outputs of the two U-Net branches in the proposed DKS-DoubleU-Net.

The potential roles of the two modules played in the task of tubule segmentation could be inferred as follows: (1) the DenseNet module, utilizing dense connections between layers through Dense Blocks, would help better extract features from the tubules with complex structures and diverse shapes and (2) the KMS module, enabling a self-adaptive kernel selecting for feature fusion, would facilitate the discovery of tubules of different sizes. Therefore, both qualitative and quantitative experiments have shown that our model achieves competitive performance on both the private clinical dataset and public BRACS dataset. Specifically, in view of quantitative assessment, the DKS-DoubleU-Net produced a Dice of 70.02% and an F1-Score of 73.13% on the private data, which has increased by 2.04% in Dice and 1.89% in F1-Score compared to the original DoubleU-Net. On the public BRACS dataset, our proposed model achieved a Dice of 92.72% and an F1-Score of 92.98%, which outperformed the DoubleU-Net by 0.53% and 0.37%, respectively. Both in terms of quantitative metrics and qualitative visual perception of the tubule segmentation results, our model showed the best performance among the state-of-the-art models, including DoubleU-Net [11], U-Net [8], ResUnet++ [9], DeepLabV3+ [10], and HRNet [12].

The limitation of the DKS-DoubleU-Net is that after the introduction of DenseNet and Kernel Selecting Module (KSM), the network structure becomes more complex and has more parameters to learn, therefore, in our future work, we will focus on optimizing the network structure for more efficient and accurate segmentation performance on diverse tubule regions. As another limitation, our network cannot handle images of random size. In our experiment, we cropped or resize the input image to the size of 512×512. Therefore, to develop an algorithm dealing with input images of random sizes will be one of our further works.

6. Conclusions

In this paper, we proposed a semantic segmentation model termed DKS-DoubleU-Net for accurate semantic segmentation of breast tubules in H&E images. The proposed DKS-DoubleU-Net adopted a DenseNet module as the second encoder of the DoubleU-Net and inserted a Kernel Selecting Module (KSM) before the output layer in each of the two U-Net branches. Our purpose is to discover, reuse, fuse, and propagate the dense features and multilevel features extracted from the coarse breast tubular regions suggested by the first U-Net subnetwork branch of DoubleU-Net. Moreover, in the experiment section, we applied the proposed DKS-DoubleU-Net to two datasets. In terms of both quantitative evaluation and visual perception quality, the

proposed DKS-DoubleU-Net achieved competitive performance in the semantic segmentation of breast tubules in the H&E images on both datasets, as compared with the state-of-the-art models. Therefore, DKS-DoubleU-Net has the potential of being a baseline for breast tubule segmentation in H&E images.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61501287, No. 61971273, No. U2001205, and No. 61877038), the Key Research and Development Program in Shaanxi Province of China (No. 2021GY-032), the State Key Laboratory of Geo-Information Engineering (SKLGIE2019-M-3-5), the Fundamental Research Funds for the Central Universities (No. GK202003077), the Research and Development Project funded by Central University/Public Health and Health Research Project (GK202007033). The authors would like to thank the Xi'an Future Artificial Intelligence Computing Center for providing the artificial intelligence computing power and technical support.

References

[1] M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, and M. F. Clarke, "Prospective identification of tumorigenic breast cancer cells," *Proceedings of the National Academy of Sciences*, vol. 100, no. 7, pp. 3983–3988, 2003.

[2] C. W. Elston and I. O. Ellis, "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histopathology*, vol. 19, no. 5, pp. 403–410, 1991.

[3] I. H. Sarker, "Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective," *SN Computer Science*, vol. 2, no. 3, pp. 1–16, 2021.

[4] J. Liu, Y. Wu, Y. Wu et al., "DAM: discrepancy alignment metric for face recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3814–3823, Montreal, QC, Canada, 2021.

[5] A. Choudhary, L. Tong, Y. Zhu, and M. D. Wang, "Advancing medical imaging informatics by deep learning-based domain adaptation," *Yearbook of Medical Informatics*, vol. 29, no. 1, pp. 129–138, 2020.

[6] M. H. Laves, S. Ihler, T. Ortmaier, and L. A. Kahrs, "Quantifying the uncertainty of deep learning-based computer-aided diagnosis for patient safety," *Current Directions in Biomedical Engineering*, vol. 5, no. 1, pp. 223–226, 2019.

[7] L. Deng, Y. Liu, Y. Shi, and H. Liu, "A deep neural network approach using distributed representations of RNA sequence and structure for identifying binding site of RNA-binding proteins," in *2019 Ieee International Conference on Bioinformatics and Biomedicine (Bibm)*, pp. 12–17, San Diego, CA, USA, 2019.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Cham, 2015.

[9] D. Jha, P. H. Smedsrud, M. A. Riegler et al., "Resunet++: An advanced architecture for medical image segmentation," in *2019 IEEE International Symposium on Multimedia (ISM)*, pp. 225–2255, San Diego, CA, USA, 2019.

[10] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, http://arxiv.org/abs/1706.05587.

[11] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "Doubleu-net: A deep convolutional neural network for medical image segmentation," in *2020 IEEE 33rd international symposium on computer-based medical systems (CBMS)*, pp. 558–564, Rochester, MN, USA, 2020.

[12] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5693–5703, Long Beach, CA, USA, 2019.

[13] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[14] L. Kong and Y. Zhang, "Multi-modal brain tumor Segmentation using cascaded 3D U-Net," in *2021 6th International Conference on Image, Vision and Computing (ICIVC)*, pp. 129–133, Qingdao, China, 2021.

[15] H. Khodajou-Chokami, S. A. Hosseini, and M. R. Ay, "PARS-NET: a novel deep learning framework using parallel residual conventional neural networks for sparse-view CT reconstruction," *Journal of Instrumentation*, vol. 17, no. 2, p. P02011, 2022.

[16] V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," 2018, http://arxiv.org/abs/1801.05746.

[17] X. Guan, G. Yang, J. Ye et al., "3D AGSE-VNet: an automatic brain tumor MRI data segmentation framework," *BMC Medical Imaging*, vol. 22, no. 1, pp. 1–18, 2022.

[18] S. Wang, J. Liu, B. I. Yao-Yao, Z. Chen, Q. H. Zheng, and H. F. Duan, "Automatic recognition of breast gland based on two-step clustering and random Forest," *Computer Science*, vol. 45, pp. 247–252, 2018.

[19] A. Basavanhally, E. Yu, J. Xu et al., "Incorporating domain knowledge for tubule detection in breast histopathology using O'Callaghan neighborhoods," *Medical Imaging 2011: Computer-Aided Diagnosis*, vol. 7963, pp. 305–319, 2011.

[20] P. Maqlin, R. Thamburaj, J. J. Mammen, and A. K. Nagar, "Automatic detection of tubules in breast histopathological images," in *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, pp. 311–321, Springer, India, 2013.

[21] C. Zhang, C. Sun, R. Davey et al., "Tubule detection in testis images using boundary weighting and circular shortest path," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 3319–3322, Osaka, Japan, 2013.

[22] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases," *Journal of Pathology Informatics*, vol. 7, no. 1, p. 29, 2016.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, Honolulu, HI, USA, 2017.

[24] Y. Zhao, Z. Jiang, A. Men, and G. Ju, "Pyramid real image denoising network," in *2019 IEEE visual communications and image processing (VCIP)*, pp. 1–4, Sydney, NSW, Australia, 2019.

[25] N. Brancati, A. M. Anniciello, P. Pati et al., "Bracs: A dataset for breast carcinoma subtyping in h&e histology images," 2021, http://arxiv.org/abs/2111.04740.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, http://arxiv.org/abs/1409.1556.

[27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Salt Lake City, UT, USA, 2018.

[28] M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision*, Academic press, 2020.

[29] F. Handoyo, *Resizing or Scaling Image from Small Image to Enlarge Image Using Bilinear Interpolation Algorithm, Median Filter Algorithm, and Unsharp Masking Algorithm*, Fakultas Ilmu Komputer, 2011.

[30] J. Ge, X. Cui, K. Xiao, C. Zou, Y. Chen, and R. Wei, "BNReLU: combine batch normalization and rectified linear unit to reduce hardware overhead," in *2019 IEEE 13th International Conference on ASIC (ASICON)*, pp. 1–4, Chongqing, China, 2019.

[31] S. Narayan, "The generalized sigmoid activation function: competitive supervised learning," *Information Sciences*, vol. 99, no. 1-2, pp. 69–82, 1997.

[32] M. Wu and B. Liu, "Data hiding in binary image for authentication and annotation," *IEEE Transactions on Multimedia*, vol. 6, no. 4, pp. 528–538, 2004.

[33] D. G. Peng, H. Zhang, L. Yang, and H. Li, "Design and realization of modbus protocol based on embedded linux system," in *2008 International Conference on Embedded Software and Systems Symposia*, pp. 275–280, Chengdu, China, 2008.

[34] K. Ramasubramanian, A. Singh, K. Ramasubramanian, and A. Singh, *Deep Learning Using Keras and Tensorflow*, Machine Learning Using R. Apress, Berkeley, CA, 2019.

[35] K. Jakhar and N. Hooda, "Big data deep learning framework using keras: a case study of pneumonia prediction," in *2018 4th International Conference on computing communication and automation (ICCCA)*, pp. 1–5, Greater Noida, India, 2018.

[36] Q. Huang, J. Sun, H. Ding, X. Wang, and G. Wang, "Robust liver vessel extraction using 3D U-net with variant dice loss function," *Computers in Biology and Medicine*, vol. 101, pp. 153–162, 2018.

[37] D. P. Kingma and B. J. Adam, "A method for stochastic optimization," 2014, http://arxiv.org/abs/1412.6980.

[38] J. Yuan, G. Xu, Y. Yu et al., "Real-time photoacoustic and ultrasound dual-modality imaging system facilitated with graphics processing unit and code parallel optimization," *Journal of Biomedical Optics*, vol. 18, no. 8, article 086001, 2013.

[39] https://engineering.case.edu/centers/ccipd/content/software.

[40] W. Wu and P. Cai, "Simulation of wavelet denoising based on MATLAB," *Information and electronic engineering*, vol. 6, no. 3, pp. 220–222, 2008.

[41] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021.

[42] P. Anilkumar and P. Venugopal, "A Survey on Semantic Segmentation of Aerial Images using Deep Learning Techniques," in *2021 Innovations in Power and Advanced computing technologies (i-PACT)*, pp. 1–7, Kuala Lumpur, Malaysia, 2021.

[43] H. H. Jebamikyous and R. Kashef, "Deep Learning-Based Semantic Segmentation in Autonomous Driving," in *2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on dependability in sensor, Cloud & big Data Systems & application (HPCC/DSS/SmartCity/DependSys)*, pp. 1367–1373, Haikou, Hainan, China, 2021.

[44] L. Yu, Z. Hu, F. Zhang, and K. Yang, "Unmanned aerial vehicle image biological soil crust recognition based on UNet++," *International Journal of Remote Sensing*, vol. 43, no. 7, pp. 2660–2676, 2022.

[45] Y. Cai, H. He, K. Yang et al., "A comparative study of deep learning approaches to rooftop detection in aerial images," *Canadian Journal of Remote Sensing*, vol. 47, no. 3, pp. 413–431, 2021.

Hindawi
BioMed Research International
Volume 2022, Article ID 6256126, 11 pages
https://doi.org/10.1155/2022/6256126

Research Article

Pulse Signal Analysis Based on Deep Learning Network

Quanyu E

Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, 999077 Hong Kong, China

Correspondence should be addressed to Quanyu E; eqy2019@connect.hku.hk

Received 28 May 2022; Revised 21 June 2022; Accepted 5 July 2022; Published 15 September 2022

Academic Editor: Chen Li

Copyright © 2022 Quanyu E. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pulse signal is one of the most important physiological features of human body, which is caused by the cyclical contraction and diastole. It has great research value and broad application prospect in the detection of physiological parameters, the development of medical equipment, and the study of cardiovascular diseases and pulse diagnosis objective. In recent years, with the development of the sensor, measuring and saving of pulse signal has become very convenient. Now the pulse signal feature analysis is a hotspot and difficulty in the signal processing field. Therefore, to realize pulse signal automatic analysis and recognition is vital significance in the aspects of the noninvasive diagnosis and remote monitoring, etc. In this article, we combined the pulse signal feature extraction in time and frequency domain and convolution neural network to analyze the pulse signal. Firstly, a theory of wavelet transform and the ensemble empirical mode decomposition (EEMD) which is gradually developed in recent years have been used to remove the noises in the pulse signal. Moreover, a method of feature point detection based on differential threshold method is proposed which realized the accurate positioning and extraction time-domain values. Finally, a deep learning method based on one-dimensional CNN has been utilized to make the classification of multiple pulse signals in the article. In conclusion, a deep learning method is proposed for the pulse signal classification combined with the feature extraction in time and frequency domain in this article.

1. Introduction

In the past two thousand years, pulse diagnosis, as one of the most important methods of disease diagnosis, has taken a significant part in traditional Chinese medicine. Traditional Chinese medicine believes that changes in the state of internal organs such as the internal organs of the human body or the limbs and facial features will produce a fixed pulse. By touching the pulse of the patient, the internal changes of various organs of the body can be judged, thereby diagnosing the physical condition and the location of the disease. Studies have shown that the pulse signal mainly comes from the heart. Under the periodic pulsation of the heart and the elastic expansion and contraction of the vascular system, blood will flow through various organs of the body [1], and eventually, this blood will be collected on the human wrist. Fingertips, ear tips, and other parts form a pulse signal; therefore, the human pulse signal contains not only heart health disease information but also blood viscosity, blood vessel wall elasticity, blood flow velocity, and other physiological and pathological information related to other human organs [2, 3]. Therefore, whether based on the traditional Chinese medicine perspective or the modern scientific perspective, the study of pulse signals has great theoretical value and practical significance. Figure 1 shows some typical pulse waves in traditional Chinese medicine.

By studying the mechanism of pulse signal generation, pressure pulses and volume pulse signals reflecting the blood pressure or volume information of human blood vessels can be successfully extracted. Compared with the volume pulse signal detection process, the pressure pulse signal detection has many advantages such as easy operation, reliable performance, safety, noninvasiveness, and strong adaptability. Therefore, it has been widely concerned by domestic and foreign researchers and engineers [4, 5]. As shown in Figure 2, the pressure sensor is placed at the superficial artery (such as the radial artery) to detect changes in the blood vessel pressure at the artery to obtain the pulse wave signal. Since the pulsation change curve of blood vessel diameter with time is approximately similar to the pulsation change curve of intravascular pressure, the curve of the change of blood vessel diameter with time detected by various pressure sensors

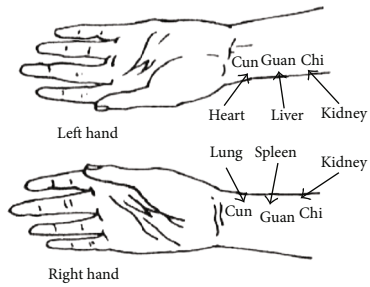


Figure 1: Traditional Chinese medicine pulse diagnosis diagram.

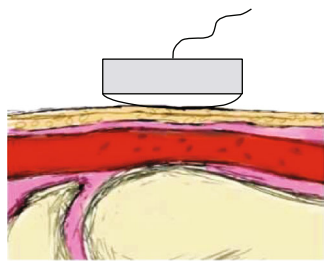


Figure 2: Mechanism of pressure sensor.

clinically reflects the pulsation change of intravascular pressure, thus achieving noninvasive. The pressure sensors that are widely used in piezoelectric, piezoresistive, piezomagnetic (inductive), etc. [6, 7]. The pressure sensor has become the most widely used sensor in the current pulse wave signal acquisition application due to its advantages such as easy manufacture, high power-electricity conversion efficiency, high accuracy, and similar use form to clinical pulse diagnosis [8].

The pulse signal is a nonlinear, nonstationary signal with an approximate period. Initially, it was mainly time-domain analysis, such as analysis through intuitive parameters of the shape, peak value, angle, and area of ascending and descending branches of the pulse waveform, but a lot of practice shows that the information contained in the pulse signal is multifaceted. It is difficult to find the rich information that has involved in the pulse signal by the intuitive analysis method alone, thus promoting the development of the pulse signal analysis method in a more comprehensive and diversified direction. In recent years, the methods that have been commonly used in pulse signal analysis are time domain analysis and time-frequency joint analysis method.

Figure 3 shows the structure diagram of the article. It summarizes the pulse signal analysis in four aspects. This dissertation starts from the collected 10 types of pulse signals and is aimed at analyzing the pulse signal through the EEMD, stationary wavelet transform, and HHT signal denoising methods, deeply excavate a large number of parameters. After the pulse signal analysis process, the preprocessed signal can be gained. Finally, through the CNN, the preprocessed pulse signal can be processed to realize the classification and recognition of various pulse signals.

The purpose of this study is to classify the collected pulse signals through a self-made pulse sensor and to cooperate with traditional Chinese medicine to extract the corresponding characteristics of the pulse through the pulse and to predict and analyze some basic diseases. The novelty of this research is that a set of standard database with 30 kinds of human pulses is established by using the touch training machine of the Faculty of Chinese Medicine of the University of Hong Kong, which can be used for classification and disease prediction.

2. Preprocessing Methods

2.1. Related Work. In 2021, the article in [9] mentioned a method about a multiscale CNN-CRF framework for environmental microorganism image segmentation. This method requires 420 scenes of EM images; however, it is difficult to gain a huge number of EM images. In order to increase the number of the dataset, 105 training images with the revolution of 256×256 pixels have been rotated by 0, 90, 180, and 270 degrees and mirroring and then divided into patches with 8×8 pixels. After this procedure, the 107520 patches have been obtained.

The paper [10] comes up with a state-of-the-art review for gastric histopathology image analysis approaches and future development. This work pointed out that machine learning needs a large number of datasets. The augmentation approaches include rotation, reverse, and scale transform, such as the degree rotation, colour vibrance, adding noise, and the reverse transformation.

2.2. Pulse Signal Data Preprocessing. The collected pulse signal usually contains all kinds of noise interference, which will cause different degrees of signal distortion and affect the quality of pulse signal under test. Therefore, before performing pulse signal analysis and processing, it needs to be preprocessed to obtain a high-quality pulse signal.

The noise interference that affects the quality of pulse signal acquisition mainly includes baseline drift and high-frequency noise, which can usually be filtered by software. The respiration of the human body and the movement of the hands on the period of signal acquisition will lead the baseline drift of the pulse signal, causing the lowest point of the pulse waveform of each cycle to deviate from the same horizontal line. It is a low-frequency signal that changes slowly and is easily mixed in the useful pulse signal [11, 12]; high-frequency noise mainly comes from random noise and environmental interference, such as electromagnetic interference of electronic equipment and thermal noise of electronic devices. These high-frequency or low-frequency noises will affect the accuracy of pulse signal measurement, so it is of great significance to take the necessary measures to filter out noise interference.

2.2.1. Denoising Based on Wavelet Transform. Wavelet transform is a time-frequency localized analysis method based on multiresolution theory. It can study any details of the signal and is an important method for signal time-frequency analysis. The specific decomposition process can be seen in Figure 4. The number of decomposition layers is determined by the analyzed data and the needs of the user [13].

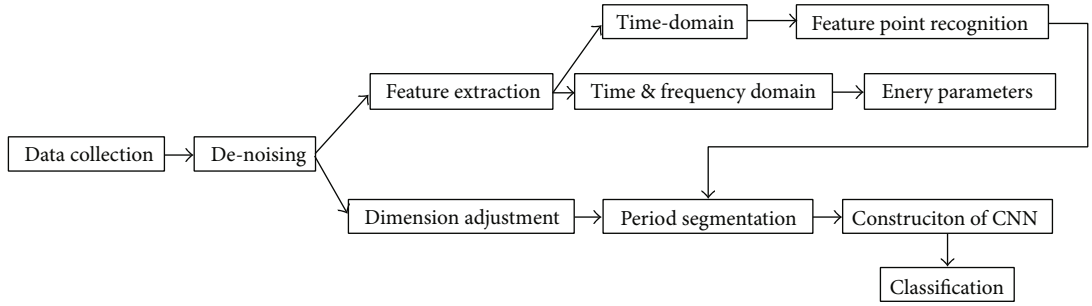


Figure 3: A framework for pulse signal analysis.

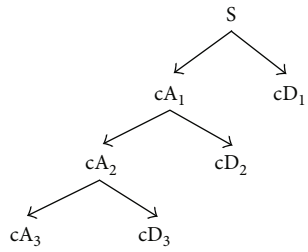


Figure 4: Three-layer wavelet decomposition tree.

The initial signal is decomposed into N layers, and the signal can be expressed as the superposition of the high frequency part and the low frequency part, which can be expressed by the following formula:

$$S = cA_1 + cD_1 = cA_2 + cD_2 + cD_1 = \cdots = cA_n + cD_{n-1} + \cdots + cD_1. \quad (1)$$

Among them, the low-frequency coefficient and the high-frequency coefficient indicate that the frequency range of the signal is different, the wavelet details representing high-frequency noise are mainly distributed in the high-frequency range, and the wavelet details representing baseline drift are mainly distributed in the low-frequency range. By filtering out the wavelet details in the frequency range where the high-frequency noise and baseline drift are located, the filtered pulse signal can be obtained by wavelet reconstruction.

The wavelet details between the maximum decomposition level of a single cycle of a periodic signal and the maximum decomposition level of the entire multicycle signal represent the baseline of the periodic signal [14]. Use db7 as the wavelet basis function and the maximum resolution scale that can be selected for wavelet decomposition of pulse signal is

$$N = \text{floor}|\log_2(L)| = \text{floor}|\log_2(1000)| = 9, \quad (2)$$

where L stands for the length of the pulse wave signal and floor represents the rounding down operation.

To calculate the maximum number of decomposition layers of a pulse signal in a single period, it is vitally important to calculate the number M of pulse waves in a piece of pulse data, which can be obtained by the differential threshold method or the Fourier transform method. The maximum decomposition level of a single-cycle signal is

$$N_1 = \text{Ceil}\left|\log_2\left(\frac{L}{M}\right)\right| = \text{Ceil}\left|\log_2\left(\frac{1000}{7}\right)\right| = 8. \quad (3)$$

Among them, ceil means rounding up. After calculating N and $N1$, the number of decomposition layers where the baseline drift is located can be determined. At the same time, the pulse signal is mainly distributed within 20 Hz. Therefore, by filtering out high-frequency noise greater than 20 Hz and low-frequency baseline drift noise, you can get a more ideal pulse signal.

From Table 1, we can know that the noise greater than 20 Hz is mainly distributed in the wavelet details of layer 1 and layer 2. Setting the high-frequency wavelet coefficients and low-frequency wavelet details to zero, that is cD_1, cD_2, cD_8, cD_9 and cA_9 in Table 1. Then, reconstruct the signal to gain the denoised signal. The original signal, reconstructed signal, and error curve are shown in Figure 5.

2.2.2. Denoising Using HHT Algorithm Based on EEMD. Hilbert-Huang transform (HHT) is another important time-frequency analysis way, which mainly includes two aspects, EEMD decomposition and Hilbert spectrum analysis. By performing the Hilbert transform on the IMF component gained by EMD decomposition, the instantaneous frequency and instantaneous amplitude of the signal will be gained [15].

The decomposition process of EEMD method [16]: add the mean value to M times ($M > 1$) in the initial signal $x(t)$, and the standard deviation of the amplitude is constant Gaussian white noise $n_i(t)(i = 1 \sim M)$, namely,

$$x_i(t) = x(t) + n_{i(t)}. \quad (4)$$

EMD decomposition is performed on $x_i(t)$ to obtain K IMF components denoted as $c_i(t)(j = 1 \sim KK)$ and remainder $f_i(t)$, where $c_i(t)$ represents the j-th IMF component gained when the Gaussian white noise is added for the i times. Utilizing the theory that the statistical mean value of the uncorrelated random sequence is zero, the IMF component $c_i(t)$ which related to the above step is averaged to offset the

Table 1: Band range of wavelet coefficients of each layer decomposed by 9-level wavelets.

Wavelet coefficients	cD_1	cD_2	cD_3	cD_4	cD_5	cD_6	cD_7	cD_8	cD_9	cA_1
Hz	50-100	25-50	12.5-25	6.3-12.5	3.1-6.3	1.6-3.1	0.8-1.6	0.4-0.8	0.2-0.4	0-0.2

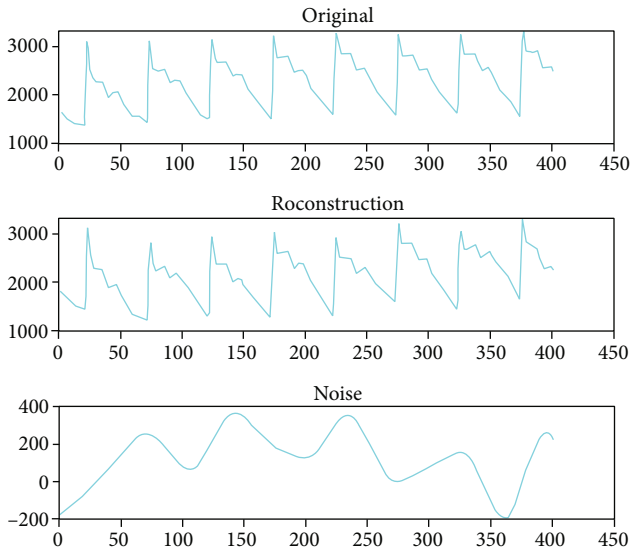


Figure 5: Denoising result based on wavelet transform.

influence of adding Gaussian white noise in the real IMF multiple times. The IMF is

$$c_j(t) = \frac{1}{M}\sum_{i=1}^{N} c_{ij}(t), \tag{5}$$

where $c_i(t)$ stands for the j-th IMF gained by EEMD decomposition of the initial signal [17] and M refers to the number of added white noise sequences.

Hilbert transform of the time domain signal$x(t)$is defined as

$$\widehat{x}(t) = H|x(t)| = \frac{1}{\pi}\sum_{-\infty}^{+\infty} x(\tau)\frac{1}{t-\tau}d\tau. \tag{6}$$

As we all know from the formula above that the Hilbert transform of $x(t)$ can also be expressed as a convolution form of $x(t)$ and $(\pi\tau)^{-1}$, namely,

$$\widehat{x}(t) = \frac{1}{\pi\tau} \times x(t). \tag{7}$$

Fourier transform of the two sides of the formula is as follows:

$$F|\widehat{x}(t)| = \frac{1}{\pi}F\left|\frac{1}{\tau}\right| f|x(t)| = jsgh(f)\cdot F|x(x)|. \tag{8}$$

The Hilbert transform of the signal $x(t)$ will be gained by inverse Fourier transform of the formula.

Hilbert transform of the IMF component is as follows:

$$\widehat{c_i}(t) = H|c_i(t)| = \frac{1}{\pi}\sum_{-\infty}^{+\infty} c_i(\tau)\frac{1}{t-\tau}d\tau. \tag{9}$$

According to the characteristics of the Hilbert transform, $c_i(t)$ can be regarded as the real part, and $c_i(t)$ can be regarded as the imaginary part, and the two parts can combine as the analytical signal $z_i(t)$:

$$z_i(t) = c_i(t) + j\widehat{c_i}(t) = a(t)e^{j\theta(i)}. \tag{10}$$

Filter the IMF component which instantaneous frequency < 0.5 Hz and >20 Hz, and combine the other IMF component to get the denoising signal. The whole debaseline drift process is shown in Figure 6.

3. Feature Extraction Methods of Pulse Signal

Due to the superficial position of the radial artery, the detection is the most convenient. It is the same as the clinical pulse diagnosis position of traditional Chinese medicine, and the radial artery is closest to the peripheral blood vessel. At present, the pulse wave is mainly measured from the radial artery. According to the mechanism of pulse wave formation, it can be divided into different heartbeat stages. The complete pulse wave waveform under normal physiological conditions is shown in Figure 7.

A complete pulse wave waveform consists of ascending branches and descending branches [18], which mainly includes 5 feature points. The meaning of each feature point is as follows.

Point b, the opening point of the aorta, which is the starting point of a pulse wave period, is the valley point of the pulse wave pattern, which marks the left ventricle contraction. Point c is the main wave, which is the maximum point in the pulse wave waveform. Point e, the tidal wave, also known as the front wave of the weighting wave, is located on the descending branch of the waveform, generally after the main wave. Point f, the through valley, is the dividing point between systole and diastole, located before the dicrotic wave and after the tidal wave. Point g is the striking wave crest, and it is a protruding wavelet, located after point f.

3.1. Pulse Wave Feature Point Recognition. For the sake of recognizing a total of three feature points from the aortic opening point (point b), the main wave peak (point b), the tidal wave (point e), and the trough valley (point f) in the pulse wave waveform, this paper uses a pulse wave feature point automatic recognition algorithm.

The characteristics of the pulse waveform are mainly represented by the extreme points and inflection points of the waveform. Therefore, the use of the differential

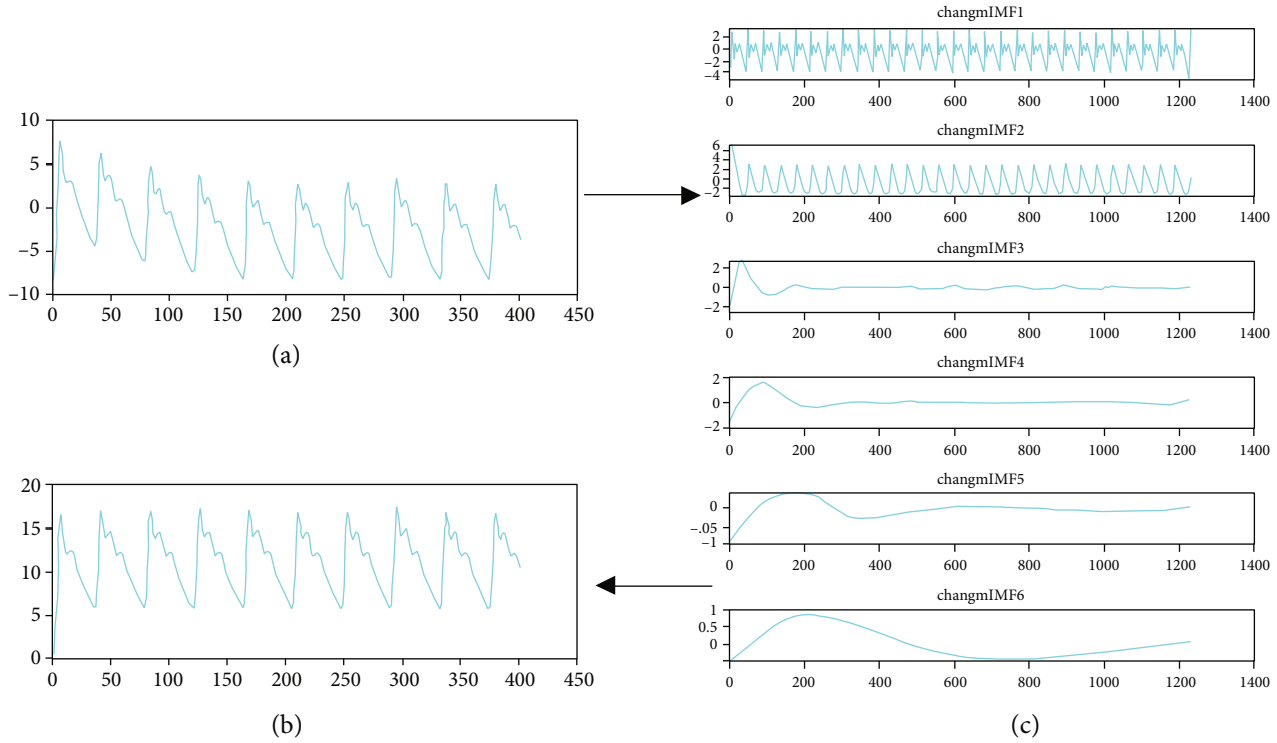


Figure 6: The debaseline drift of pulse signal: (a) original pulse wave; (b) denoised pulse wave; (c) each IMF component.

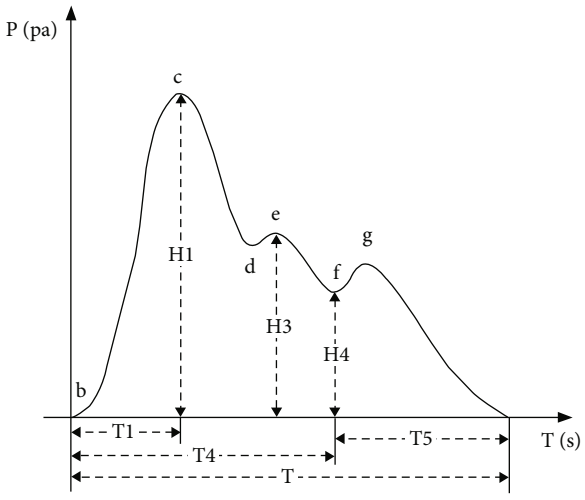


Figure 7: Typical pulse waveform.

threshold method can easily extract the pulse characteristic parameters.

The first-order and second-order differential processing of the pulse signal can obtain the maximum point of the signal. Since the height of the main wave of the pulse signal is the maximum point in a cycle, set the maximum value of the signal greater than m times by adjusting the value of parameter m which can get better recognition effect. In the same way, when the height of the dicrotic wave is set to be greater than n times and less than m times the maximum value, a better recognition effect can also be obtained.

Through the transformation law of the first-order and second-order differential signals, the minimum point of the signal is found. When the maximum figure of the signal with the height of the lowering gorge is greater than k times, a better feature parameter extraction effect can be obtained. At the same time, by observing the collected pulse signal, we can know that the pulse signal is distributed around the minimum point of the starting point. When the differential threshold method is used to extract the position information of the point, it is easy to introduce confusion and cannot accurately distinguish the pulse signal feature points. When the first-order differential value at this point is required to be greater than a certain value a, better results can be obtained. The result diagram for discovering feature points based on differential threshold method is shown in Figure 8.

3.2. Characteristic Parameters in Time and Frequency Domain. Through wavelet transform, not only the denoising of the pulse signal can be achieved but also the characteristic parameters of the waveform in the time or frequency domain can be extracted. As a result, it has been commonly concerned by researchers. Shi et al. pointed out a signal feature detection way through wavelet transform [19], and simulation proved that the method has a higher detection accuracy and a more accurate positioning effect. Hongbiao proposed a feature extraction method suitable for nonstationary pulse signals [20]. Use wavelet decomposition to realize the classification and identification of healthy people and patients with atherosclerosis.

Perform 5-layer decomposition on samples 1, 2, and 3 to obtain decomposition coefficients cA_5cD_5, cD_4, cD_3, cD_2, and cD_1. According to the law of wavelet decomposition,

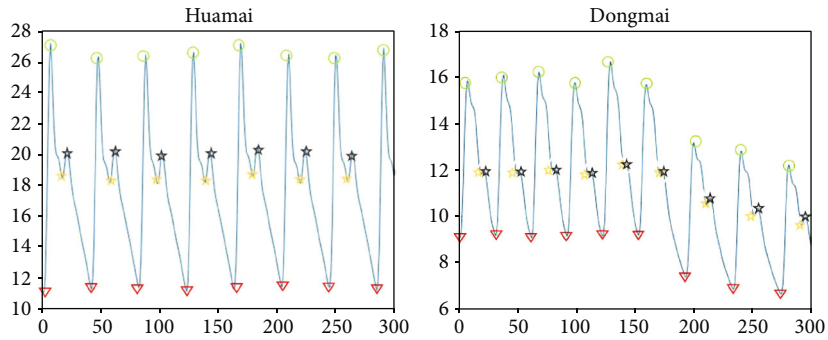


Figure 8: Result diagram for discovering feature points based on differential threshold method.

Table 2: Normalized eigenvector T_0 of the wavelet energy.

	E_{cA5}	E_{cD5}	E_{cD4}	E_{cD3}	E_{cD2}	E_{cD1}
Sample 1	0.9584	0.0386	0.0028	0.0002	0.0000	0.0000
Sample 2	0.9273	0.0641	0.0084	0.0002	0.0000	0.0000
Sample 3	0.9832	0.0134	0.0034	0.0000	0.0000	0.0000
Sample 4	0.9752	0.0201	0.0046	0.0001	0.0000	0.0000
Sample 5	0.9943	0.0048	0.0008	0.0001	0.0000	0.0000
Sample 6	0.9813	0.0162	0.0024	0.0001	0.0000	0.0000

the frequency ranges corresponding to cA_5cD_5, cD_4, cD_3, cD_2, and cD_1 are 0-3.13, 3.13-6.25, 6.25-12.5, 12.5-25, 25-50, and 50-100 Hz. Since information greater than 20 Hz is filtered out during preprocessing, the amplitudes of cD_2 and cD_1 are close to zero. At the same time, we can know from the wavelet decomposition diagram that the amplitude of the wavelet coefficients of low-frequency information is large, and the amplitude of the wavelet coefficients of high-frequency information is small.

Then, according to the relationship between the wavelet coefficients and the signal energy distribution, the square sum of the wavelet coefficients of each layer is calculated. We can get the feature vector T:

$$T = [E_{cA5}, E_{cD5}, E_{cD4}, E_{cD3}, E_{cD2}, E_{cD1}]. \quad (11)$$

Furthermore, the normalized feature vector TT_0 can be obtained:

$$T_0 = \frac{[E_{cA5}, E_{cD5}, E_{cD4}, E_{cD3}, E_{cD2}, E_{cD1}]}{E_0}. \quad (12)$$

We can see from Table 2 that the energy ratios of data 1-6 at 0-6.25 Hz are all higher than 99%, indicating that the collected pulse signals are all from healthy people. At the same time, using statistical methods to analyze the wavelet coefficients in each layer in detail, we can study their relationship with specific diseases and promote the objective development of pulse diagnosis.

4. Experiments and Results

4.1. Experiment Setting

4.1.1. Pulse Signal Dataset. The dataset consists of the raw pulse data collected by the homemade pulse sensing device of Lee Ultrasound Lab from the palpation training machine in Traditional Chinese Medicine School of The University of Hong Kong. There are 10 types of pulse waves (or Mais) which contain in the dataset. The names of those Mais are as follows: Changmai, Chimai, Dongmai, Huamai, Huanmai, Jimai, Pingmai, Shimai, Weimai, and Xuanmai. Every type of Mais was collected in 3 different applied pressures. There is no quantitative standard for the applied pressure of the palpation training machine, nor the quantitative math model for the applied pressure and the pulse signal. Therefore, the 3 applied pressures were not unified cross different Mais. When no extra pressure forced on the sensor, there is still an initial pressure of the sensor. The initial pressure was recalled as the baseline. It was caused by the produce process that needs to be subtracted from the collected signal. Each data signal contains two types of variables. The x axis (the horizontal axis) represents the pressure sample points which units is "Pa." The y axis (the vertical axis) represents the time points when collecting the corresponding pressure points.

4.1.2. Experimental Environment. The preprocessing part of the experiment in this paper is implemented on MATLAB R2019a, and the deep learning part is implemented with Python 3.7. The deep learning uses CNN network. The CNN model is built and trained on TensorFlow on an 8 GB memory computer.

The experiment runs on macOS 10.15.

4.2. Training and Test Data Setting. Preprocess the collected signals in the above process to denoise and solve the problem

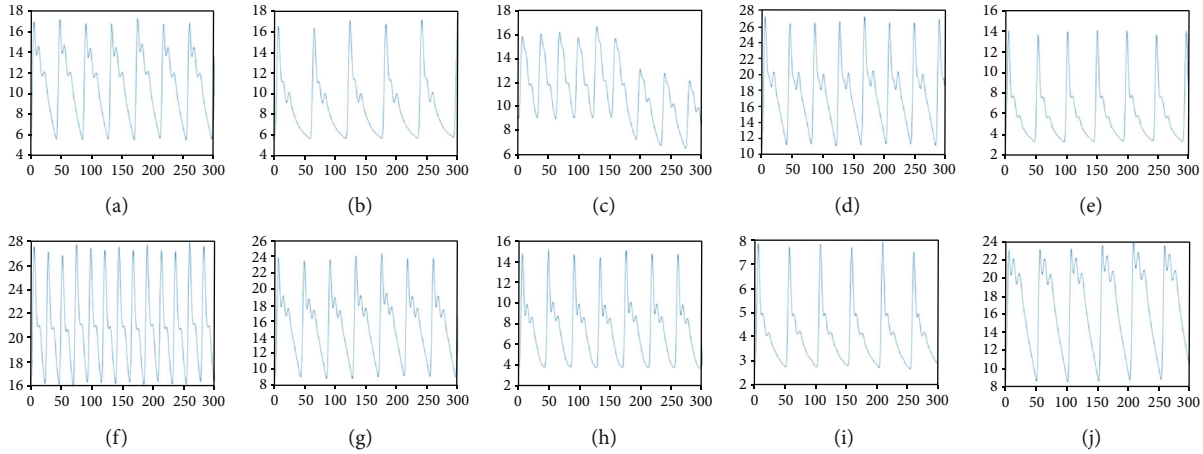


Figure 9: Database of the pulse wave: (a) Changmai; (b) Chimai; (c) Dongmai; (d) Huamai; (e) Huanmai; (f) Jimai; (g) Pingmai; (h) Shimai; (i) Weimai; (j) Xuanmai.

of baseline drift. Here, the wavelet transform method which has been mentioned in Chapter 2.2.1 is used for preprocessing.

4.2.1. Pulse Data Dimension Adjustment. During signal acquisition, the saved data variables are time and pressure, because the hardware reading is not absolutely uniform, resulting in different time intervals. When performing deep learning on data, it is necessary to ensure that each input data has the same length, so cubic spline interpolation is performed on the original data to expand the original data volume tenfold. After that, resampled the interpolated data; that is, 50 time slices (50 points per second) are intercepted every second, thereby ensuring that the time interval of the sampling points is uniform.

4.2.2. Pulse Segmentation and Augmentation. Because the collected signals have different lengths, the signals must be segmented. Use the feature points of the pulse signal which has been mentioned in Chapter 2.2.1 to find the lowest point of the pulse signal (that is, point b, the opening point of the aorta), use it as the starting point of the signal, and cut the resampled data. Intercept the pulse signal with a length of 6 seconds (that is, 300 time slices). At the same time, due to insufficient data volume and in order to not waste data, prevent the overfitting phenomenon, and better describe each type of pulse type, we cross-intercept the pulse signal by intercepting a segmentation of pulse signal in every other feature point (point b) to expand the dataset.

Through the operations above, 677 sets of pulse data have been obtained in the experiment: Changmai (59), Chimai (49), Dongmai (79), Huamai (63), Huanmai (63), Jimai (130), Pingmai (54), Shimai (60), Weimai (50), and Xuanmai (70). As we can see from Figure 9, it is the database of the pulse signal for each pulse wave.

4.2.3. Arrangement of Datasets. The collected 10 types of pulse signals have been divided into 677 sets of data. And then, the dataset is randomly allocated into training set and test set by the ratio of 7 : 3. As a result, the training set contains 473 sets of data, and the test set involves 204 sets of data. The arrangement of the training and test datasets can be seen in Table 3.

Table 3: The arrangement of training and test datasets.

Dataset/class	Train	Test	Total
Changmai	43	16	59
Chimai	31	18	49
Dongmai	55	24	79
Huamai	47	16	63
Huanmai	42	21	63
Jimai	88	42	130
Pingmai	38	16	54
Shimai	45	15	60
Weimai	34	16	50
Xuanmai	50	20	70
Total	473	204	677

4.3. Related Work. In [19], the authors mentioned a multiscale CNN-CRF framework for environmental. In this work, a novel method called mU-Net-BXs is utilized to optimize the adaptability of U-Net convolution neural network. The original inception using different size convolution filters of 1×1, 3×3, and 5×5 (called BOLCK-I) is the direct way to optimize the adaptability of U-Net. Inspired by Inception V2 and Inception V3, a 5×5 convolution filter can be replaced by a sequence of two 3×3 convolution filter, and a 7×7 convolution filter actually resembles a sequence of three 3×3 convolution filters (called BLOCK-II). In the same way, a 3×3 convolution filter can also be replaced by a sequence of 1×3 and 3×1 convolution filters as BLOCK-III. And then, deploy the BLOCK-III in mU-Net architecture. The whole architecture of the network U-Net convolution neural network is shown in Figure 10.

In [21], the authors proposed an approach of the identification of COVID-19 samples from X-ray images using

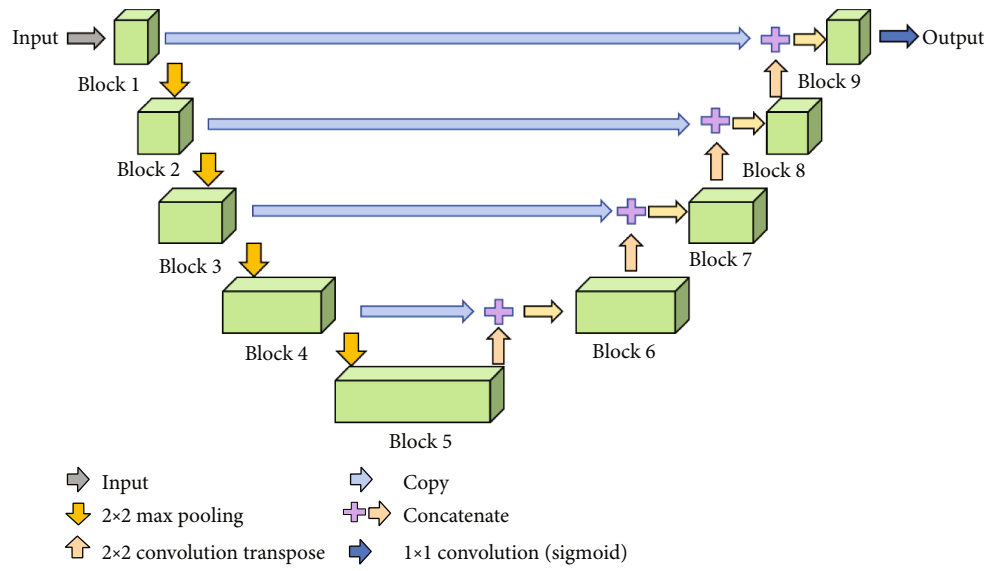


Figure 10: The architecture of mU-Net.

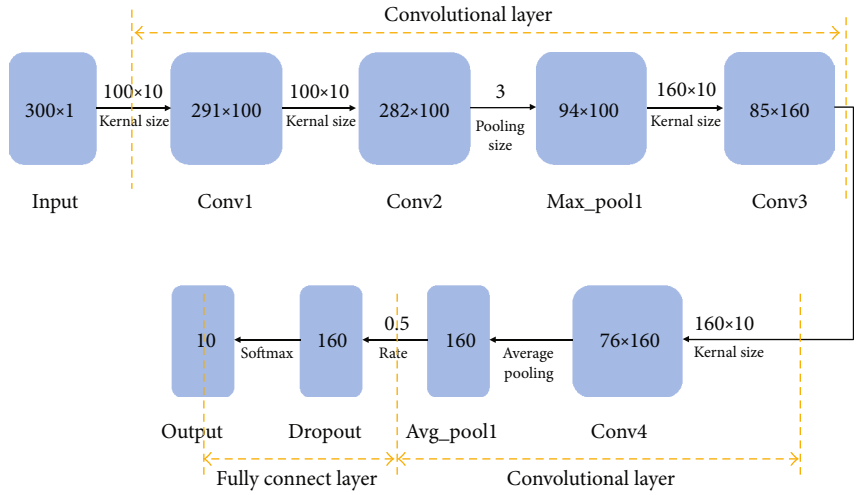


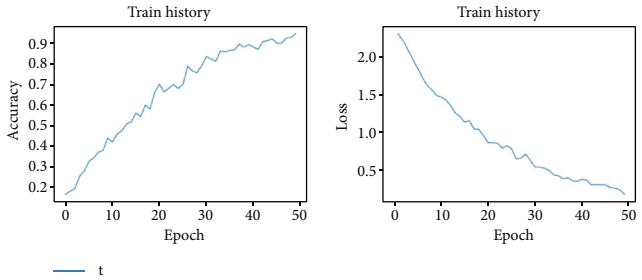
Figure 11: CNN model framework based on pulse classification.

deep learning in which collected COVID CXR images from the GitHub repository with various sizes of pixels and kept target size to the same pixel. Then, apply VGG series, Xception, ResNet V1 and ResNet V2 series, Inception series, DenseNet series, and MobileNet networks in the transfer learning process, where the weights are pretrained on ImageNet dataset.

4.4. Construction of the Classification of the Pulse Signal Algorithm Based on the CNN Model. The constructed convolutional neural network consists of two parts, the convolutional layer and the fully connected layer. The convolutional layer includes four convolutional layers and two pooling layers. In the connection layer, the selected activation function is the softmax function. The loss function selects the categorical_crossentropy function suitable for multiclassification, and the metric selects the commonly used accuracy.

The whole model framework is shown in Figure 11; it contains 2 parts. For input data, the data has been preprocessed, and each data record contains 300 time slices (the pulse wave data has been divided into 6 seconds pulse wave; each second includes 50 time slices). In each time slices, three values for the left horizontal axis, right horizontal axis, and vertical axis will be saved. This will produce a 300×1 matrix. The data should be put into the model with a size of 300. The first layer of the network should be embedding the data to its initial size, that is, 300×1.

The first convolutional layer: the first layer can be seen as a filter (or feature detector) with a size of 10 (also named the kernel size). However, only one filter is not enough for the model to learn all the features in the first convolution layer, so we select 100 feature detectors. This makes us to train 100 different functions at the first layer of the model. The result of this CNN layer is a 291×100 matrix. Each column of the output matrix contains a filter



(a) (b)

Figure 12: Result of the test set of the training model: (a) accuracy of the model in 50 times; (b) loss rate of the model in 50 times.

Table 4: The evaluation metrics for pulse signal classification.

Metric	Definition	Metric	Definition
Precision	$\text{Precision} = \frac{TP}{TP + FP}$	Recall	$\text{Re call} = \frac{TP}{TP + FN}$
F1-score	$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$		

weight. Using the defined kernel size and considering the length of the input matrix, each feature detector involves 291 weights.

Second convolutional layer: results from the last layer should put into the second convolution layer. We can select 100 different feature vectors to train at this layer as well. Having the same steps as the last layer, the size of the result will be 282×100.

Maximum pool layer: for the sake of reducing the intricacy of the result and avoiding the data of overfitting, the pool layer is usually used after the CNN layer. In this study, we select 3 sizes. As a result, the input matrix's size is 3 times larger than the output matrix.

Third and fourth convolution layers: for the sake of the model to have higher level of functions, the next is another sequence of 2 convolution layers. However, the length of the feature detector is still 10; we select 160 feature detectors. As a result, the output matrix after these two layers is a 76×160 matrix.

Average pooling layer: the reason for this layer is make one more layer of pooling to further avoid overfitting. Meanwhile, instead of taking the max value, the average of the two weights in the CNN model is taken. The size of the output matrix is 1×160 neuron matrix. Each feature detector has only one weight left in this layer of neural network.

Dropout layer: the dropout layer will randomly assign 0 weights to the neurons in the network. When the ratio we select is 0.5, the weight of 50% of neurons is zero. After this step, the CNN model will become less sensitive to small changes of data. Therefore, it should further improve the accuracy of our invisible data. The result of this layer is still a 1×160 neuron matrix.

Fully connected layer: the model has to predict 10 categories ("Changmai," "Chimai," "Dongmai," "Huamai," "Huanmai," "Jimai," "Pingmai," "Shimai," "Weimai," "Xuanmai"). The activation function we select is softmax.

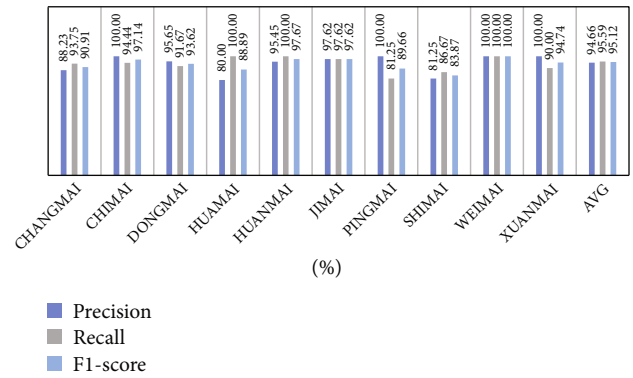


Figure 13: Classification result of different pulse wave based on the CNN model.

It makes the sum of the total 10 results of the CNN model to be 1. Therefore, the result will stand for the probability of each of the ten types.

4.5. Analysis of the Experimental Results. Input the processed data into the built CNN model for training and testing. The training times for the training set are set to 50 times (epoch = 50). Figure 12 shows the accuracy and loss rate of the training set during the training process. As the information shown in Figure 12, when the training time is more than 40 times, the accuracy rate gradually stabilizes at around 90%. The final accuracy rate stabilizes at around 94%, and the loss rate decreases to within 0.2 with the increase of the training times.

Then, put the test set data into the trained CNN model, and the accuracy rate of the test set is 94.12%. For the training result, we select precision, recall, and F1 score those three evaluation parameters to evaluate the classifier performance. The equation of the three parameters is shown in Table 4.

Figure 13 shows that the evaluation parameters of the test set. From Figure 13, we know that the accuracy value, recall rate, and F1 score of each group are all above 80%, the individual pulse wave accuracy rate can reach 100%, and the average recall rate and F1 score are above 95%; the average precision rate is above 94%. The average of the three evaluation parameters is not much different. In conclusion, the model has good performance in accuracy and stability.

The test set contains a total of 204 sets of data, which are allocated to the training set according to 7 : 3. For the sake of enhancing the universality of the results and reducing the chance, in the process of assigning the training set and the validation set, we did not choose to distribute each pulse in equal proportion according to the 10 types of pulses but chose to do all the data randomly assigned. However, due to the small sample size, it may result in less pulse wave data of a certain type in the test set, which may lead to a lower performance in the evaluation parameters of the results, but they are all stable at more than 80%. If you can expand the data capacity or increase the number of training times, this phenomenon can also be greatly improved.

204/1 – 1s – loss: 0.1886 – accuracy: 0.9412										
Predict Label	0	1	2	3	4	5	6	7	8	9
0	15	0	0	0	0	0	0	1	0	0
1	0	17	0	0	0	1	0	0	0	0
2	0	0	22	0	0	0	0	2	0	0
3	0	0	0	16	0	0	0	0	0	0
4	0	0	0	0	21	0	0	0	0	0
5	1	0	0	0	0	41	0	0	0	0
6	1	0	0	2	0	0	13	0	0	0
7	0	0	1	0	1	0	0	13	0	0
8	0	0	0	0	0	0	0	0	16	0
9	0	0	0	2	0	0	0	0	0	18

Figure 14: Confusion matrix of the CNN model (test set).

Compared with the evaluation parameters, the confusion matrix can more intuitively see the classification results of each type of pulse wave.

The confusion matrix is an indicator for judging the results of the model and is part of the model evaluation indicators. The confusion matrix is mostly used to indicate the quality of the classifier. The confusion matrix is also called the error matrix [22]. In the error matrix, each row of the matrix stands for the class prediction of the data by the classifier, and each column of the matrix represents the true category to which the version belongs [23]. The reason why it is called the confusion matrix is that it can intuitively see whether the results of the samples are confused in the results of the machine learning output.

From the diagonal of the confusion matrix in Figure 14, it can intuitively see the number of pulse prediction pairs of each type and the result of erroneous prediction. It can adjust the deep learning model or adjust the feature extraction according to the part of the result set that is wrongly predicted. Evaluation parameters can also be calculated based on the confusion matrix.

5. Conclusions

This research collects pulse signals through self-made sensors. Then, the pulse signal is subjected to wavelet transform to denoise. Analyze the pulse signal in time frequencies and perform feature extraction. The time frequencies include the extraction of pulse signal feature points, the energy extraction of pulse wavelet coefficients, and the energy extraction of the HHT transform. Hope to use statistical methods to analyze the wavelet coefficients in each layer in detail, we can study their relationship with specific diseases and promote the objective development of pulse diagnosis. In the signal acquisition, a total of ten different pulse signals are collected, and this data is used to classify the pulse signals based on deep learning. The classification model selects the CNN model. The initial signal is classified by signal segmentation and labeling. Build a one-dimensional CNN model, utilize the model to train the data, and then get a pulse signal classifier with an accuracy of about 95%.

In the future, we consider combining a CNN module and a VT module proposed in [24] for local and global feature extraction, respectively, to improve the model performance. Meanwhile, the strategy used in LCU-net [25] is considered to optimize the model, where the dense CRF is applied to reduce the memory cost. Besides, we can use MRF and CRF methods in [26] to mask the annotations of the various groups of imaging functions and regions of interest with the data augmentation. Furthermore, we can also apply the GasHisTransformer [27] to capture long-range correlation considering the global and local associations of the pulse signal in a unified context.

Data Availability

The data used to support the findings of this study are available from the first author upon request.

Conflicts of Interest

There are no conflicts of interest.

Acknowledgments

The author would like to express sincere thanks to Dr. Yuchen Tang for the previous cooperation in this work.

References

[1] Y. Chen, L. Zhang, D. Zhang, and D. Zhang, "Wrist pulse signal diagnosis using modified Gaussian models and fuzzy C-means classification," *Medical Engineering and Physics*, vol. 31, no. 10, article S1350453309001830, pp. 1283–1289, 2009.

[2] Y. Chen, L. Zhang, D. Zhang, and D. Zhang, "Computerized wrist pulse signal diagnosis using modified auto-regressive models," *Journal of Medical Systems*, vol. 35, no. 3, pp. 321–328, 2011.

[3] S. Xianghe and Y. Wang, "Pulse wave: the bridge connecting traditional Chinese medicine with western medicine," *Journal of Integrated Traditional Chinese and Western Medicine*, vol. 6, no. 9, pp. 891–896, 2008.

[4] Z. Lieliang, J. Zhu, and L. Xu, "Progress in clinical application research of photoelectric volume pulse wave," *Journal of Clinical Anesthesiology*, vol. 29, no. 11, 2013.

[5] H. Liu, Y. Wang, and L. Wang, "A review of non-contact, low-cost physiological information measurement based on photoplethysmographic imaging," in *In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2088–2091, San Diego, CA, USA, 2012, August.

[6] T. Weichang, L. Congying, and F. Gairong, "Research on ZM-300 intelligent pulse analyzer," *Proceedings of the 12th Annual Conference on Chinese Medicine Diagnosis*, pp. 26–29, 2011.

[7] E. Kaniusas, H. Pfutzner, L. Mehnen et al., "Method for continuous nondisturbing monitoring of blood pressure by magnetoelastic skin curvature sensor and ECG," *IEEE Sensors Journal*, vol. 6, no. 3, pp. 819–828, 2006.

[8] Y. Zhang, C. Wang, and J. Cui, *Advance in the Application Research of PVDF Piezoelectric Film*, Engineering Plastics Application, 2003.

[9] J. Zhang, C. Li, F. Kulwa et al., "A multiscale CNN-CRF framework for environmental microorganism image segmentation," *BioMed Research International*, vol. 2020, Article ID 4621403, 27 pages, 2020.

[10] S. Ai, C. Li, X. Li et al., "A state-of-the-art review for gastric histopathology image analysis approaches and future

development," *BioMed Research International*, vol. 2021, Article ID 6671417, 19 pages, 2021.

[11] W. H. Lin, H. Wang, O. W. Samuel, G. Liu, Z. Huang, and G. Li, "New photoplethysmogram indicators for improving cuffless and continuous blood pressure estimation accuracy," *Physiological Measurement*, vol. 39, no. 2, article 025005, 2018.

[12] C. Fischer, B. Domer, T. Wibmer, and T. Penzel, "An algorithm for real-time pulse waveform segmentation and artifact detection in photoplethysmograms," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, pp. 372–381, 2017.

[13] D. Ni, F. Hu, and C. Yue, "Image retrieval technology based on wavelet transform," *Image Coding and Software*, vol. 25, no. 12, 2012.

[14] H. Qingyang, L. Bingyu, and X. Wang, "Simultaneous elimination of pulse wave signal highfrequency noise and respiratory baseline based on wavelet transform," *Beijing Biomedical Engineering*, vol. 33, no. 3, pp. 247–253, 2014.

[15] Y. Jia, G. Li, X. Dong, and K. He, "A novel denoising method for vibration signal of hob spindle based on EEMD and grey theory," *Measurement*, vol. 169, p. 108490, 2021.

[16] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: the Hilbert spectrum," *Annual Review of Fluid Mechanics*, vol. 31, pp. 417–457, 1999.

[17] S. Zhang, L. Zhi, and T. Zhou, "Medical image retrieval using empirical mode decomposition with deep convolutional neural network," *BioMed Research International*, vol. 2020, Article ID 6687733, 12 pages, 2020.

[18] R. Sahni, "Noninvasive monitoring by photoplethysmography," *Clinics in Perinatology*, vol. 39, no. 3, pp. 573–583, 2012.

[19] Z. Shi, D. Kai, and J. Dong, "Pulse wave landmark detection method based on wavelet transform," *Data Acquisition and Processing*, vol. 21, 2006.

[20] Z. Hongbiao, "Pulse signal feature extraction based on wavelet transform," *Electronic Measurement Technology*, vol. 32, no. 9, 2009.

[21] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[22] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997.

[23] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[24] W. Liu, C. Li, N. Xu et al., *CVM-Cervix: A Hybrid Cervical Pap-Smear Image Classification Framework Using CNN*, Pattern Recognition, Visual Transformer and Multilayer Perceptron, 2022.

[25] J. Zhang, C. Li, S. Kosov et al., "LCU-Net: A novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, p. 107885, 2021.

[26] Y. Li, C. Li, X. Li et al., "A comprehensive review of Markov random field and conditional random field approaches in pathology image analysis," *Archives of Computational Methods in Engineering*, vol. 29, pp. 609–639, 2022.

[27] H. Chen, C. Li, G. Wang et al., "GasHis-Transformer: A multi-scale visual transformer approach for gastric histopathological image detection," *Pattern Recognition*, vol. 130, p. 108827, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 8955227, 7 pages
https://doi.org/10.1155/2022/8955227

Research Article

A Machine Learning Model Based on Unsupervised Clustering Multihabitat to Predict the Pathological Grading of Meningiomas

Xinghao Wang,[1] **Jia Li**,[1] **Jing Sun**,[1] **Wenjuan Liu**,[1] **Linkun Cai**,[2] **Pengfei Zhao**,[1] **Zhenghan Yang**,[1] **Han Lv**,[1] **and Zhenchang Wang**[1]

[1]*Department of Radiology, Beijing Friendship Hospital, Capital Medical University, Beijing, China*
[2]*School of Biological Science and Medical Engineering, Beihang University, No. 37 XueYuan Road, Beijing 100191, China*

Correspondence should be addressed to Han Lv; chrislvhan@126.com and Zhenchang Wang; cjr.wzhch@vip.163.com

Received 19 July 2022; Accepted 1 September 2022; Published 12 September 2022

Academic Editor: Chen Li

Copyright © 2022 Xinghao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. We aim to develop and validate a machine learning model by enhanced MRI to determine the pathological grading of meningiomas with unsupervised clustering image analysis method, which are multihabitat to reflect the inherent heterogeneity of tumors. *Materials and Methods*. A total of 120 patients with meningiomas confirmed by postoperative pathology were included in the study, including 60 patients with low-grade meningiomas (WHO grade I) and 60 patients with high-grade meningiomas (WHO grade II and WHO grade III). All patients underwent complete head enhanced magnetic resonance scans before surgery or any anti-tumor treatment. Enrolled patients in the group received surgical resection and obtained postoperative pathological data. The patients in the training group (84 people) and the test group (36 people) were randomly divided into two groups according to the ratio of 7 to 3. Multi-habitat features were extracted from MRI images based on enhanced T1. Machine learning method was used to model, which was used to distinguish high-grade meningioma from low-grade meningioma. At the same time, the obtained machine learning model was calibrated and evaluated. *Results*. In patients with low-grade meningioma and high-grade meningioma, we found significant differences in Silhouette coefficient ($P<0.05$). In the machine learning model, the area under the curve was 0.838 in the training group (sensitivity, 67.65%; specificity, 88.82%) and 0.73 in the test group (sensitivity, 69.05%; specificity, 71.43%). After the analysis of calibration curve and decision curve analysis, the model had shown the potential of great application value. *Conclusions*. Multi-habitat analysis based on enhanced MRI (T1) could accurately predict the pathological grading of meningiomas. This unsupervised image-based method could reflect the direct heterogeneity between high-grade meningiomas and low-grade meningiomas, which is of great significance for patients' treatment and prevention of recurrence.

1. Introduction

Meningioma is one of the most common intracranial tumors, accounting for more than one-third of all primary central nervous system tumors [1]. Meningiomas could be usually diagnosed by accident and be generally considered less malignant than other intracranial tumors [2]. Meningiomas originate from arachnoid cells located on the inner surface of the dura mater, usually from meningeal precursor cells derived from the mesoderm and neural crest. Among meningiomas, most meningiomas belong to WHO grade I benign tumors, and the recurrence rate is very low, while nearly 5% are WHO grade II/III tumors, showing a higher nature of invasion and recurrence [2, 3]. The diagnosis of meningioma largely depends on radiology. For example, imaging strongly suggests that in meningioma, biopsy is not necessary [4]. Under normal conditions, asymptomatic meningiomas grow linearly, with a growth rate of 2-4 mm/year, but there are also cases of constant volume and exponential rapid growth [5], so it is very necessary to monitor asymptomatic meningiomas. It is estimated that the 10-year overall survival rate of nonmalignant meningiomas is 81.4%, while that of grade II meningiomas is 53%, and that of grade III meningiomas is 0% [1]. The 5-year recurrence

rate after total resection of grade I meningioma is 7-23%, grade II meningioma is 50-55%, and grade III meningioma is 72-78%[1, 2], which is a huge difference. At present, it is difficult to accurately judge the pathological grade of meningiomas before surgery, but it is very important for patients to choose treatment methods, formulate follow-up plans, and guide prognosis.

Magnetic resonance imaging (MRI) is widely used in the diagnosis of central nervous system diseases, including the diagnosis and detection of diseases and many other aspects. For the imaging research of meningiomas, some studies [6–8] have been carried out by using the methods of multiparameter magnetic resonance imaging and radiomics. And unsupervised clustering multihabitat [9–12] has been widely used in the evaluation of tumor heterogeneity to obtain the heterogeneity information within the tumor. Habitat analysis is an important research method gradually developed in the image field. In this study, we collected preoperative enhanced MR images of patients with meningiomas of different pathological grades for multihabitat analysis to reflect the differences in tumor heterogeneity levels among patients with meningiomas of different grades and help determine their pathological grades.

2. Materials and Methods

2.1. Patient. A total of 120 patients with meningiomas were retrospectively analyzed from May 2016 to June 2019. And all patients underwent preoperative enhanced MRI scanning and successfully obtained pathological tissue samples. Our entire experimental plan had been approved by the ethics committee, in which the images and clinical data of patients had been approved and exempted from informed consent. In the process of processing the data, considering the sensitivity and ethical requirements of the head data, we stored all the data confidentially and unlabeled it to comply with the relevant subject protection principles. The exclusion criteria were as follows: (1) patients with a history of surgery, (2) patients with a history of tumor embolism or gamma knife surgery before MRI, and (3) patients whose T1-weighted images were not clear enough to be analyzed. The patients in the training group (84 patients) and the test group (36 patients) were randomly divided into two groups according to the ratio of 7 to 3.

2.2. MR Imaging. All patients underwent complete enhanced MR scanning. T1-weighted, T2-weighted, and enhanced T1-weighted imaging was imaged with a 3.0 MRI system (Signa, HDxt, General Electric Healthcare, Milwaukee, WI, USA), with an 8-channel array coil. Routine sequence imaging was performed for all subjects and included axial T2WI (TR = 3520 ms, TE = 102 ms, ETL = 20, matrix size = 320 × 256) and FLAIR (TR = 8000 ms, TE = 165 ms, TI = 750 ms, matrix size = 256 × 192) and axial with coronal contrast-enhanced (CE) T1WI after injection of contrast agent.

2.3. Habitat Generation and Feature Extraction. The workflow of the multihabitat is shown in Figure 1. We selected the enhanced T1 image for processing. First, two medical imaging diagnostic doctors with more than 5 years of experience outlined the overall edge of the tumor, and the two found that it was time to introduce a third person who has more than 10 years of experience in the diagnosis of the central nervous system to judge the dispute. Although it was not a radiomics study, we refer to the Image Biomarker Standardization Initiative [13] principles for image processing methods. All tumor edge confirmation work was completed through 3D slicer software (https://www.slicer.org/). After obtaining the complete region of interest of meningioma, multihabitat [14] was implemented through self-built code in Python (3.8.5). K-means algorithm [15] was widely used in unsupervised image segmentation, and our heterogeneous region segmentation also adopts this kind of algorithm. In this paper, we choose a clustering parameter with K equal to three. We selected five clustering indicators to evaluate the tumor intrinsic heterogeneity of meningiomas: (1) inertia, within cluster sum of square error; (2) Calinski-Harabasz Index; (3) Silhouette coefficient; (4) separation; and (5) Davies-Bouldin Index.

2.4. Model Establishment. Through statistical judgment, we screen out meaningful clustering features for final classification. For the features of modeling, we adopt the Z-score method to reduce the adverse effects on the model caused by different feature distributions. For feature screening, we choose many ways of parallel or serial, such as Lasso, PCA, PCC, and analysis of variance. For the selection of machine learning classifiers, we choose many ways, including linear regression, linear regression, logistic regression, linear discriminant analysis, naive Bayes, KNN, random forest, and Gaussian regression. For the model, we used the experimental set for modeling and the verification set for verification to explore the fitting degree of the machine learning model.

2.5. Statistical Analysis. MedCalc version 15.2.2 (http://www.medcalc.org) and R (version 4.0.3) were used for analyses. We extracted labels from the postoperative pathological reports of patients with meningiomas, in which WHO grade I was a low-grade meningioma, and WHO grade II and grade III were a high-grade meningioma. For clustering features, first judge the data type and select the corresponding t-test or U-test. The ROC curve was used to evaluate the model, and the decision curve analysis (DCA) curve and calibration curve of the model are calculated to further describe the model. Bilateral P values less than 0.05 were considered statistically significant.

3. Results

3.1. Clinical Characters. We retrospectively recruited 120 patients with meningiomas, including 56 males and 64 females, 60 patients with low-grade meningiomas (grade I) and 60 patients with high-grade meningiomas (grade II and III). More information could be summarized in Table 1. There was no significant difference in clinical characteristics between the training group and the validation group ($P > 0.05$).

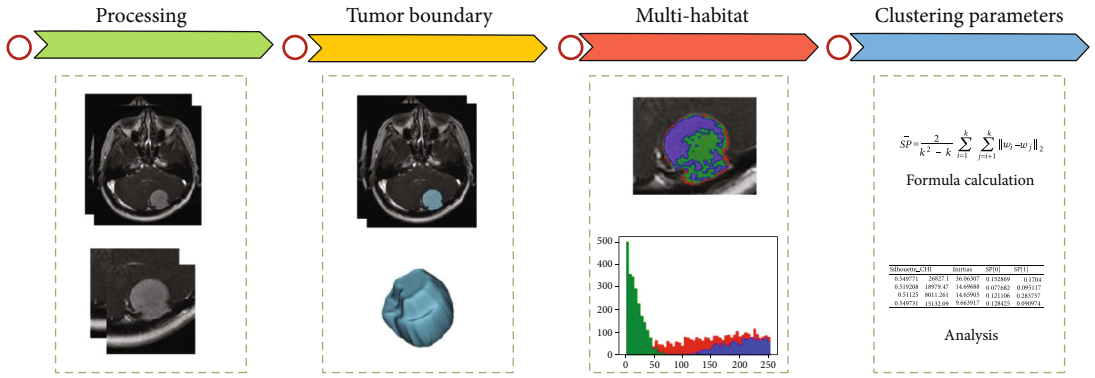


Figure 1: The workflow of the multihabitat.

Table 1: The clinical characteristics of meningiomas.

Characteristic	Training group ($n = 84$)	Test group ($n = 36$)
Age (average year)	52.3	52.9
Gender		
Male	40	16
Female	44	20
PR		
Positive	39	16
Negative	45	20
Ki-67		
Positive	34	17
Negative	50	19
WHO grading		
Grade I	41	19
Grade II	34	13
Grade III	9	4

3.2. Clustering Parameters and Machine Learning Model. For all clustering parameters, we performed the operation of confusion matrix (for the label with different pathological grades). The detailed description of each parameter is shown in Table 2.

Through the screening and dimensionality reduction of clustering features, we had successfully obtained a machine learning model based on inertia, Calinski-Harabasz Index, Silhouette coefficient, and Davies-Bouldin Index. We used Gaussian process as the classifier. Gaussian process combines the features to build a joint distribution to estimate the probability of the classification. The machine learning model had shown a great ability to distinguish different levels of meningiomas (Figure 2). We also calculated its calibration curve (Figure 3) and DCA curve (Figure 4). Through the calibration of the model, we can visually describe the fitting degree of the model. As shown in Figure 4, the decision curve of our model is all on the horizontal axis, and there is no "negative return" area, which is considered to be one of the good clinical application values.

4. Discussion

In this study, unsupervised clustering was used to evaluate the heterogeneity of meningiomas at different pathological levels for enhanced magnetic resonance image (T1), and a machine learning model for regression prediction was established. For the multihabitat method based on K-means, we have successfully divided the subregions of enhanced magnetic resonance images. Unsupervised clustering parameters were used to participate in the establishment of the machine learning model. The final machine learning model could better distinguish high-grade meningiomas from low-grade meningiomas.

Compared with low-grade (benign) meningiomas, high-grade (atypical or anaplastic) tumors have invasive biological behavior, increased risk of recurrence, and increased mortality [16]. The classification of meningioma preoperative prediction is crucial because it affects various treatment plans, including surgical resection strategies. According to the guidelines issued by the European Association of Neuro-Oncology, meningiomas found accidentally and speculated radiologically can be treated only by observation [4], so judging the pathological grade also plays an important role in the formulation of follow-up observation strategies. Due to the limitations of biopsy, in some cases, the histological grade may not be determined. At the same time, according to previous cohort studies, it is found that the correlation between clinical information such as age and gender and tumor grade is weak [16], and it is difficult to have a single clinical feature to predict the pathological grade of tumor. Previous studies [17–20] have focused on the imaging characteristics of meningiomas, diffusion and perfusion imaging, amide proton imaging, and PET to reflect the grading of meningiomas. However, there are great limitations in its grading value and application scenarios. In this paper, we found that clustering parameters and the machine learning model were related to the grading of meningiomas, which may be related to the tumor heterogeneity represented by the three signal areas of MRI, and it is also consistent with previous studies that high-grade meningiomas show more complex texture patterns [21] than low-grade meningiomas on MRI. A previous study [22] also showed that irregular

Table 2: Distinguishing ability of clustering parameters.

	AUC (95% CI)	Specificity	Sensitivity	Youden index
Inertia	0.597 (0.504-0.685)	46.67%	73.33%	0.20
Calinski-Harabasz Index	0.596 (0.502-0.684)	41.67%	76.67%	0.183
Silhouette coefficient	0.754 (0.668-0.828)	83.33%	56.67%	0.40
Separation	0.571 (0.477-0.661)	46.67%	71.67%	0.184
Davies-Bouldin Index	0.674 (0.583-0.757)	48.33%	80.0%	0.283

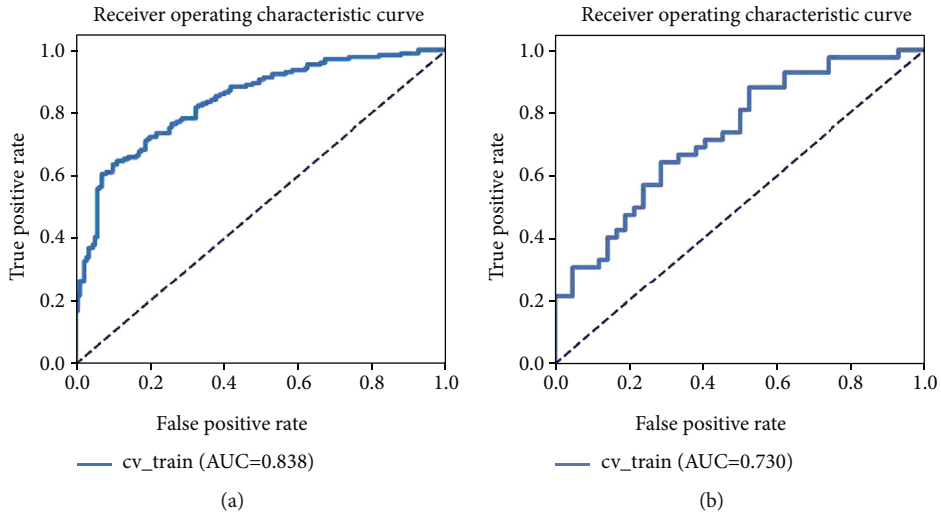


Figure 2: ROC of the machine learning model: (a) the performance of the training group; (b) the performance of the validation group.

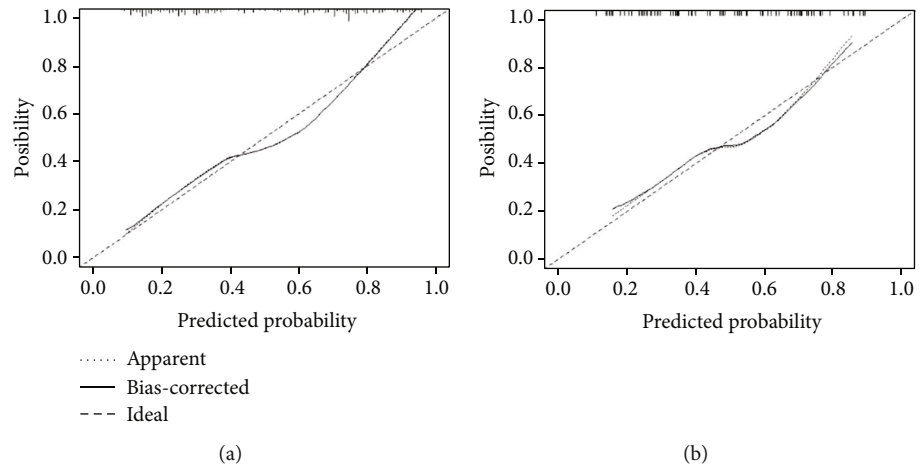


Figure 3: Calibration curve of the machine learning model: (a) the calibration curve of the training group; (b) the calibration curve of the validation group.

tumor boundaries were associated with high-grade meningiomas, which was consistent with our results.

Medical imaging can provide full volume evaluation of the continuous nature of tumors by generating spatial resolution maps of subunits called "voxels" [23–25]. Malignant tumors have complex biology and show significant spatial variation in gene expression, biochemistry, histopathology, and macro structure. Cancer cells not only evolved from clones of single progenitor cells to more aggressive and treatment resistant cells but also showed branching evolution, so that each tumor developed and retained multiple different subclonal populations. There was a wide consensus among various theories of tumor formation that there are differences between tumor cells, progenitor cells, and cells

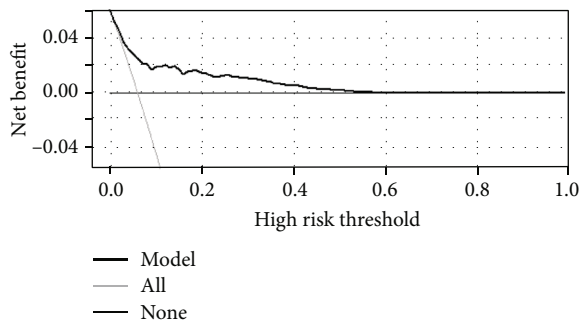


Figure 4: The decision curve analysis of the machine learning model.

between tumors. This genetic heterogeneity combined with spatial variation in different environments can lead to a variety of regional differences in matrix structure, oxygen consumption, energy metabolism, and growth factor expression. Therefore, with the development of different regions of the tumor, each region has spatially different blood perfusion, hypoxia, cell proliferation, apoptosis, and other characteristics [26–28]. Compared with benign meningiomas, the high-grade meningiomas [29, 30] are characterized by genomic instability, which indicates that regional changes in chromosome structure may be an important feature of recurrence, treatment resistance, and invasiveness. An innovative study [28] combining the spatial transcription information of meningiomas shows that although the MRI signals of high-grade meningiomas and low-grade meningiomas are generally the same, the signals of high-grade meningiomas are more variable. The above research shows that the MRI-based heterogeneity of meningiomas is related to the regional transcriptome differences of tumor tissues, in which the regions with high signal are rich in developmental gene expression, and these gene expression programs may be the basis of meningioma cell proliferation and tumor recurrence. Meningiomas with high proliferative potential may show a highly heterogeneous distribution of proliferating cells in tumors, and this heterogeneity may produce irregular shapes. In our experimental protocol, our unsupervised classification of red areas (high signal areas) is always located at the edge of the tumor, which is considered to be related to tumor cell proliferation and infiltration [31]. The data-driven method [14, 32–34] successfully distinguishes survival tumors from nonsurvival tumors using multiparameter MRI and verifies the method against H&E histology [35].

Different from traditional radiomics, we unsupervised the three-dimensional tumor voxels into subregions and analyzed their internal or mutual relationships to explain the spatial heterogeneity. While fully considering the internal differences of their subclassifications, clustering parameters are designed to evaluate the correlation or eccentricity of different subgroups (spatial subregions). For different parameters, they represent different perspectives to evaluate the heterogeneity components within tumors, which is undoubtedly consistent with previous studies [36–39]. In addition to the absolute characterization of tumors, we pay attention to the direct internal relations of different subregions, and their ability to respond to internal relations is stronger than direct indicators. Although many factors were considered in the study design, there were some deficiencies in this study (retrospective study). First of all, for the lack of control over information sources, one of the common problems of retrospective research is that we cannot design relevant data collection methods or other confounding factors in advance. Secondly, for the processing of brain MRI images, segmentation technology is one of the key technologies [40], but due to the limitations of experimental conditions, our preprocessing methods and segmentation methods are not accurate and automatic. Third, our research data is single center, which will undoubtedly weaken the persuasion of the results, which undoubtedly needs to be further explored in the real world of large samples and multicenters.

5. Conclusions

Multihabitat analysis based on K-means by enhanced MRI could distinguish high-grade meningioma from low-grade meningioma accurately. This unsupervised image-based method could reflect the direct heterogeneity between different grade meningiomas, which is of great importance for patients' treatment and prevention of recurrence.

Data Availability

The MRI images of the study are available from the corresponding authors upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was sponsored in part by the Beijing Scholar 2015 (Zhenchang Wang) and Beijing Key Clinical Discipline Funding (No. 2021-135).

References

[1] Q. T. Ostrom, H. Gittleman, P. Liao et al., "CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014," *Neuro-Oncology*, vol. 19, Supplement_5, pp. v1–v88, 2017.

[2] I. Maggio, E. Franceschi, A. Tosoni et al., "Meningioma: not always a benign tumor. A review of advances in the treatment of meningiomas," *CNS Oncology*, vol. 10, no. 2, 2021.

[3] R. A. Buerki, C. M. Horbinski, T. Kruser, P. M. Horowitz, C. D. James, and R. V. Lukas, "An overview of meningiomas," *Future Oncology*, vol. 14, no. 21, pp. 2161–2177, 2018.

[4] R. Goldbrunner, G. Minniti, M. Preusser et al., "EANO guidelines for the diagnosis and treatment of meningiomas," *The Lancet Oncology*, vol. 17, no. 9, pp. e383–e391, 2016.

[5] T. Hashiba, N. Hashimoto, S. Izumoto et al., "Serial volumetric assessment of the natural history and growth pattern of incidentally discovered meningiomas," *Journal of Neurosurgery*, vol. 110, no. 4, pp. 675–684, 2009.

[6] K. R. Laukamp, G. Shakirin, B. Baeßler et al., "Accuracy of radiomics-based feature analysis on multiparametric magnetic resonance images for noninvasive meningioma grading," *World Neurosurgery*, vol. 132, pp. e366–e390, 2019.

[7] J. Zhang, K. Yao, P. Liu et al., "A radiomics model for preoperative prediction of brain invasion in meningioma non-invasively based on MRI: a multicentre study," *eBioMedicine*, vol. 58, article 102933, 2020.

[8] G. Hamerla, H. J. Meyer, S. Schob et al., "Comparison of machine learning classifiers for differentiation of grade 1 from higher gradings in meningioma: a multicenter radiomics study," *Magnetic Resonance Imaging*, vol. 63, pp. 244–249, 2019.

[9] L. Beer, P. Martin-Gonzalez, M. Delgado-Ortet et al., "Ultrasound-guided targeted biopsies of CT-based radiomic tumour habitats: technical development and initial experience in metastatic ovarian cancer," *European Radiology*, vol. 31, no. 6, pp. 3765–3772, 2021.

[10] D. Cherezov, D. Goldgof, L. Hall et al., "Revealing tumor habitats from texture heterogeneity analysis for classification of lung cancer malignancy and aggressiveness," *Scientific Reports*, vol. 9, no. 1, p. 4500, 2019.

[11] E. Sala, E. Mema, Y. Himoto et al., "Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging," *Clinical Radiology*, vol. 72, no. 1, pp. 3–10, 2017.

[12] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[13] A. Zwanenburg, M. Vallières, M. A. Abdalah et al., "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.

[14] T. Du and H. Zhao, "Habitat analysis of breast cancer-enhanced MRI reflects BRCA1 mutation determined by immunohistochemistry," *BioMed Research International*, vol. 2022, Article ID 9623173, 9 pages, 2022.

[15] M. Dias, B. Rocha, J. F. Teixeira, and H. P. Oliveira, "Automatic sternum segmentation in thoracic MRI," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1018–1021, Berlin, Germany, 2019.

[16] V. R. Kshettry, Q. T. Ostrom, C. Kruchko, O. Al-Mefty, G. H. Barnett, and J. S. Barnholtz-Sloan, "Descriptive epidemiology of World Health Organization grades II and III intracranial meningiomas in the United States," *Neuro-Oncology*, vol. 17, no. 8, pp. 1166–1173, 2015.

[17] S. Peng, Z. Cheng, and Z. Guo, "Diagnostic nomogram model for predicting preoperative pathological grade of meningioma," *Translational Cancer Research*, vol. 10, no. 9, pp. 4057–4064, 2021.

[18] J. W. Lee, K. W. Kang, S. H. Park et al., "18F-FDG PET in the assessment of tumor grade and prediction of tumor recurrence in intracranial meningioma," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 36, no. 10, pp. 1574–1582, 2009.

[19] B. Joo, K. Han, Y. S. Choi et al., "Amide proton transfer imaging for differentiation of benign and atypical meningiomas," *European Radiology*, vol. 28, no. 1, pp. 331–339, 2018.

[20] Y. Zhu, C. Man, L. Gong et al., "A deep learning radiomics model for preoperative grading in meningioma," *European Journal of Radiology*, vol. 116, pp. 128–134, 2019.

[21] Y. W. Park, S. Kim, S. S. Ahn et al., "Magnetic resonance imaging-based 3-dimensional fractal dimension and lacunarity analyses may predict the meningioma grade," *European Radiology*, vol. 30, no. 8, pp. 4615–4622, 2020.

[22] Y. Kawahara, M. Nakada, Y. Hayashi et al., "Prediction of high-grade meningioma by preoperative MRI assessment," *Journal of Neuro-Oncology*, vol. 108, no. 1, pp. 147–152, 2012.

[23] J. P. O'Connor, C. J. Rose, J. C. Waterton, R. A. Carano, G. J. Parker, and A. Jackson, "Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome," *Clinical Cancer Research*, vol. 21, no. 2, pp. 249–257, 2015.

[24] J. Zhang, C. Li, S. Kosov, M. Grzegorzek, and H. Li, "LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, article 107885, 2021.

[25] H. Chen, C. Li, X. Li et al., "GasHis-Transformer: A multi-scale visual transformer approach for gastric histopathology image classification," 2021, https://arxiv.org/abs/2104.14528.

[26] Y. P. Choi, H. S. Shim, M. Q. Gao, S. Kang, and N. H. Cho, "Molecular portraits of intratumoral heterogeneity in human ovarian cancer," *Cancer Letters*, vol. 307, no. 1, pp. 62–71, 2011.

[27] A. J. Craig, J. von Felden, and A. Villanueva, "Molecular profiling of liver cancer heterogeneity," *Discovery Medicine*, vol. 24, pp. 117–125, 2017.

[28] S. T. Magill, H. N. Vasudevan, K. Seo et al., "Multiplatform genomic profiling and magnetic resonance imaging identify mechanisms underlying intratumor heterogeneity in meningioma," *Nature Communications*, vol. 11, no. 1, p. 4803, 2020.

[29] W. L. Bi, N. F. Greenwald, M. Abedalthagafi et al., "Genomic landscape of high-grade meningiomas," *NPJ Genomic Medicine*, vol. 2, no. 1, 2017.

[30] E. A. Williams, S. Santagata, H. Wakimoto et al., "Distinct genomic subclasses of high-grade/progressive meningiomas: NF2-associated, NF2-exclusive, and NF2-agnostic," *Acta Neuropathologica Communications*, vol. 8, no. 1, p. 171, 2020.

[31] D. Kalasauskas, Y. Tanyildizi, M. Renovanz et al., "Evaluation of resection margin after image-guided dural tail resection in convexity meningiomas," *Journal of Clinical Medicine*, vol. 10, no. 6, p. 1177, 2021.

[32] O. Jalnefjord, M. Montelius, J. Arvidsson, E. Forssell-Aronsson, G. Starck, and M. Ljungberg, "Data-driven identification of tumor subregions based on intravoxel incoherent motion reveals association with proliferative activity," *Magnetic Resonance in Medicine*, vol. 82, no. 4, pp. 1480–1490, 2019.

[33] P. D. Tar, N. A. Thacker, M. Babur et al., "Habitat imaging of tumors enables high confidence sub-regional assessment of response to therapy," *Cancers (Basel)*, vol. 14, no. 9, p. 2159, 2022.

[34] X. Wang, K. Wu, X. Li, J. Jin, Y. Yu, and H. Sun, "Additional value of PET/CT-based radiomics to metabolic parameters in diagnosing lynch syndrome and predicting PD1 expression in endometrial carcinoma," *Frontiers in Oncology*, vol. 11, article 595430, 2021.

[35] Y. Li, C. Li, X. Li, K. Wang, and Q. Wang, "A comprehensive review of Markov random field and conditional random field approaches in pathology image analysis," *Archives of Computational Methods in Engineering*, vol. 29, pp. 609–639, 2021.

[36] A. K. Syed, J. G. Whisenant, S. L. Barnes, A. G. Sorace, and T. E. Yankeelov, "Multiparametric analysis of longitudinal quantitative MRI data to identify distinct tumor habitats in preclinical models of breast cancer," *Cancers (Basel)*, vol. 12, no. 6, p. 1682, 2020.

[37] A. K. Syed, R. Woodall, J. G. Whisenant, T. E. Yankeelov, and A. G. Sorace, "Characterizing trastuzumab-induced alterations in intratumoral heterogeneity with quantitative imaging and immunohistochemistry in HER2+ breast cancer," *Neoplasia*, vol. 21, no. 1, pp. 17–29, 2019.

[38] A. S. Kazerouni, D. A. Hormuth 2nd, T. Davis et al., "Quantifying tumor heterogeneity via MRI habitats to characterize microenvironmental alterations in HER2+ breast cancer," *Cancers (Basel)*, vol. 14, no. 7, p. 1837, 2022.

[39] X. Wang, C. Xu, and H. Sun, "DNA damage repair-related genes signature for immune infiltration and outcome in cervical cancer," *Frontiers in Genetics*, vol. 13, article 733164, 2022.

[40] S. B. Martins, J. Bragantini, A. X. Falcão, and C. L. Yasuda, "An adaptive probabilistic atlas for anomalous brain segmentation in MR images," *Medical Physics*, vol. 46, no. 11, pp. 4940–4950, 2019.

Hindawi
BioMed Research International
Volume 2022, Article ID 8504149, 9 pages
https://doi.org/10.1155/2022/8504149

Research Article

Efficient Synchronous Real-Time CADe for Multicategory Lesions in Gastroscopy by Using Multiclass Detection Model

Yiji Ku, Hui Ding, and Guangzhi Wang

Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Guangzhi Wang; wgz-dea@tsinghua.edu.cn

Received 2 June 2022; Accepted 21 July 2022; Published 31 August 2022

Academic Editor: Chen Li

Copyright © 2022 Yiji Ku et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Often more than one category of lesions in patients' gastrointestinal tracts need to be found in the endoscopic examination. Therefore, there is a need to establish an efficient synchronous real-time computer-aided detection (CADe) system for multicategory lesion detection. This paper proposes to build a system with a multiclass detection model based on the YOLOv5 to detect multicategory lesions synchronously in real-time. Two joint detection CADe systems using multiple single-class detection models with the same structure in parallel or series are established for comparison. A retrospective dataset containing 31117 images from 3747 patients is used in this study. To train the model, various online data augmentation methods and multiple loss functions are used. The proposed CADe system can synchronously detect cancers, gastrointestinal stromal tumours, polyps, and ulcers from different quality input images with 98% precision, 89% recall, and 90.2% mAP. The detection speed is 47 frames per second with a 0.04 s latency on a PC workstation. Compared to the two joint detection CADe systems, the proposed system is more accurate with faster speed and lower latency. Two extra experiments indicated that the lesion detection model based on YOLOv5x could provide better performance than other common YOLO structures and that different accuracy metrics and lesion categories have different requirements for the number of training images. The proposed synchronous real-time CADe system with the multiclass detection model can detect multicategory lesions with high accuracy and speed and low latency on limited hardware. It expands the clinical application of CADe in endoscopy and uses expensive labelled medical images more efficiently than multiple single-category lesion models for joint detection.

1. Introduction

Diseases of the gastrointestinal tract are common in humans [1]. Early examination and treatment can significantly improve their effects [2]. The endoscope is the main instrument used to diagnose and treat gastrointestinal lesions in a clinic [3]. Owing to the large number of images generated during the endoscopy, finding all the lesions accurately and timely is difficult. Using computer-aided detection (CADe) system to assist doctors in endoscopy can reduce the rate of misdiagnosis and increase the rate of detection to improve efficiency [4].

Classical CADe systems mainly use the active-window method with artificially designed features [5, 6]. However, it is unreliable and time-consuming to extract complex pathological features. Deep learning has better feature extraction and classification capabilities [7–9]. It significantly improves the performance of CADe systems [10]. Nowadays, CADe in endoscopy can accurately detect single category lesions in real-time using deep learning models [11–14].

However, there may be more than one category of lesions in a patient's gastrointestinal tract, which should be detected during an endoscopy [15]. This requires the CADe system to detect multiple lesion categories synchronously in real-time. Using multiple single-category lesion detection models parallelly or serially to compose such a system will further raise the hardware requirements with the increase of lesion category without a limit for real-time detection. This will in turn increase the system's building cost. A multicategory lesion detection model may have better detection speed and latency with limited hardware resources. However, whether it can obtain similar accuracy or not remains unanswered. YOLO is a series of one-stage detection models with a good balance between speed and accuracy [16]. These

models have shown good performance in single-category lesion detection under endoscopy [13, 14, 17–19]. However, a detailed study on the synchronous real-time detection of multicategory lesions is required.

This paper proposes a scheme to use a multicategory lesion detection model based on YOLOv5 [20] to establish a multicategory lesion CADe system. It can realize synchronous low-delay real-time detection in limited hardware resources. The study also tested two CADe systems for multicategory lesion joint detection using multiple single-category lesion detection models with the same structure. The models used in all three systems are trained on a large and complicated local dataset. Compared with the two joint detection systems, the proposed system obtains better speed and latency on a PC workstation and achieves higher accuracy, which proves that using a multicategory lesion detection model is more effective for detecting multicategory lesions in expensive endoscopic images. To the best of our knowledge, this is the first study on synchronous real-time multicategory detection for gastroscopy. It can help researchers establish more efficient CADe systems for endoscopy.

2. Materials and Methods

2.1. Dataset. A large and complicated dataset was collected from the hospital for retrospective study in this work. It contains 31117 images from 3747 patients without private information. Experienced doctors manually annotate 27241 lesions in images into four categories: cancer, ulcer, polyp, and gastrointestinal stromal tumour (GIST). The number of images and lesions are shown in Table 1. All images are unmodified and taken by multiple operators using Olympus 260 series endoscopes. The sizes, aspect ratios, and quality of these images are not selected. Some examples of low-quality images in the dataset are shown in Figure 1. The annotated bounding box (bbox) area ratios also have a significant difference. Due to the statistic of the ratio of the target bbox over the entire image area in the dataset, the ratio less than 0.04 is defined as the little target, and the ratio greater than 0.25 is defined as the large target, and the rest are medium targets. The distribution of targets is shown in Figure 2. Such a dataset makes the model more difficult to train but more robust in reality.

The images in the dataset are numbered and assigned to a training set, validation set, and test set according to the archiving order. The training set contains 3600 images of each category lesion and 3600 negative samples, a total of 18000 images. The validation set contains 1200 images of each image type, totalling 6000 images. All the other 7117 images are included in the test set. The distribution of the category of lesions in these sets is shown in Figure 3, which can be considered balanced. Negative samples included in the dataset ensure that the model would not be overconfident, which is confirmed by other studies [21].

2.2. Synchronous Multicategory Lesion Detection System. The proposed multicategory lesion detection system (MDS) is shown in Figure 4. It can synchronously detect cancers, polyps, GISTs, and ulcers in gastroscopy. This system can also obtain low latency and high FPS using limited hardware resources.

Table 1: The number of images and lesions in the dataset.

Illusion	Cancer	GIST	Polyp	Ulcer	Negative sample
Number of images	6238	6239	6227	6207	6206
Number of lesions	7483	6239	7183	6336	—

Single images are streamed into a YOLOv5-based multiclass detection model (MDM, see Figure 5) to detect multicategory lesions after being preprocessed by scaling and padding to 640×640. The HardSwish function [22], expressed in Equation (1), is used as an activation function in the model. The model outputs prediction vectors containing the corner coordinates of detected bboxes of lesions, the categories of lesions, and the confidence scores. IoU-based non-maximum suppression is used to postprocess the prediction vectors. The system converts the vectors into bboxes, selects the one with the highest confidence from multiple overlapping bboxes as the final result, and draws it in the image.

$$\text{HardSwish}(x) = x\frac{\min\left(6, \max\left(0, x+3\right)\right)}{6}. \tag{1}$$

2.3. Multicategory Lesion Joint Detection Systems. MDS is compared with multiple single-category lesion detection models to detect multicategory lesions jointly. A single-class detection model (SDM) is trained for each lesion category. Four SDMs are run parallelly or serially to replace the MDM in MDS to obtain two modified systems. According to the running order of SDMs, the two systems are named multicategory lesion parallel detection system (MPDS) and multicategory lesion serial detection system (MSDS). The system runs SDMs parallelly for MPDS and serially for MSDS to generate prediction vectors. Running N SDMs in parallel needs N times the computing resources of one model to obtain similar speed and latency. Running N SDMs in series needs the same computing resources as one model but nearly N times the running time and latency. The preprocessing and postprocessing in the two systems are identical to MDS.

2.4. Training Strategy for the Lesion Detection Model. All the systems are implemented on Python 3.8 and PyTorch 1.8, and the models are trained on a workstation equipped with NVIDIA RTX3090 24 GB. The training processes are shown in Figure 6.

This study uses various online data augmentation methods in training to improve the robustness and suppress the negative impact of low-quality images. The methods and parameters used are listed in Table 2. Only one category lesion label is reserved when training an SDM. Others are set to empty as negative samples. When training an MDM, all labels are reserved.

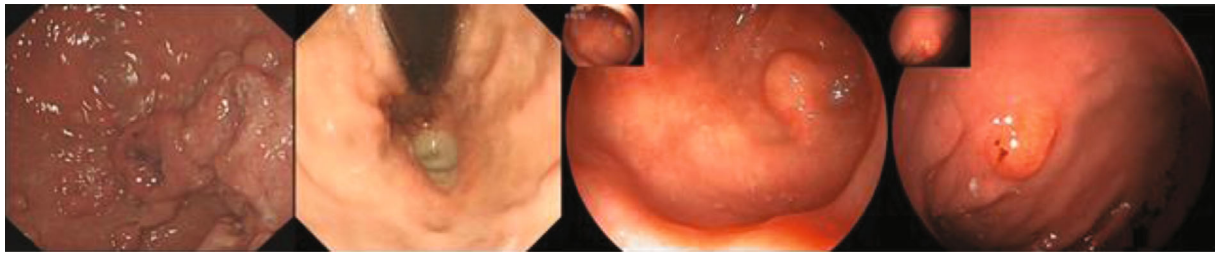


Figure 1: Examples of low-quality images in the dataset. The formats of images are various, and there may be various artefacts, such as highlights and blur. The aspect ratios are also not the same.

These models are trained from weights pretrained on the COCO dataset [24]. Using pretrained weights in endoscopic image analysis can improve the accuracy and save training time [25, 26]. All layers are not frozen. The initial learning rate is 0.01. The warmup strategy [27] is adopted to decay the learning rate to 0.001 in the form of cosine. The optimizer uses SGD momentum, and the momentum parameter is set to 0.937. This study uses Box loss, Conf loss, and Class loss to optimize the model. Box loss is calculated based on CIoU [28] written as Equation (2). Conf loss and Class loss are calculated by cross-entropy written as Equation (3). Since the SDMs do not judge the lesion category, their Class losses are always set to 0. Due to the large value of Box loss in training, its weight is adjusted to 0.1. The weights of the other loss functions are 1.0 to make the training process smoother. This study uses mini-batch to set the batch size as 72 to obtain a better training effect. Under the above settings, these models are close to convergence after training for 1000 epochs.

$$\begin{cases} \text{Loss}_{\text{Box}} = 1 - \text{IoU} + \dfrac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v, \\ \text{IoU} = \dfrac{\text{Box} \cap \text{Box}^{gt}}{\text{Box} \cup \text{Box}^{gt}}, \\ \alpha = \dfrac{v}{1 - \text{IoU} + v}, \\ v = \dfrac{4}{\pi^2}\left(\arctan \dfrac{w^{gt}}{h^{gt}} - \arctan \dfrac{w}{h}\right)^2, \end{cases} \tag{2}$$

$$\begin{cases} \text{Loss}_{\text{Conf/Class}} = \text{mean}\{l_1, \cdots l_n\}, \\ l_n = -(y_n * \log(p_n) + (1 - y_n) * \log(1 - p_n)), \end{cases} \tag{3}$$

where the Box^{gt} and Box are the bbox of ground-truth and prediction, respectively. $\rho^2(b, b^{gt})$ is the distance between the centre points of the Box^{gt} and Box. c is the diagonal length of the smallest box covering the Box^{gt} and Box. w^{gt} and w are the width of the Box^{gt} and Box. h^{gt} and h are the height of the Box^{gt} and Box, respectively. y_n and p_n are the label and prediction, respectively.

2.5. Metrics. This study uses multiple objective metrics to evaluate accuracy and speed performance. The precision, recall, and mAP are used to evaluate the accuracy, and the frame per second (FPS) is used to evaluate the detection speed.

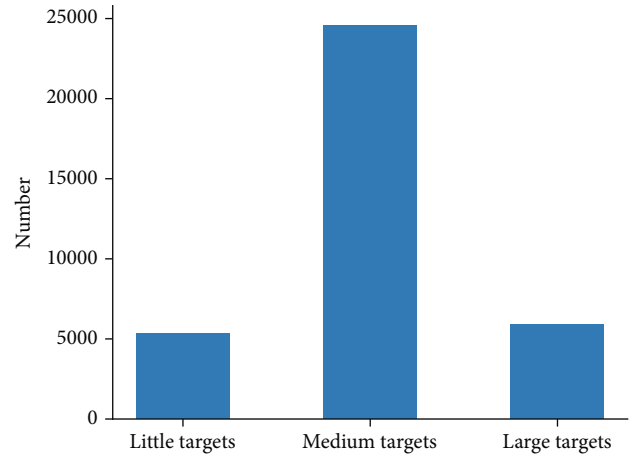


Figure 2: The distribution of different target sizes in the dataset.

$$\text{Presicion} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

$$\text{mAP} = \sum_{\text{confidence_threshold}=0}^{\text{confidence_threshold}=1} \text{Presicion} \times \text{Recall}, \tag{6}$$

where TP, FP, and FN are true positive, false positive, and false negative, respectively. When $\text{IoU} \geq 0.9$, the prediction is considered to have detected a lesion.

3. Results

3.1. Multicategory Lesion Detection Performance. The models with the highest accuracy on the validation set were selected to test on the test set to examine the performance of MDS, MPDS, and MSDS. These systems were run on an NVIDIA TITAN V 12 GB to simulate a PC workstation. The input image was scaled to 640×640 by padding, and the batch size was set to 1.

In the test, the detection speed of the MDS is 47 FPS, the latency is 0.04 s, the MPDS is 17FPS and 0.07 s, and the MSDS is 13FPS and 0.1 s. Therefore, only the MDS can

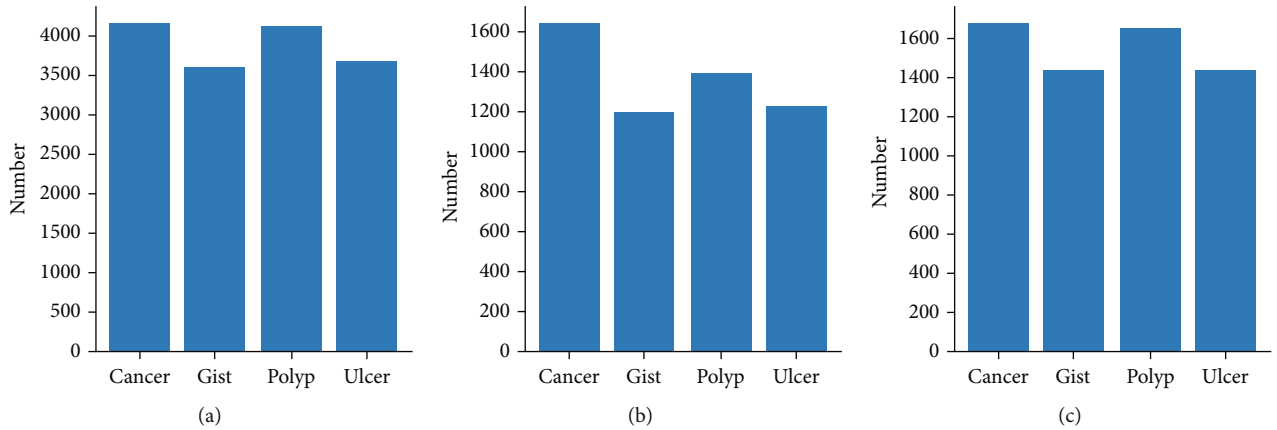


Figure 3: The number of different lesion categories in the dataset. (a) Training set. (b) Validation set. (c) Test set.

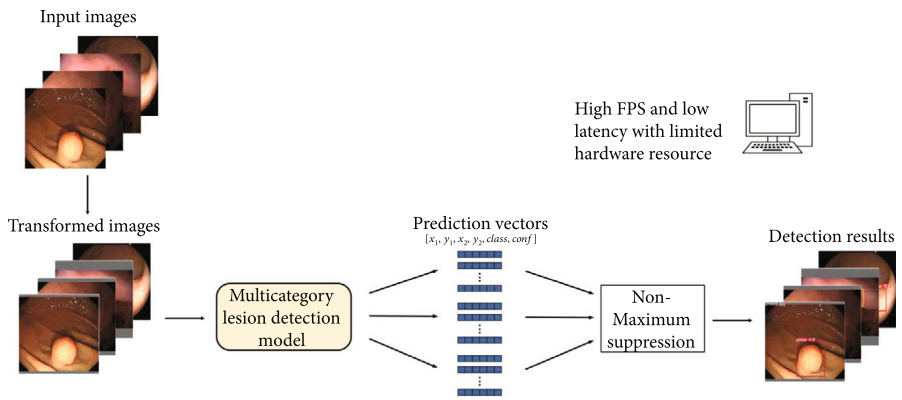


Figure 4: The synchronous multicategory lesion detection system for gastroscopy. The system is designed to run with limited hardware resources.

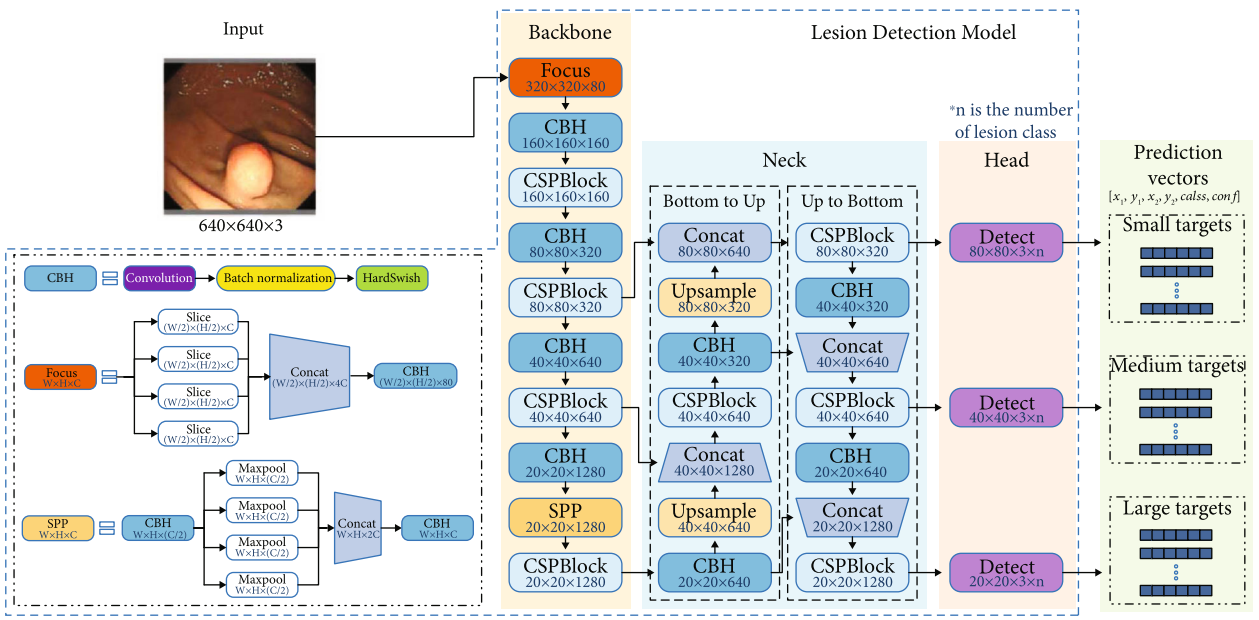
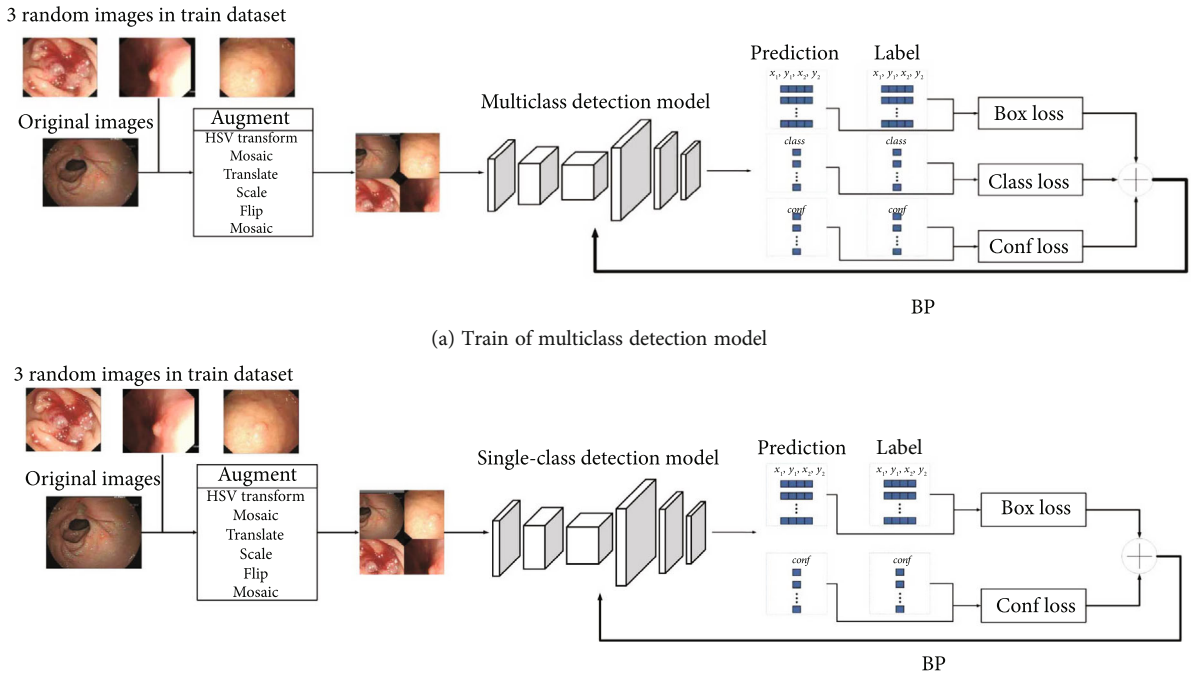


Figure 5: Structure of the lesion detection model based on YOLOv5. All models in this paper use the same design.

run in real-time on a PC workstation. The load of running MDS is approximately 55% on the GPU and 100% on one CPU core, while the MPDS is 80% on the GPU and 65% on four CPU cores, and the MSDS is 65% on the GPU and 100% on one CPU core. It can be observed that the MDS and MSDS need similar hardware resources, but the MPDS



(a) Train of multiclass detection model

(b) Train of single-class detection model

Figure 6: The training processes of detection models (a) MDM for MDS and (b) SDM for MPDS and MSDS.

Table 2: Data augmentation parameters in training.

Augmentation	Parameter
Transformation in HSV colour space	Transform range in hue: ±1.5% Transform range in saturation: ±70% Transform range in value: ±40%
Translation	Range: ±10%
Scale	Range: ±50%
Flip	Probability of flip up-down: 50% Probability of flip left-right: 50%
Mosaic [23]	Number of images: 4

Table 3: Detection performance of all lesion detection models.

	Precision		Recall		mAP	
Model	MDM	SDMs	MDM	SDMs	MDM	SDMs
All	**98%**	95.5%	**89%**	88.5%	**90.2%**	88.3%
Cancer	**95.4%**	91.7%	**66.3%**	63.7%	**67.3%**	62.9%
GIST	**99.9%**	98.5%	96.9%	**97.9%**	**98.7%**	98.6%
Polyp	**99.3%**	97.5%	**98%**	**98%**	**99%**	98.3%
Ulcer	**97.5%**	94.3%	**94.8%**	94.6%	**96%**	93.4%

needs many times. For MPDS, the hardware pressure and speed lower than the theory may be due to other hardware bottlenecks.

The accuracy of the model used in the system represents the system's accuracy, as shown in Table 3. For all lesions, MDM has higher accuracy with 98% precision, 89% recall, and 90.2% mAP, whereas that of SDM is 95.5%, 88.5%, and 88.3%, respectively. The accuracies of the two types of models for GIST and polyp are similar. However, the accuracy of MDM is significantly better for the other two categories of lesions, especially in cancer.

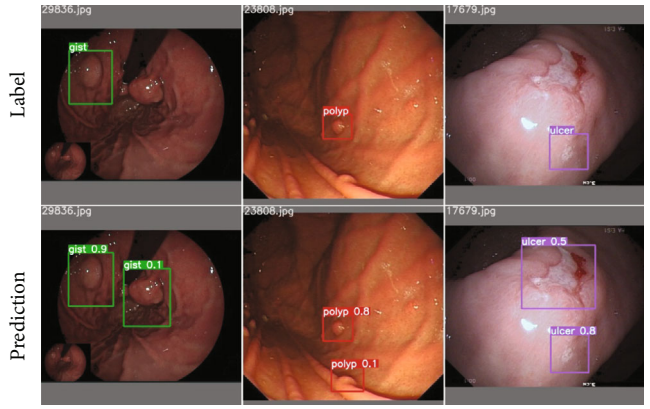


Figure 7: MDM can detect some unannotated lesions. The first line is the label, and the second line is the model prediction.

For GIST, polyp, and ulcer, the MDM has achieved over 97% precision and over 94% recall, which are extremely high. The model can detect some targets missed in the annotation for these categories, as shown in Figure 7. However, MDM and SDM have a low recall for cancer with many false detections, as shown in Figure 8(a). The reason may be that cancer has more complex features than the others to be learned in a similar number of lesions in the training set. The other three categories have only a negligible number of missed predictions. Their features of which are usually quite different in the dataset, as shown in Figures 8(b)–8(d).

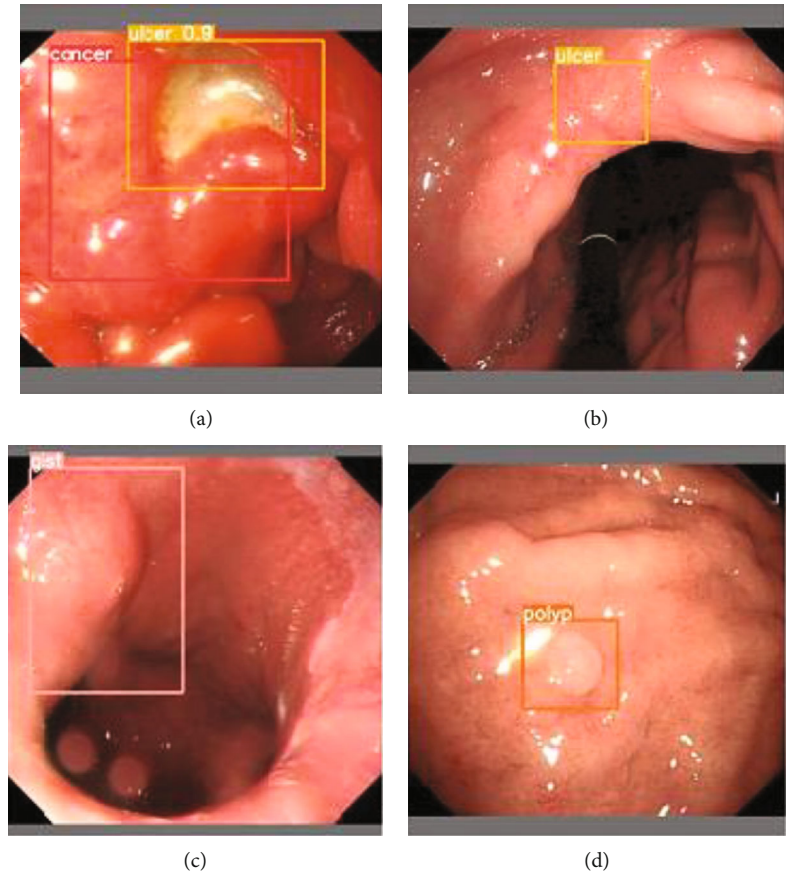


Figure 8: Typical detection failure examples of each lesion category. The lesion in (a) is labelled as cancer which is detected as an ulcer because of the vitiligo. The lesions labelled in (b–d) are false positive.

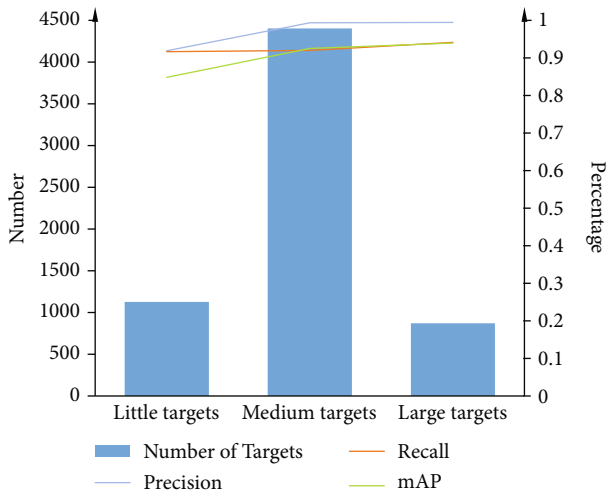


Figure 9: The accuracy of MDM for different size target lesions. The target size is bigger, and the accuracy is higher.

This study also evaluated the performance of MDM on different size targets defined in the "Dataset" section, as shown in Figure 9. It can be observed that the model has high accuracy for different size lesions. However, the model has higher accuracy for larger targets, which may be because larger lesions contain more features and are therefore easier to be detected.

Table 4: Performance of different YOLO model designs.

Model	Precision	Recall	mAP	FPS
YOLOv5x	**0.985**	0.885	**0.898**	47
YOLOv5l	0.903	**0.887**	0.881	51
YOLOv4	0.897	0.843	0.826	54
YOLOv3-SPP	0.879	0.831	0.802	69

3.2. Performance Comparison of Different YOLO Model Designs. In addition to YOLOv5x, this paper compared the performance of YOLOv5l, YOLOv4 [23], and YOLOv3-SPP based on MDS. The training and testing are the same as the "Training Strategy for the Lesion Detection Model" section, and the results are listed in Table 4. It can be found that the model based on YOLOv5x is significantly better than others in accuracy and can also be run in real-time. YOLOv5x can provide better detection capabilities.

3.3. The Influence of Train Image Number. The number of training images affects the accuracy and robustness of the deep learning model significantly. It may be one of the reasons why the model cannot detect cancers better. However,

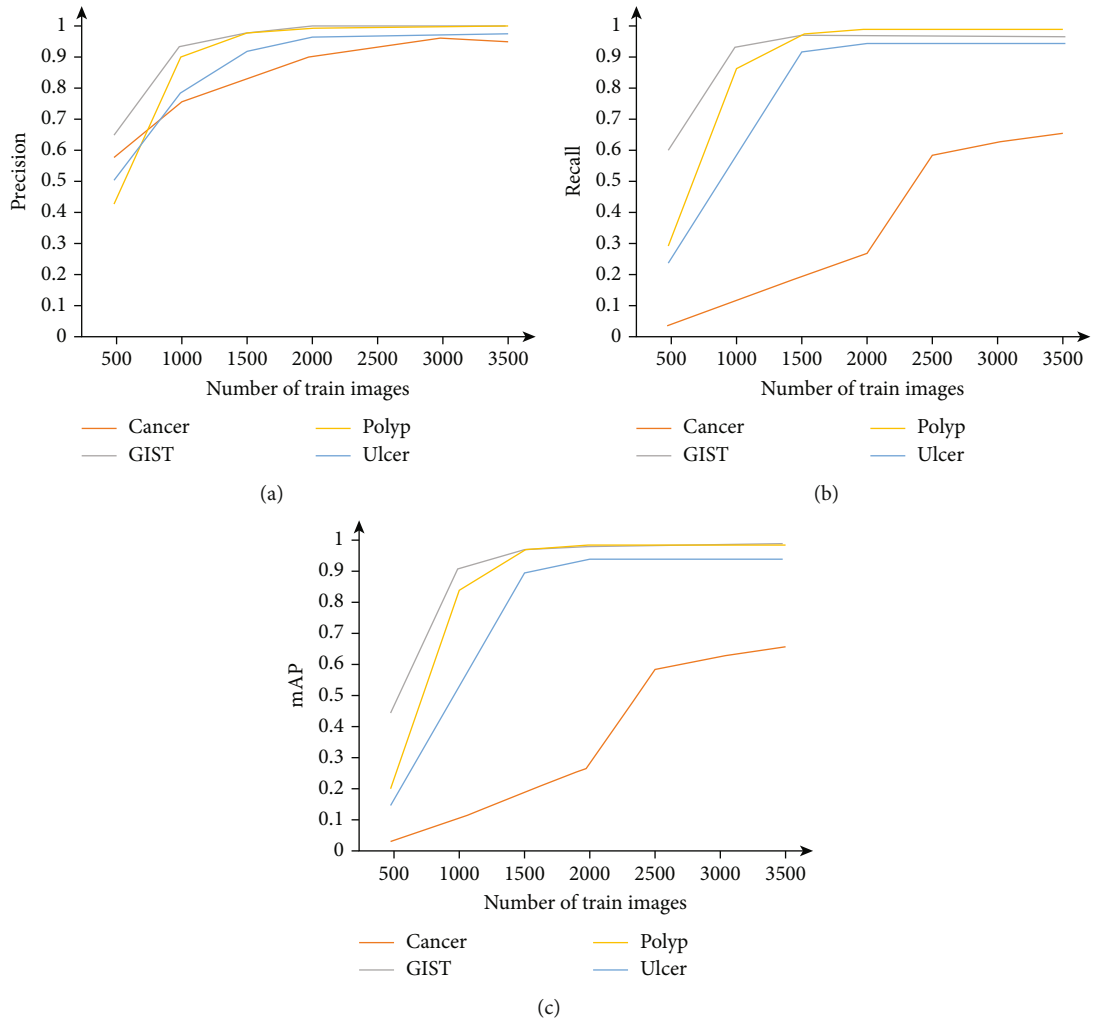


Figure 10: The effect of the number of training images on MDS accuracy metrics of each lesion category. The effects on the precision, recall, and mAP are shown in (a), (b), and (c), respectively.

its impact needs to be clarified. This study explored the impact of the YOLOv5x based on MDS by setting the number of each lesion train image to n ($n = 500$, 1000, 1500, 2000, 2500, 3000, and 3500). Other settings are the same as above. The results are shown in Figure 10. It can be found that the number of training images affects different metrics of different lesions differently. For all lesion categories, the number of images hardly affects the precision after exceeding 2500 for each category. After exceeding 2000 images, other accuracy metrics improve negligibly for GIST, polyp, and ulcer but continue rising significantly for cancer. It supports the previous assumption that cancers have more complex features to be learned, making the number of training images insufficient.

4. Discussion

For multicategory lesion detection, this study proposes and establishes a synchronous real-time CADe system using MDM basing YOLOv5 named MDS and two joint detection systems with multiple SDMs named MPDS and MSDS. MDS can detect cancers, GISTs, polyps, and ulcers from single images with 98% precision, 89% recall, and 90.2% mAP better than MPDS and MSDS. MDS significantly outperforms MPDS and MSDS by achieving 47 FPS with 0.04 s delay on a PC workstation. It is the only one in three systems realizing online detection. The results indicate that using MDM is more meaningful and efficient for costly labelled endoscopic images. This study also explores the influences of the model design and the training image number on the MDS performance.

The existing works [15, 29] about multicategory lesion detection mostly focus on intestinal lesions and lack of detailed study of performance. Compared with them, this study achieves real-time and more accurate detection on a large and complicated dataset and studies some factors that affect performance. The performance comparison of the MDM and SDMs is also a novelty.

The model structure used in all systems is identical to facilitate the performance comparison. Modifying the structure of SDMs may improve the speed or accuracy. However, it is not easy to improve both simultaneously. If SDM uses a

more efficient structure, the model can be easily modified and trained on a dataset for multicategory lesion detection. In summary, compared with the systems combining multiple SDMs directly, the advantages of the CADe system with an MDM are comprehensive in both accuracy and speed.

Better speed is an expected advantage, but the higher accuracy is worth discussing. It is generally believed that a binary classification is easier than classifying more categories. However, the MDM is more accurate in the test, which may be due to some similar features of different lesion categories interfere with the SDM. This makes it difficult to judge whether these lesions belong to the target or not. The MDM eliminates the interference as much as possible through multicategory label supervision. It is similar to the auxiliary learning [30] for each category.

The dataset used is designed as a quantity balance of multiple lesion categories, which can eliminate the preference of models and fairly evaluate the detection performance of different categories. However, there should be differences in the distribution of four diseases in practice. This will not affect the conclusion of system designs from a technical perspective shown in this paper; however, it may affect the CADe performance in the clinic. As far as we know, there is a lack of research on the distribution of these lesions under gastroscopy, and the existing works, including the clinical studies of CADe, have not considered it. Therefore, it is difficult to design a dataset close to clinical distribution at present. In the future, if more extensive and large-scale clinical image data can be obtained in subsequent research, we will further explore the impact of lesion distribution on CADe performance from a clinical perspective.

5. Conclusions

The constructed system can run on limited hardware and detect multicategory lesions accurately and synchronously in real-time. It can further expand the application of CADe in clinic and more efficiently use expensive-labelled medical images. We expect that this work will be a pioneering reference for future studies and researchers. The system can detect GISTs, polyps, and ulcers with extremely high accuracy, and the next step will be to examine its clinical effectiveness. The ways of improving the accuracy of cancer detection will also be explored in future works. In addition, we also hope to improve clinical applicability by detecting more categories or adding new features.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Key R&D Program of China (2019YFC0119503) and NSFC (U20A20389).

References

[1] M. Q. H. Meng, M. Tao, P. Jiexin, C. Hu, X. Wang, and Y. Chan, "Wireless robotic capsule endoscopy: state-of-the-art and challenges," in *Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788)*, vol. 5566, pp. 555a–5561, 2004.

[2] D. S. Joshua, G. F. Paul, and G. Peter, "Early detection of cancer: past, present, and future," *American Society of Clinical Oncology Educational Book*, vol. 35, no. 35, pp. 57–65, 2015.

[3] O. Onaizah, Z. Koszowska, C. Winters et al., "Guidelines for robotic flexible endoscopy at the time of COVID-19," *Frontiers in Robotics and AI*, vol. 8, article 612852, 2021.

[4] S. Pecere, S. M. Milluzzo, G. Esposito, E. Dilaghi, A. Telese, and L. H. Eusebi, "Applications of artificial intelligence for the diagnosis of gastrointestinal diseases," *Diagnostics*, vol. 11, no. 9, p. 1575, 2021.

[5] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, and P. C. de Groen, "Polyp-alert: near real-time feedback during colonoscopy," *Computer Methods and Programs in Biomedicine*, vol. 120, no. 3, pp. 164–179, 2015.

[6] T. Kanesaka, T. C. Lee, N. Uedo et al., "Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging," *Gastrointestinal Endoscopy*, vol. 87, no. 5, pp. 1339–1344, 2018.

[7] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[8] Y. Li, C. Li, X. Li et al., "A comprehensive review of Markov random field and conditional random field approaches in pathology image analysis," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 609–639, 2022.

[9] H. Chen, C. Li, G. Wang et al., "GasHis-transformer: a multi-scale visual transformer approach for gastric histopathological image detection," *Pattern Recognition*, vol. 130, article 108827, 2022.

[10] C. Joonmyeong, S. Keewon, J. Jinhoon et al., "Convolutional neural network technology in endoscopic imaging: artificial intelligence for endoscopy," *Clinical Endoscopy*, vol. 53, no. 2, pp. 117–126, 2020.

[11] P. Wang, T. M. Berzin, J. R. Glissen Brown et al., "Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study," *Gut*, vol. 68, no. 10, pp. 1813–1819, 2019.

[12] K. Yang, S. Chang, Z. Tian et al., "Automatic polyp detection and segmentation using shuffle efficient channel attention network," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 917–926, 2022.

[13] G. Urban, P. Tripathi, T. Alkayali et al., "Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy," *Gastroenterology*, vol. 155, no. 4, pp. 1069–1078.e8, 2018.

[14] J. Y. Lee, J. Jeong, E. M. Song et al., "Real-time detection of colon polyps during colonoscopy using deep learning:

systematic validation with four independent datasets," *Scientific Reports*, vol. 10, no. 1, p. 8379, 2020.

[15] P. M. Vieira, N. R. Freitas, V. B. Lima, D. Costa, C. Rolanda, and C. S. Lima, "Multi-pathology detection and lesion localization in WCE videos by using the instance segmentation approach," *Artificial Intelligence in Medicine*, vol. 119, article 102141, 2021.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[17] Z. Guo, R. Zhang, Q. Li et al., "Reduce false-positive rate by active learning for automatic polyp detection in colonoscopy videos," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 1655–1658, 2020.

[18] M. Misawa, S. E. Kudo, Y. Mori et al., "Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video)," *Gastrointestinal Endoscopy*, vol. 93, no. 4, pp. 960–967.e3, 2021.

[19] S. Durak, B. Bayram, T. Bakirman et al., "Deep neural network approaches for detecting gastric polyps in endoscopic images," *Medical & Biological Engineering & Computing*, vol. 59, no. 7-8, pp. 1563–1574, 2021.

[20] G. Jocher and A. Stoken, *ultralytics/yolov5: v3.0*, Zenodo, 2020.

[21] I. Pacal, A. Karaman, D. Karaboga et al., "An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets," *Computers in Biology and Medicine*, vol. 141, article 105031, 2022.

[22] A. Howard, M. Sandler, B. Chen et al., "Searching for MobileNetV3," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, 2019.

[23] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, http://arxiv.org/abs/2004.10934.

[24] L. Tsung-Yi, M. Michael, B. Serge et al., "Microsoft COCO: common objects in context," in *European Conference on Computer Vision(ECCV 2014)*, pp. 740–755, Springer International Publishing, Champions, 2014.

[25] Y. Wang, Z. Feng, L. Song, X. Liu, and S. Liu, "Multiclassification of endoscopic colonoscopy images based on deep transfer learning," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 2485934, 12 pages, 2021.

[26] J. Yogapriya, V. Chandran, M. G. Sumithra, P. Anitha, P. Jenopaul, and C. Suresh Gnana Dhas, "Gastrointestinal tract disease classification from wireless endoscopy images using pretrained deep learning model," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 5940433, 12 pages, 2021.

[27] A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher, "A closer look at deep learning heuristics: learning rate restarts, Warmup and Distillation," 2018, http://arxiv.org/abs/1804.03999.

[28] Z. H. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: faster and better learning for bounding box regression," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.

[29] J. Gao, Q. Xiong, C. Yu, and G. Qu, "White-light endoscopic colorectal lesion detection based on improved YOLOv5," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 9508004, 11 pages, 2022.

[30] A. Valada, N. Radwan, and W. Burgard, "Deep auxiliary learning for visual localization and odometry," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6939–6946, 2018.

Hindawi
BioMed Research International
Volume 2022, Article ID 4864485, 6 pages
https://doi.org/10.1155/2022/4864485

Research Article

Application of Deep Learning on the Prognosis of Cutaneous Melanoma Based on Full Scan Pathology Images

Anhai Li,[1] Xiaoyuan Li,[2] Wenwen Li,[3] Xiaoqian Yu,[4] Mengmeng Qi,[5] and Ding Li[1,2]

[1]*Department of Dermatology, Qingdao Huangdao District Central Hospital, Qingdao, Shandong, China*
[2]*Department of Traditional Chinese Medicine, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, China*
[3]*Department of Hematology, Qingdao Women and Children's Hospital, Qingdao, Shandong, China*
[4]*Department of Dermatology, Qingdao Haici Hospital (Qingdao Traditional Chinese Medicine Hospital), Qingdao, Shandong, China*
[5]*Department of Endocrinology, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, China*

Correspondence should be addressed to Ding Li; 907325791@qq.com

Received 20 June 2022; Revised 13 August 2022; Accepted 17 August 2022; Published 28 August 2022

Academic Editor: Chen Li

Copyright © 2022 Anhai Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction. The purpose of this study is to use deep learning and machine learning to learn and classify patients with cutaneous melanoma with different prognoses and to explore the application value of deep learning in the prognosis of cutaneous melanoma patients. *Methods*. In deep learning, VGG-19 is selected as the network architecture and learning model for learning and classification. In machine learning, deep features are extracted through the VGG-19 network architecture, and the support vector machine (SVM) model is selected for learning and classification. Compare and explore the application value of deep learning and machine learning in predicting the prognosis of patients with cutaneous melanoma. *Result*. According to receiver operating characteristic (ROC) curves and area under the curve (AUC), the average accuracy of deep learning is higher than that of machine learning, and even the lowest accuracy is better than that of machine learning. *Conclusion*. As the number of learning increases, the accuracy of machine learning and deep learning will increase, but in the same number of cutaneous melanoma patient pathology maps, the accuracy of deep learning will be higher. This study provides new ideas and theories for computational pathology in predicting the prognosis of patients with cutaneous melanoma.

1. Introduction

Cutaneous melanoma (CM) is the most dangerous type of skin cancer; it is the fifth most common cancer in men and women in the United States [1]. The incidence of CM increases with age [2]. It is estimated that there will be 106110 new patients in 2021 and result in 7180 deaths [3]. Thanks to the application of immunotherapy and targeted therapy drugs, the survival rate of CM patients has been continuously improved [4]. But it still cannot change the poor prognosis of patients with advanced CM [5]. Therefore, the early diagnosis of CM is essential to improve the survival rate of CM patients and save lives, because the survival rate of CM patients depends on the stage of the disease at the time of diagnosis [6].

In recent years, bioinformatics analysis methods based on whole genome sequencing have made great progress in the discovery of CM diagnosis markers, prognosis prediction, CM subtypes classification, and personalized treatment development [7, 8]. However, the expensive cost of whole genome sequencing brings a greater economic and social burden and cannot be widely cited [9]. With the deepening of deep learning research in the field of medicine. New deep learning network structure methods continue to emerge, deep learning can choose different neural network architectures to study and classify different types of medical examination results, including whole genome sequencing result, medical imaging data, pathological picture, common laboratory indicators and patient written electronic health record [10–13]. The research of Rahaman et al. and Chen et al. also

proved the great value of deep learning in clinical direction, and with the deepening of algorithm research, the application value of deep learning in medical treatment is limitless [14–17]. Among them, the VGG-19 network structure was found to have high accuracy in applications for diagnosing retinal diseases and identifying thyroid tumor types [18, 19]. Thus, the development of a new analysis method based on pathological images to predict the prognosis of CM patients based on deep learning is of great significance to the diagnosis and prognosis of CM patients.

This research is based on the CM patient cancer tissue pathology full scan from the TCIA database through deep learning method study and detects whether deep learning can effectively distinguish patients with different prognoses. This research is aimed at discussing the apply value of deep learning in the prediction of CM patient prognosis and offering a new and cheap detection method for the clinical dermatologist.

Table 1: Clinical characteristics of the CM patients.

Parameter	Classification	Individuals	Living	Deceased
Gender	Male (1)	29	23	6
	Female (0)	24	18	6
Age	<60	19	16	3
	≥60-80≤	28	21	7
	>80	5	3	2
BMI	<18.5	0	0	0
	≥18.5-25≤	18	11	7
	>25-30≤	19	17	2
	>30	16	13	3
Stage	Stage I	7	7	0
	Stage II	29	23	6
	Stage III	14	8	6
	Stage IV	3	3	0

2. Materials and Method

2.1. Data Collection and Preprocessing. The cutaneous melanoma pathology images (ID: CPTAC-CM) [20] were downloaded from The Cancer Imaging Archive (TCIA) database (https://www.cancerimagingarchive.net/) [21], including a total of 12 deceased cutaneous melanoma patients (51 melanoma pathological pictures) and 59 living cutaneous melanoma patients (272 melanoma pathological pictures), and divided into deceased cutaneous melanoma patients group (DCM) and living cutaneous melanoma patients group (LCM). In order to better perform deep learning of the pathology map, the svs format of the pathology map is converted to the png format through the reaConverter 7 pro software (https://www.reaconverter.com/).

2.2. VGG-19 Model Construction. In this section, we choose the VGG-19 model as the network architecture and learning model. The VGG-19 model was constructed using Python 3.6 (https://www.python.org/). The patients in our research cohort are also completely randomly assigned. 80% of the total sample is set as the training set, 20% is set as the validation set, and 20% of the training set is set as the test set. This binary classification mode, batch size was set as 2048, epoch was set as 200, weight was chosen as ImageNet model, and activation function was set as sigmoid. The optimizer was selected as Adam (learning rate = 0.01, beta_1 = 0.9, beta_2 = 0.999, epsilon = $1e-08$, decay = 0.0). The 5-fold cross-validation was used to evaluate the accuracy of this model.

2.3. Deep Feature Extraction and SVM Model Construction. In this section, we choose VGG-19 model as the network architecture to extract these pathology scan deep features and to quantify these pathology maps for follow-up research on machine learning; weight was chosen as ImageNet model. In order to better perform subsequent machine learning analysis, filter the acquired depth features before importing them into lasso regression analysis to, and delete the blank values and unqualified depth features. Finally, a support vector machine (SVM) model is established based on the optimal depth features selected, and the prediction model is verified in the test set.

2.4. Statistical Analysis. Statistical analysis was performed using Python 3.6 and GraphPad Prism 9.0 software (https://www.graphpad.com/; San Diego, United States). Using the prognosis results as the gold standard, we calculated the sensitivity and specificity of the machine learning model and deep learning model, plotted the receiver operating characteristic (ROC) curve, and calculated the area under the ROC curve (AUC), thus evaluating the prediction performance of the model.

The chi-square test or Fisher's exact test was used to compare the differences in count data between the two groups, and the independent sample t-test or a nonparametric test was used to compare the differences in measurement data. A value of $P < 0.05$ was considered statistically significant.

3. Result

3.1. CM Patient Detail. There are a total of 53 CM patients after deleting the patients who did not indicate the survival status, including 12 deceased CM patients (51 melanoma pathological pictures) and 41 living CM patients (271 melanoma pathological pictures) (Table 1).

3.2. Diagnostic Performance of Deep Learning Model. In this section, we through VGG-19 network architecture directly learn and classify CM patients with good prognosis and CM patients with poor prognosis. The result of the 5-fold cross-validation showed the ROC fold 1 (AUC = 0.713), the ROC fold 2 (AUC = 0.789), the ROC fold 3 (AUC = 0.811), the ROC fold 4 (AUC = 0.739), the ROC fold 5 (AUC = 0.793), and the mean ROC of 5 ROC fold (AUC = 0.769 ± 0.037). The results suggest that the best AUC for 5-fold cross-validation is fold 3 (Figure 1(a) and Tables 2 and 3).

3.3. Diagnostic Performance of Machine Learning Model. In this section, we first extract a total of 512 deep features from the pathological pictures of CM patients through the VGG-19 neural network and imported 512 deep features into lasso

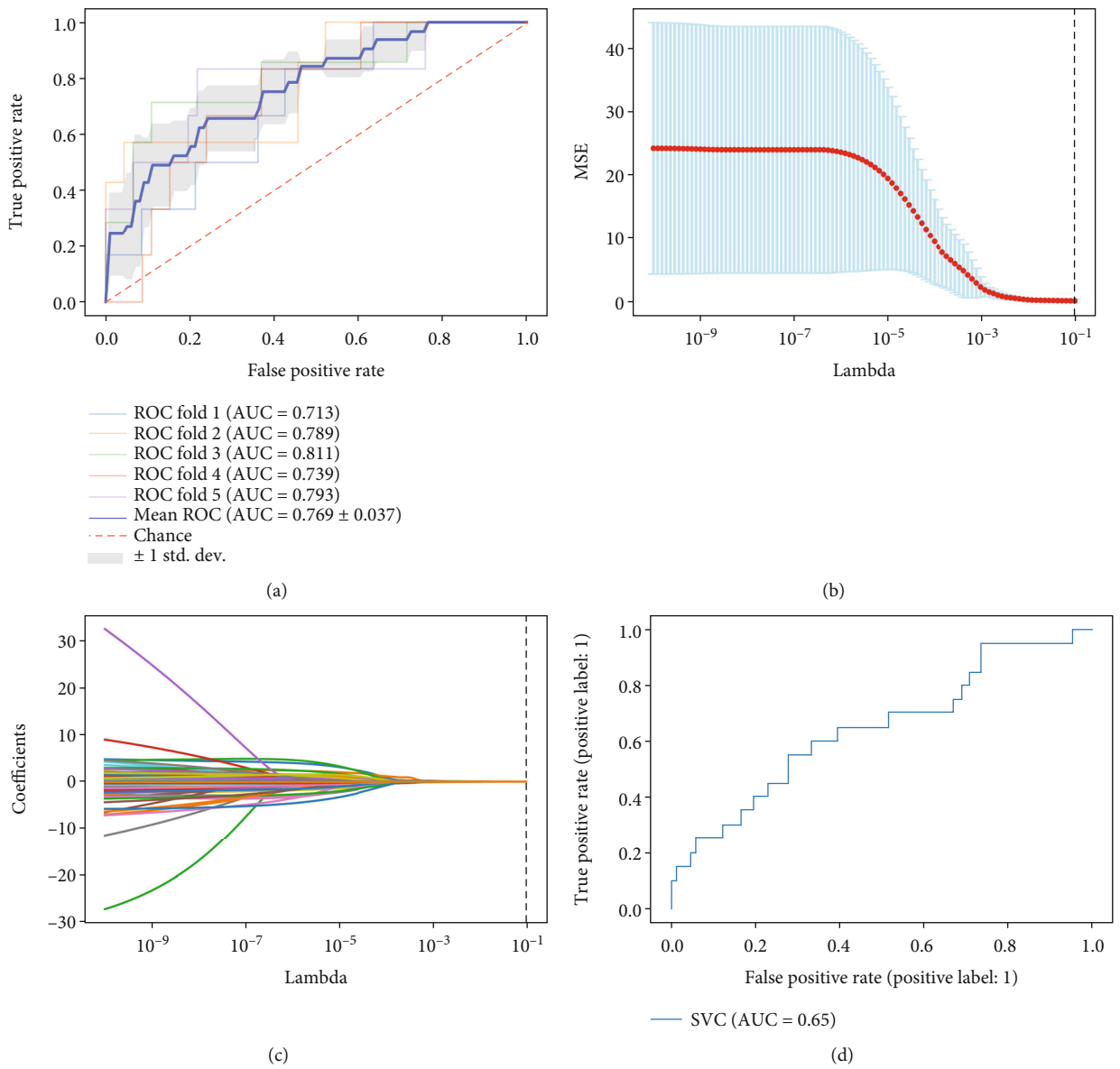


Figure 1: (a) Receiver operating characteristic (ROC) curves and area under the curve (AUC) of the nomogram for the accuracy of deep learning model diagnosis. (b, c) Plots for lasso regression coefficients over different values of the deep feature from pathology scan images. (d) Receiver operating characteristic (ROC) curves and area under the curve (AUC) of the nomogram for the accuracy of machine learning model diagnosis in the testing set.

Table 2: The result of 5-fold cross-validation.

Fold	Classification	Precision	Recall	F1-score	Support
Fold 1 valid	Living	0.838	0.886	0.861	35
	Deceased	0.000	0.000	0.000	6
Fold 2 valid	Living	0.892	0.943	0.917	35
	Deceased	0.500	0.333	0.400	6
Fold 3 valid	Living	0.889	0.914	0.901	35
	Deceased	0.400	0.333	0.364	6
Fold 4 valid	Living	0.838	0.912	0.873	34
	Deceased	0.250	0.143	0.182	7
Fold 5 valid	Living	0.846	0.971	0.904	34
	Deceased	0.500	0.143	0.222	7

Table 3: The accuracy of each fold cross-validation.

Fold	F1-score accuracy	Support
Fold 1 valid	0.756	41
Fold 2 valid	0.854	41
Fold 3 valid	0.829	41
Fold 4 valid	0.780	41
Fold 5 valid	0.829	41

regression analysis for screening. Finally, we build the SVM model through the screened depth features, and the result showed that the AUC in the validation set is 0.65 (Figures 1(b)–1(d)).

4. Discussion

CM is one of the most dangerous types of skin cancer and the main leading cause of cancer-related mortality due its metastatic power and late diagnosis [22]. Thus, predicting the prognosis of melanoma patients in advance is essential to identify CM patients with potentially poor prognoses [23]. How to predict the prognosis of CM patients in advance has also become one of the research hotspots of CM. Bioinformatics analysis methods, metabolomics and intestinal flora, and other methods have proposed methods for predicting the prognosis of CM patients [24–26], which have played an important role in the prognosis research of CM. The application of CNN in clinical medicine provides a new direction and new dimension for the prognosis prediction of CM patients. At present, CNN-related research is mainly based on learning and predicting the skin surface of CM patients, such as the research of Yang et al. and Haenssle et al. [27, 28]. There are only a few studies and reviews on pathological pictures, such as Wang et al.'s article and Laury et al.'s article [13, 29]. There are very few studies based on CNN to predict the prognosis of CM patients by learning from pathological scans.

The histopathological analysis is one of the main gold standards of CM diagnosis, and there are numerous histologic criteria used to diagnose melanoma, but none alone is sufficient to establish CM diagnosis [30]. In order to improve the accuracy of CM diagnosis based on histopathological, multiple researchers offer multiple methods such as immunohistochemical and immunofluorescence [31, 32]. The application of immunohistochemistry and immunofluorescence has improved the accuracy of CM diagnosis to a limited extent, but at the same time, it has also brought an excessively high economic burden. The appearance of deep learning in the medical maybe a better way to solve above questions. Deep learning is a computer model that extracts information from images on a computer. It is possible to extract special data from medical images that are invisible to the human eye and can be used to inform molecular status, prognosis, and treatment sensitivity [33, 34]. Related research has found that deep learning can not only provide the accuracy and objectivity of diagnosis but also reduce the workload and inspection cost of pathologists, because deep learning can learn pathology full scan pictures without the pathologist providing any part of the comment area [35, 36]. It is precise because the integration of deep learning and histopathology has promoted the rapid development of computational pathology. In addition to the routine diagnosis of histopathology, computational pathology also has the ability to identify and extract new features of diseases [37]. Related research results show that computational pathology can effectively predict lymph node metastasis detection and improve breast cancer Ki67 score, prostate cancer Gleason score, and melanoma tumor infiltrating lymphocyte score [38–40].

In this study, we through machine learning and deep learning constructed a two-classification model to study and classify the pathological scan images separately and to compare the application value of two learning methods on the pathological scan images. The accuracy of the two models showed that the accuracy of the CNN model (AUC = 0.769) is significantly higher than that of the SVM model (AUC = 0.65). We initially thought that compared with the SVM model algorithm, the CNN model algorithm is more suitable for predicting the prognosis of CM patients.

This study is based on CNN learning and classifying pathological scan images of CM patients with different prognoses to evaluate the prognostic value of CNN in predicting CM patients. Compared with the conventional prognostic prediction methods at this stage, CNN can let clinicians know the prognosis of CM patients earlier so that targeted treatment of CM patients can be carried out earlier to improve the prognosis of CM patients. Although in this study, the CNN accuracy ratio did not reach a very satisfactory 90% or more, we believe that as the number of subsequent pathological images increases, the accuracy will increase with the increase in the amount of learning.

Data Availability

The data used in this study come from the CPTAC-CM dataset of the TCIA database (https://www.cancerimagingarchive.net/).

Ethical Approval

This study was conducted after the approval of institutional ethical review board.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Anhai Li and Ding Li conceived and designed the experiments. Anhai Li and Ding Li performed the experiments. Xiaoyuan Li, Wenwen Li, and Xiaoqian Yu analyzed and interpreted the data. Mengmeng Qi and Ding Li contributed the materials, analysis tools, or data. Anhai Li and Ding Li wrote the paper. Wenwen Li critically revised its important intellectual content. All authors read and agreed to submit the final version of the manuscript.

References

[1] D. Schadendorf, A. C. J. van Akkooi, C. Berking et al., "Melanoma," *The Lancet*, vol. 392, no. 10151, pp. 971–984, 2018.

[2] S. Ribero, L. S. Stucci, E. Marra et al., "Effect of age on melanoma risk, prognosis and treatment response," *Acta Dermato-Venereologica*, vol. 98, no. 7, pp. 624–629, 2018.

[3] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.

[4] C. Karimkhani, R. Gonzalez, and R. P. Dellavalle, "A review of novel therapies for melanoma," *American Journal of Clinical Dermatology*, vol. 15, no. 4, pp. 323–337, 2014.

[5] K. D. Miller, R. L. Siegel, C. C. Lin et al., "Cancer treatment and survivorship statistics, 2016," *CA: a Cancer Journal for Clinicians*, vol. 66, no. 4, pp. 271–289, 2016.

[6] M. S. Brady, S. A. Oliveria, P. J. Christos et al., "Patterns of detection in patients with cutaneous melanoma," *Cancer*, vol. 89, no. 2, pp. 342–347, 2000.

[7] R. Kamps, R. D. Brandão, B. J. Bosch et al., "Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification," *International Journal of Molecular Sciences*, vol. 18, no. 2, p. 308, 2017.

[8] P. F. Cheng, "Medical bioinformatics in melanoma," *Current Opinion in Oncology*, vol. 30, no. 2, pp. 113–117, 2018.

[9] A. Kozińska, P. Seweryn, and I. Sitkiewicz, "A crash course in sequencing for a microbiologist," *Journal of Applied Genetics*, vol. 60, no. 1, pp. 103–111, 2019.

[10] J. Zou, M. Huss, A. Abid, P. Mohammadi, A. Torkamani, and A. Telenti, "A primer on deep learning in genomics," *Nature Genetics*, vol. 51, no. 1, pp. 12–18, 2019.

[11] G. Eraslan, Ž. Avsec, J. Gagneur, and F. J. Theis, "Deep learning: new computational modelling techniques for genomics," *Nature Reviews. Genetics*, vol. 20, no. 7, pp. 389–403, 2019.

[12] A. Esteva, A. Robicquet, B. Ramsundar et al., "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.

[13] S. Wang, D. M. Yang, R. Rong, X. Zhan, and G. Xiao, "Pathology image analysis using segmentation deep learning algorithms," *The American Journal of Pathology*, vol. 189, no. 9, pp. 1686–1698, 2019.

[14] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[15] J. Zhang, C. Li, S. Kosov et al., "LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, no. 4, article 107885, 2021.

[16] H. Chen, C. Li, X. Li et al., "Gas his-transformer: a multi-scale visual transformer approach for Gastric Histopathology Image Classification," *Pattern Recognition*, vol. 130, article 108827, 2022.

[17] J. Zhang, C. Li, Y. Yin, J. Zhang, and M. Grzegorzek, "Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer," *Artificial Intelligence Review*, vol. 4, pp. 1–58, 2022.

[18] Y. Wang, Q. Guan, I. Lao et al., "Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: a large-scale pilot study," *Annals of Translational Medicine*, vol. 7, no. 18, p. 468, 2019.

[19] K. M. Kim, T. Y. Heo, A. Kim et al., "Development of a fundus image-based deep learning diagnostic tool for various retinal diseases," *Journal of Personalized Medicine*, vol. 11, no. 5, p. 321, 2021.

[20] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), "Radiology data from the clinical proteomic tumor analysis consortium cutaneous melanoma [CPTAC-CM] collection [data set]," *The Cancer Imaging Archive*, 2018.

[21] K. Clark, B. Vendt, K. Smith et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[22] G. C. Leonardi, L. Falzone, R. Salemi et al., "Cutaneous melanoma: from pathogenesis to therapy (review)," *International Journal of Oncology*, vol. 52, no. 4, pp. 1071–1080, 2018.

[23] A. Testori, J. Soteldo, D. Sances et al., "Cutaneous melanoma in the elderly," *Melanoma Research*, vol. 19, no. 3, pp. 125–134, 2009.

[24] D. Donnelly 3rd, P. P. Aung, and G. Jour, "The "-OMICS" facet of melanoma: heterogeneity of genomic, proteomic and metabolomic biomarkers," *Seminars in Cancer Biology*, vol. 59, pp. 165–174, 2019.

[25] B. Huang, W. Han, Z. F. Sheng, and G. L. Shen, "Identification of immune-related biomarkers associated with tumorigenesis and prognosis in cutaneous melanoma patients," *Cancer Cell International*, vol. 20, no. 1, p. 195, 2020.

[26] A. Limeta, B. Ji, M. Levin, F. Gatto, and J. Nielsen, "Meta-analysis of the gut microbiota in predicting response to cancer immunotherapy in metastatic melanoma," *JCI Insight*, vol. 5, no. 23, article e140940, 2020.

[27] C. H. Yang, J. H. Ren, H. C. Huang, L. Y. Chuang, and P. Y. Chang, "Deep hybrid convolutional neural network for segmentation of melanoma skin lesion," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 9409508, 15 pages, 2021.

[28] H. A. Haenssle, C. Fink, R. Schneiderbauer et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.

[29] A. R. Laury, S. Blom, T. Ropponen, A. Virtanen, and O. M. Carpén, "Artificial intelligence-based image analysis can predict outcome in high- grade serous carcinoma via histology alone," *Scientific Reports*, vol. 11, no. 1, article 19165, 2021.

[30] B. A. Shoo, R. W. Sagebiel, and M. Kashani-Sabet, "Discordance in the histopathologic diagnosis of melanoma at a melanoma referral center," *Journal of the American Academy of Dermatology*, vol. 62, no. 5, pp. 751–756, 2010.

[31] M. S. Dinehart, S. M. Dinehart, S. Sukpraprut-Braaten, and W. A. High, "Immunohistochemistry utilization in the diagnosis of melanoma," *Journal of Cutaneous Pathology*, vol. 47, no. 5, pp. 446–450, 2020.

[32] D. Hartmann, S. Krammer, S. Vural et al., "Immunofluorescence and confocal microscopy for ex-vivo diagnosis of melanocytic and non-melanocytic skin tumors: a pilot study," *Journal of Biophotonics*, vol. 11, no. 3, article e201700211, 2018.

[33] A. B. Levine, C. Schlosser, J. Grewal, R. Coope, S. J. M. Jones, and S. Yip, "Rise of the machines: advances in deep learning for cancer diagnosis," *Trends in Cancer*, vol. 5, no. 3, pp. 157–169, 2019.

[34] D. Komura and S. Ishikawa, "Machine learning approaches for pathologic diagnosis," *Virchows Archiv*, vol. 475, no. 2, pp. 131–138, 2019.

[35] Y. Jiang, M. Yang, S. Wang, X. Li, and Y. Sun, "Emerging role of deep learning-based artificial intelligence in tumor pathology," *Cancer Communications*, vol. 40, no. 4, pp. 154–166, 2020.

[36] P. Courtiol, C. Maussion, M. Moarii et al., "Deep learning-based classification of mesothelioma improves prediction of

patient outcome," *Nature Medicine*, vol. 25, no. 10, pp. 1519–1525, 2019.

[37] B. Acs, M. Rantalainen, and J. Hartman, "Artificial intelligence as the next step towards precision pathology," *Journal of Internal Medicine*, vol. 288, no. 1, pp. 62–81, 2020.

[38] F. Klauschen, S. Wienert, W. D. Schmitt et al., "Standardized Ki67 diagnostics using automated scoring–clinical validation in the Gepar Trio breast cancer study," *Clinical Cancer Research*, vol. 21, no. 16, pp. 3651–3657, 2015.

[39] A. Uryvaev, M. Passhak, D. Hershkovits, E. Sabo, and G. Bar-Sela, "The role of tumor-infiltrating lymphocytes (TILs) as a predictive biomarker of response to anti-PD1 therapy in patients with metastatic non-small cell lung cancer or metastatic melanoma," *Medical Oncology*, vol. 35, no. 3, p. 25, 2018.

[40] J. I. Epstein, M. J. Zelefsky, D. D. Sjoberg et al., "A contemporary prostate cancer grading system: a validated alternative to the Gleason score," *European Urology*, vol. 69, no. 3, pp. 428–435, 2016.

Hindawi
BioMed Research International
Volume 2022, Article ID 3168503, 17 pages
https://doi.org/10.1155/2022/3168503

Research Article

Computer Image Analysis Reveals C-Myc as a Potential Biomarker for Discriminating between Keratoacanthoma and Cutaneous Squamous Cell Carcinoma

Xinyun Fan,[1,2,3] **Xueli Niu**,[1,2,3] **Ze Wu**,[4] **Lu Yao**,[1,2,3] **Shirui Chen**,[5] **Wenyu Wan**,[1,2,3] **Bo Huang**,[5] **Rui-Qun Qi**,[1,2,3] **and Tao Zhang**[6]

[1]*Key Laboratory of Immunodermatology, Ministry of Education, Department of Dermatology, The First Hospital of China Medical University, Shenyang 110001, China*
[2]*Key Laboratory of Immunodermatology, National Health Commission of the People's Republic of China, The First Hospital of China Medical University, Shenyang 110001, China*
[3]*National and Local Joint Engineering Research Center of Immunodermatological Theranostics, The First Hospital of China Medical University, Shenyang 110001, China*
[4]*Department of Dermatology, General Hospital of Northern Theater Command, Shenyang 110000, China*
[5]*Department of Pathology, Cancer Hospital of China Medical University, Liaoning Cancer Hospital & Institute, Shenyang, 110042 Liaoning Province, China*
[6]*Department of Stem Cells and Regenerative Medicine, Shenyang Key Laboratory of Stem Cell and Regenerative Medicine, China Medical University, Shenyang 110122, China*

Correspondence should be addressed to Bo Huang; huangbo@cancerhosp-ln-cmu.com, Rui-Qun Qi; xiaoqiliumin@163.com, and Tao Zhang; cmuzhangtao@163.com

Received 27 April 2022; Revised 10 July 2022; Accepted 25 July 2022; Published 23 August 2022

Academic Editor: Chen Li

Copyright © 2022 Xinyun Fan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The distinction between Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC) is critical yet usually challenging to discriminate clinically and histopathologically. One approach to differentiate KA from cSCC is through assessing the immunohistochemical staining patterns of the three indicators, β-catenin, C-Myc, and CyclinD1, which are critical molecules that play important roles in the Wnt/β-catenin signaling pathway. Ki-67, as a proliferation biomarker for human tumor cells, was also assessed as an additional potential marker for differentiating KA from cSCC. In this report, these four indicators were analyzed in 42 KA and 30 cSCC cases with the use of the computer automated image analysis system. Computer automated image analysis is a time-based and cost-effective method of determining IHC staining in KA and cSCC samples. We found that C-Myc staining was predominantly localized in the nuclei of basal cells within KA patients, whereas cSCC staining was predominantly localized in the nuclei of diffuse cells. This C-Myc staining pattern has a sensitivity of 78.6% and a specificity of 66.7% for identifying KA. Moreover, positive rates of distinct expression patterns of C-Myc and Ki-67 may also serve as a means to clinically distinguish KA from cSCC. Taken together, our results suggest that these markers, in particular C-Myc, may be useful in differentiating KA from cSCC.

1. Introduction

Keratoacanthoma (KA) is a proliferative, fast-growing squamous epithelial skin tumor which shows a spontaneous regression. It most commonly occurs in the elderly, mainly in sun-exposed areas, with the face being particularly vulnerable, while multiforms and rash types are less common. According to the clinical manifestation of KA, the biological development of KA can be divided into proliferative, mature, and regressing stages. In all of these cases, there is a tendency

to progress to a well-differentiated Cutaneous Squamous Cell Carcinoma (KA-cSCC) and KA with complete progression to cSCC [1].

KA shares many clinical and histological features with the well-differentiated cutaneous condition of Cutaneous Squamous Cell Carcinoma (cSCC). In specific, the morphological and histopathological characteristics of KA are very similar to those of cSCC, making it difficult to distinguish between these two conditions. In fact, some clinicians consider KA to be a variant of cSCC [2], which contributes to the controversy regarding the definition and diagnosis of KA. However, genomic differences, prognosis, and metastasis rates that exist between KA and cSCC suggest that these two conditions do not belong to the same category [3–5]. Moreover, reports showing a high incidence of KA (409/100,000 person-years) concentrated among residents of Queensland, Australia [6] also suggest that notable differences are present between KA and cSCC. Alternatively, such findings may likely grossly underestimate the incidence of KA due to the misclassification of these lesions as well-differentiated cSCC, underreporting, and/or spontaneous regression of KA prior to diagnosis. Given these issues, a timely and accurate differential diagnosis between these two conditions is crucial.

An important factor which can aid in such a differential diagnosis is the identification of specific, discriminatory clinicopathological biomarkers. Some clinicians have recommended that Ki-67 may possess this potential as it shows a positive percent of >20% in cSCC and ≤20% in KA [7, 8]. It has also been reported that MYC gene copy number aberrations are more common in cSCC [9, 10]. MYC genes include C-Myc, N-Myc, and L-Myc, among which C-Myc represents one of the key target genes in the Wnt/β-catenin signaling pathway [11]. Activation of the Wnt/β-catenin signaling pathway is closely associated with the development and metastasis of different tumors, including skin, colon, and breast cancers [12, 13]. In addition, in this Wnt/β-catenin signaling pathway, β-catenin is essential for controlling hair follicle formation [14–16] and the three pathological cycles of KA are inextricably linked to the cyclic growth process after hair follicle formation. These relationships have led to the hypothesis that KA originates from hair follicles and may be a follicle-associated tumor, with the spontaneous regression of KA undergoing an apoptotic process similar to that of hair follicle degeneration [17, 18]. In support of this hypothesis are findings in the chemically induced KA mouse model which show that the Wnt/β-catenin signaling pathway was functionally active during the growth phase of the KA tumor and that retinoic acid- (RA-) mediated Wnt inhibition was sufficient to induce spontaneous regression of KA [19].

The diverse capacities of the Wnt/β-catenin signaling pathway in the development of cSCC and KA led us to conjecture that key signature molecules within this pathway may play different roles and show different expression patterns in KA versus cSCC. Accordingly, it may be possible to identify and isolate specific diagnostic biomarkers to assist in the differentiation of KA from cSCC. Given this potential, we selected critical molecules within the Wnt/β-catenin signaling pathway cascade, including β-catenin, C-Myc, and CyclinD1, for analyses (Figure 1). Ki-67, being one of the most commonly used molecules in immunohistochemical studies to mark cell cycle regulation and proliferation ability [20], was also assessed. Expression patterns of these potential markers were evaluated in samples from KA and cSCC patients, using computer automated image analysis software to assess whether any differences in their protein expression profiles would be present within these two types of cancer. With these analyses, it is possible to describe the specific molecular and functional changes during tumor progression and initially explore the notion that KA and cSCC are distinct squamous epithelial proliferative skin tumors, as well as to determine whether specific molecules can be used as biomarkers for the differentiation of KA versus cSCC. Such information will be critical for the stimulation of new directions in research needed for the development of future targeted therapies.

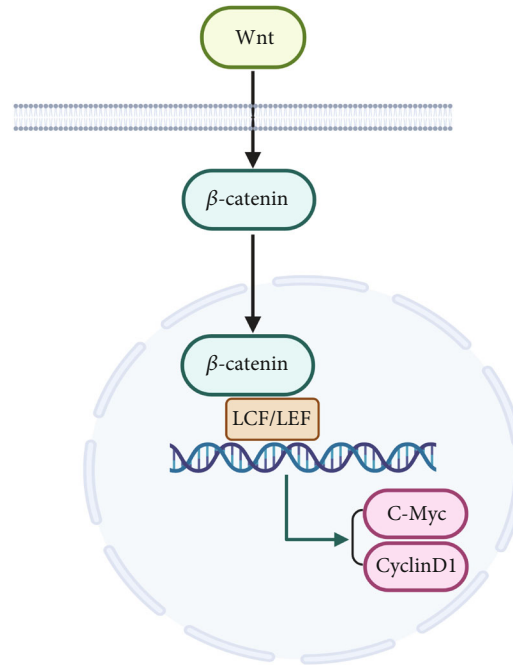


Figure 1: The Wnt/β-catenin signaling pathway.

Pathology evaluation represents the "gold standard" for tumor diagnosis. With traditional pathology assessments, the pathologist's subjective decision is based on microscopic observations, with the pathology image information stored on glass slides. Unfortunately, the accuracy involved with such procedures is not ideal and the information stored on the glass slides leads to difficulties in information processing and transfer. Moreover, the image information interpretation software currently available is relatively inconvenient. However, pathology slides can be fully digitalized with the use of whole slide imaging (WSI) technology, which utilizes the immense computing power of computers to perform computer-aided diagnosis (CAD). CAD is a technique that can quantify the shape, size, and color information of pathology images, as well as enable image retrieval, pattern recognition, computer learning, and deep learning. In this

way, it is possible to construct CAD mathematical models and extract corresponding quantitative features and/or identify a specific region of interest (ROI) within a pathology image. Such quantitative features or regions obtained with CAD represent potentially important tools to assist pathologists in quantitatively evaluating indicators to achieve a more rapid, accurate, and highly reproducible medical diagnosis [21–24]. Our current experiments involving the use of inForm® automated image analysis software have indicated that this provides an accurate means to quantify biomarkers in tissue sections. Therefore, in this report, we utilized this computer analysis to identify potential biomarker differences between samples from KA and cSCC patients and, in this way, provide a more robust protocol to discriminate between these two conditions.

2. Methods

2.1. Subjects. Surgical removal of skin KA (42 cases) and well-differentiated cSCC (30 cases) samples was performed at the First Hospital of China Medical University over the period from January 2016 to May 2021. None of these patients received radiotherapy, chemotherapy, and/or biological therapy prior to surgery. All cases were assessed and confirmed independently by two pathologists. Informed consent was obtained from patients according to the Declaration of Helsinki, and the Ethics Committee of the First Hospital of China Medical University approved these experiments.

2.2. Histology and Immunohistochemistry. Formalin-fixed, paraffin-embedded sections were cut into 4 μm sections. The sections were deparaffinized and rehydrated using xylene and graded alcohol solutions. They were blocked with 2% bovine serum albumin and incubated with specific primary antibodies for 12 hours at 4°C, followed by incubation with a biotinylated secondary antibody. Subsequently, diaminobenzidine (DAB) was added dropwise over a 3- to 5-minute period and sections were then counterstained with hematoxylin. The primary antibodies included rabbit polyclonal antibodies of CyclinD1 (dilution 1 : 200, cat. no.: ab134175, Abcam), β-catenin (dilution 1 : 500, cat. no.: ab32572, Abcam), Ki-67 (dilution 1 : 500, cat. no.: ab92742, Abcam), and C-Myc (cat. no.: GT220607, Gene Tech).

2.3. Sample Assessment. Images were generated using the Vectra™ 3.0 multiplexing image technology platform (Akoya Biosciences, DE, USA). The entire pathology section was scanned for 10x photos, then labeled using the Phenochart Whole Slide Viewer (Akoya Biosciences, DE, USA), with each tissue core scanned for production of 20x photos. The inForm® 2.4 image analysis software program (Akoya Biosciences, DE, USA) has a powerful self-learning feature enabling the identification and segmentation of specific tissue types through a "circle training" approach that allows for the measurement or scoring of biomarker expression in specific tissues, as well as within nuclear, cytoplasmic, and cytosolic regions. With this program, it is possible to convert subjective interpretational assessments of the pathology within sections into accurate and consistent objective morphological records.

A brief description of the computer software analysis steps is contained in Figure 2 with β-catenin and Ki-67 expressions in cSCC samples presented as examples and presented in Figure 3.

2.4. Statistical Analysis. All data were analyzed using the SPSS 23.0 statistical software program (IBM Corporation, Armonk, NY, USA). Results are presented as the means ± standard deviations (SD). Student's t-tests, chi-squared tests, or one-way ANOVAs were used for these analyses. Graphs were generated using GraphPad Prism 9.0 software (GraphPad Software, Inc., San Diego, CA), Visio Drawing (Microsoft Corporation, SEA, USA), and BioRender® software (USA) with BioRender® software publication agreement number VB23ZIU9YD. A $P < 0.05$ was required for results to be considered statistically significant.

3. Results

3.1. Patient Characteristics. A summary of the demographic and clinicopathological data of patients included in this study is contained in Table 1. No statistically significant differences were obtained between the KA and cSCC patients or between the three subtypes of KA (mature KA, KA-cSCC, and KA degenerative) in terms of age, sex, and site of tumor onset (Supplemental Table 1). There was a significantly greater prevalence of KA versus cSCC at the site of sun-exposed tissue, a result that is consistent with previous findings regarding the prevalence of KA at sun-exposed sites ($P = 0.000$).

3.2. β-Catenin Expression in KA and cSCC Patients. The expression of β-catenin in KA and cSCC patients was found to be located in cell membranes and/or the cytoplasm of the tumor island (Figure 4(a)). β-Catenin expression, with Score 2+, was significantly greater in cSCC versus KA patients ($P_{2+} = 0.026$, Figure 4(b)). β-Catenin H-Scores were not significantly different between the KA and cSCC samples ($P = 0.062$, Figure 4(c)). Results of β-catenin expression, obtained with crosstab comparisons among the three subtypes of KA, revealed that KA-regressing with Score 3+ was significantly greater than the KA-well developed and KA-cSCC subtypes ($P_{3+} = 0.035$, Figure 4(d)). Moreover, results of the within-group analysis revealed that the KA-regressing subtype was significantly greater than the KA-well developed and KA-cSCC subtypes ($P = 0.011$), while KA-cSCC, with Score 3+, was significantly greater than the KA-regressing subtype ($P = 0.019$, Figure 4(e)).

3.3. C-Myc Expression in KA and cSCC Patients. C-Myc expression in KA cases was mainly located in the nuclei of basal cells at the initial portion of tumor infiltration, with a few KA cases showing positive C-Myc expression within nuclei of suprabasal cells. C-Myc expression was found in the nuclei and/or cytosol of most cells in cSCC patients, with this expression being significantly greater in these regions as compared with that in other regions (Figures 5(a) and 5(b)). The C-Myc expression pattern within samples from KA

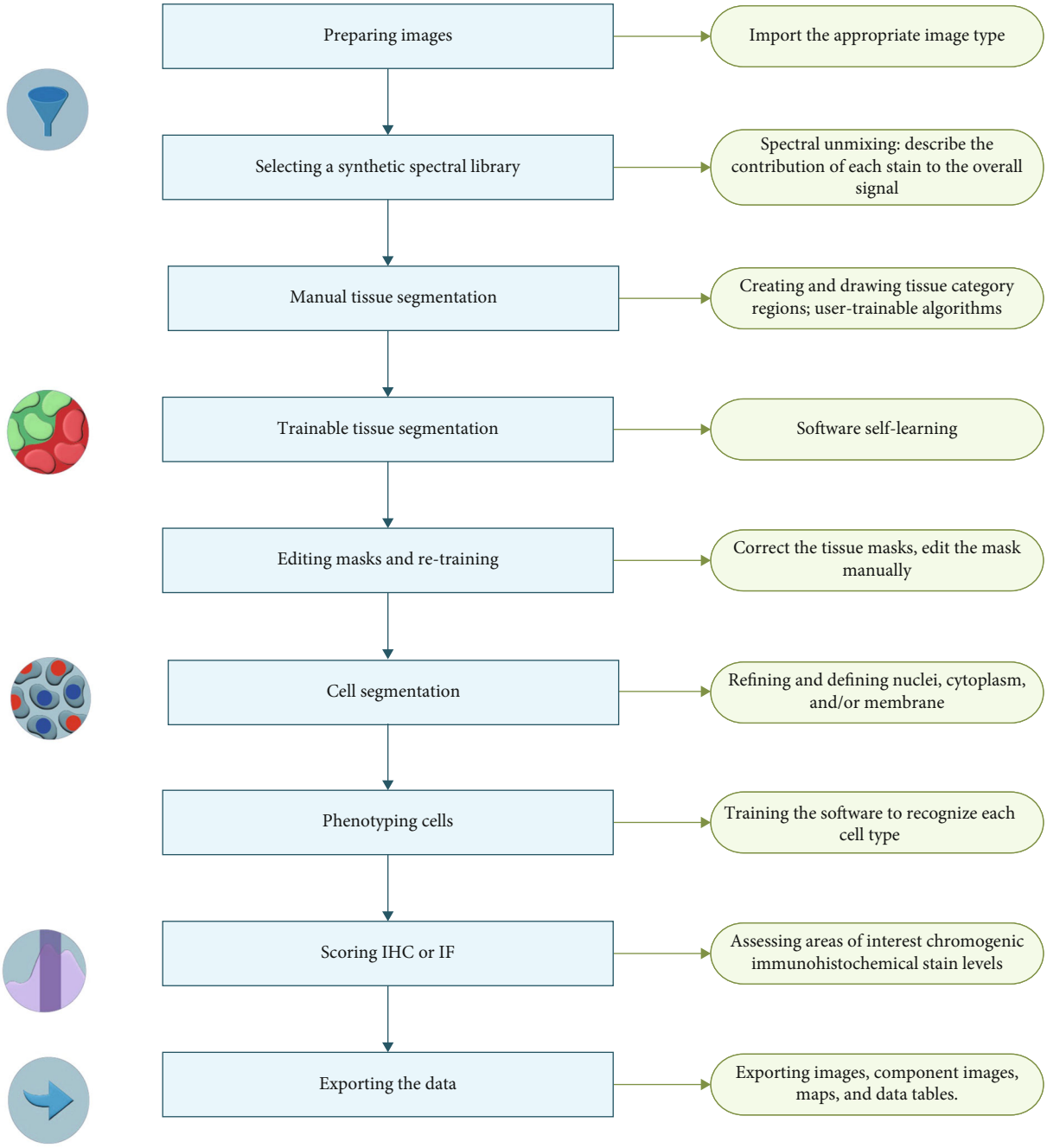


Figure 2: The InForm® image analysis process.

patients was significantly greater than that from cSCC patients ($P = 0.000$, Table 2), with significantly increased C-Myc Scores 0 and 2+ ($P_0 = 0.034$ and $P_{2+} = 0.018$, respectively, Figure 5(c)) and H-Scores ($P = 0.024$, Figure 5(d)). No statistically significant differences in C-Myc expression were obtained among the three subtypes of KA patients (Figure 5(e)).

3.4. CyclinD1 Expression in KA and cSCC Patients. CyclinD1 expression in KA patients showed an intense amount of positive staining within basal cells and/or portions of suprabasal cell nuclei and corresponding cytoplasm, while samples from cSCC patients showed diffuse positive staining within the nucleus and cytoplasm (Figure 6(a)). The overall expression of CyclinD1 was significantly greater in KA versus cSCC patients, with Score 1+ ($P_{1+} = 0.041$, Figure 6(b)). Results of the H-Score and other comparisons as performed among the three KA subtypes failed to show any statistically significant differences (Figures 6(c) and 6(d)).

3.5. Ki-67 Expression in KA and cSCC Patients. Ki-67 expression within samples from KA patients was mainly localized in the nuclei of basal cells at the initial portions of tumor infiltration, with this expression being observed in the nuclei

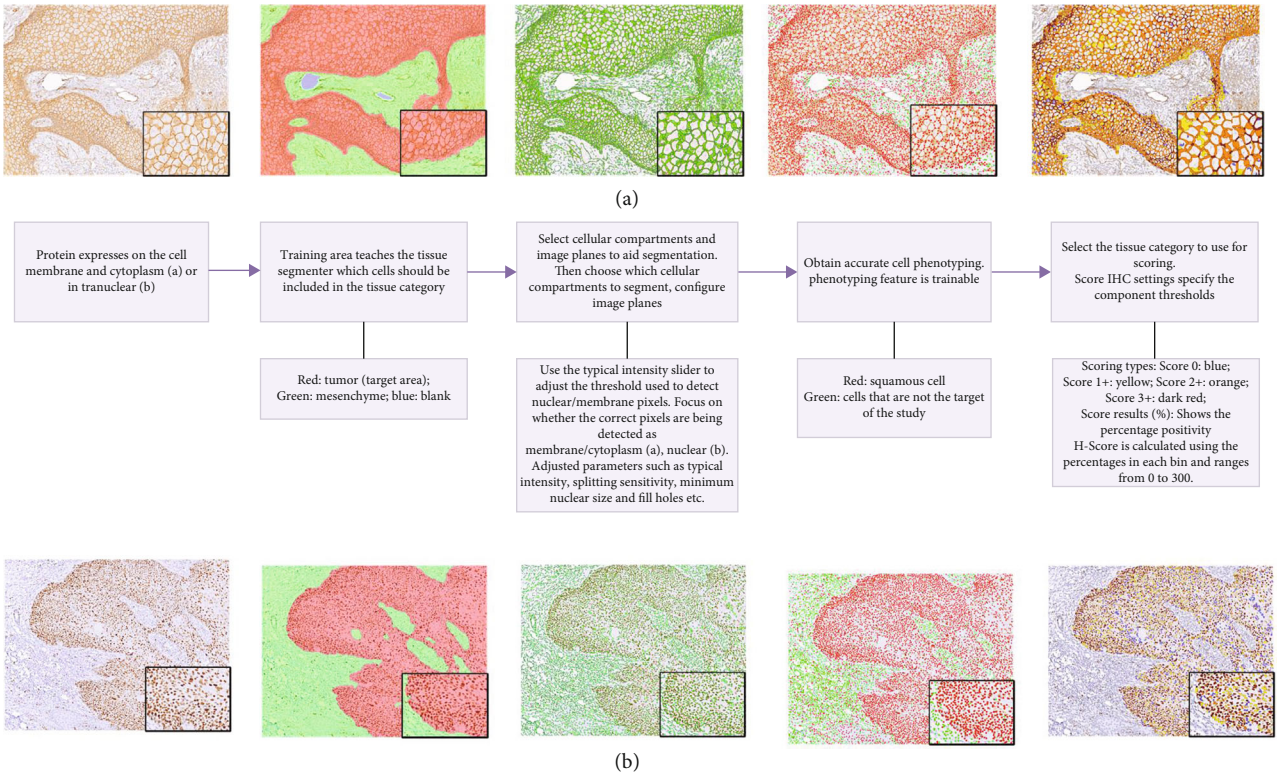


Figure 3: Description of the software analysis process as presented with two examples.

Table 1: Summary of clinical characteristics of Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC) patients.

	KA	cSCC	*P* value
Patients	42	30	
Age (years)	60.67 ± 11.29	60.97 ± 9.84	0.907
Gender (*n*, %)			0.690
Female	19 (45.24%)	15 (50.0%)	
Male	23 (54.80%)	15 (50.0%)	
Site (*n*, %)			0.000
Head and neck	35 (83.3%)	11 (36.7%)	
Body	4 (9.5%)	11 (36.7%)	
Arms and legs	39 (7.1%)	8 (26.7%)	

of only a few basal cells, while in cSCC patients, this Ki-67 expression was localized within the nuclei of most cells in the tumor region (Figure 7(a)). Statistically significant Ki-67 Scores 0 to 3+ ($P_0 = 0.000$, $P_{1+} = 0.011$, $P_{2+} = 0.005$, and $P_{3+} = 0.000$, Figure 7(b)) and *H*-Scores ($P = 0.000$, Figure 7(c)) were obtained in samples from KA versus cSCC patients. No statistically significant differences were obtained among the three KA subtypes (Figure 7(e)). The mean and percent of positive Ki-67 DAB in KA samples were significantly greater than those in cSCC samples, results which were consistent with previous findings ($P = 0.000$, Figures 8(a) and 8(b)).

3.6. Ki-67 and C-Myc in KA Are Correlated and Consistent. The C-Myc *H*-Score was positively correlated with the Ki-67 *H*-Score in KA patients (Figures 9(a) and 9(b)). These findings of a positive correlation between Ki-67 and C-Myc prompted us to assess the diagnostic sensitivity and specificity of these two markers. We utilized the suggestion of diagnosing Ki-67 based on a positive percent of >20% as cSCC and ≤20% as KA [7, 8]. Moreover, we assigned a diagnosis of KA when the C-Myc expression mode was located in the nucleus of the basal cells at the initial portion of tumor infiltration and used an opposite criterion for cSCC. Moreover, similar diagnostic sensitivities and specificities were obtained when either tandem or parallel diagnostic methods were applied for KA and cSCC. The specificity for a diagnosis of KA with a Ki-67 of ≤20% was 96.7%. Based on the presence of the statistically significant expressions of C-

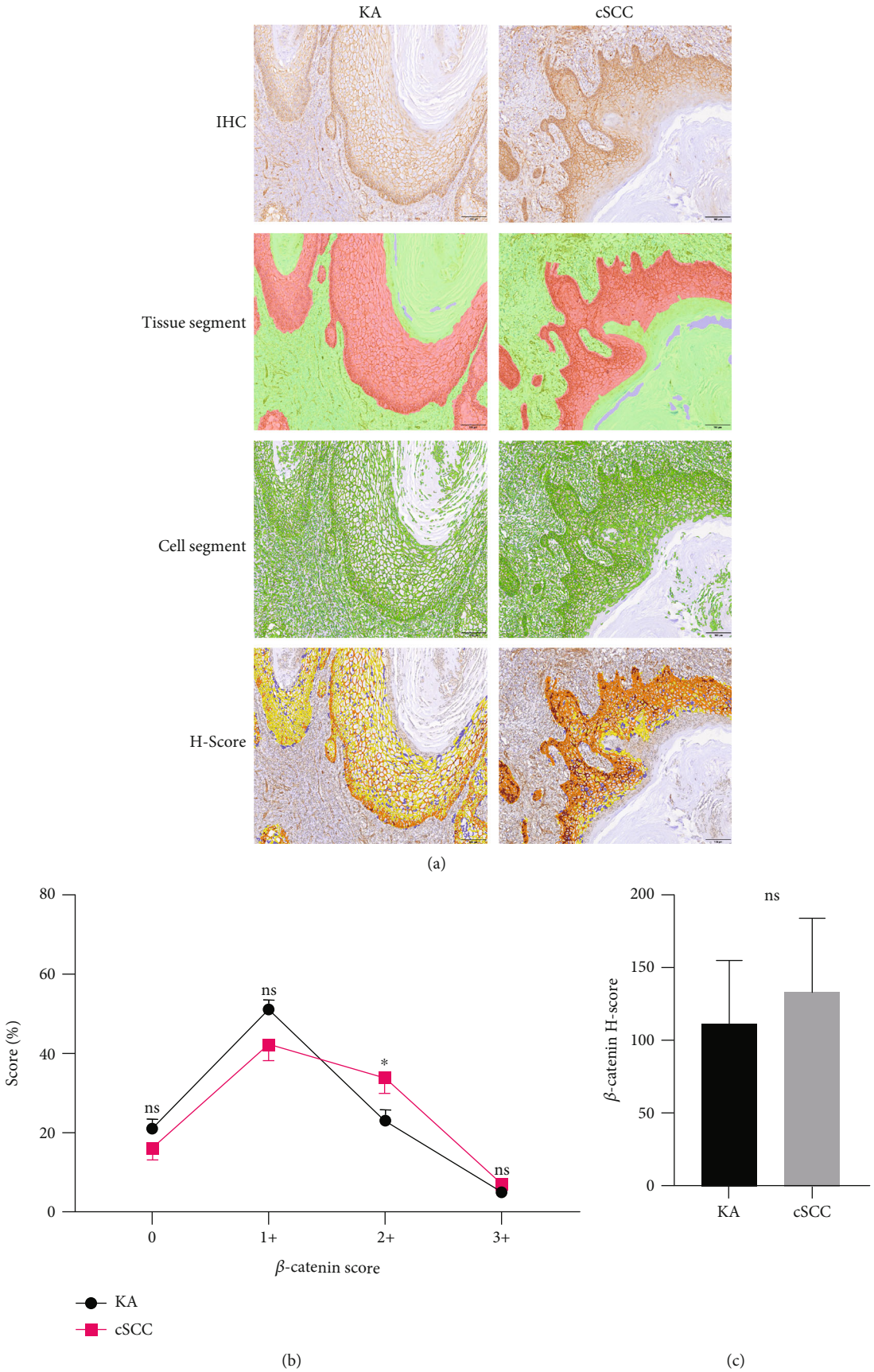


Figure 4: Continued.

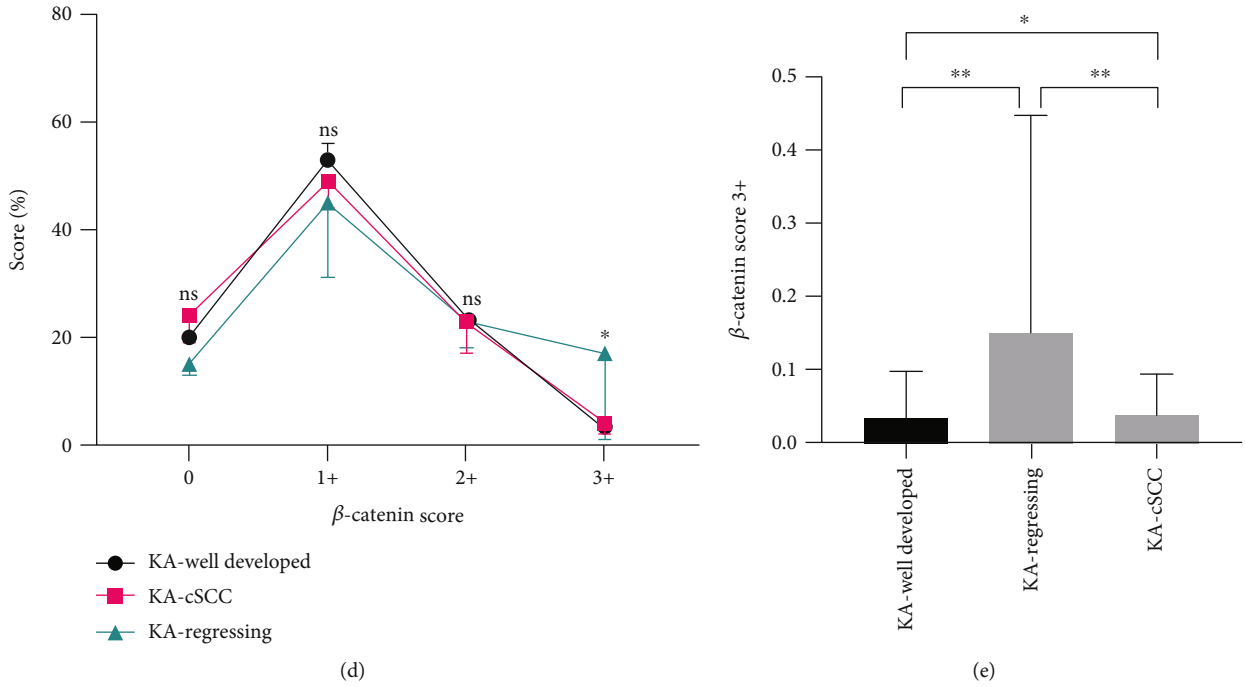


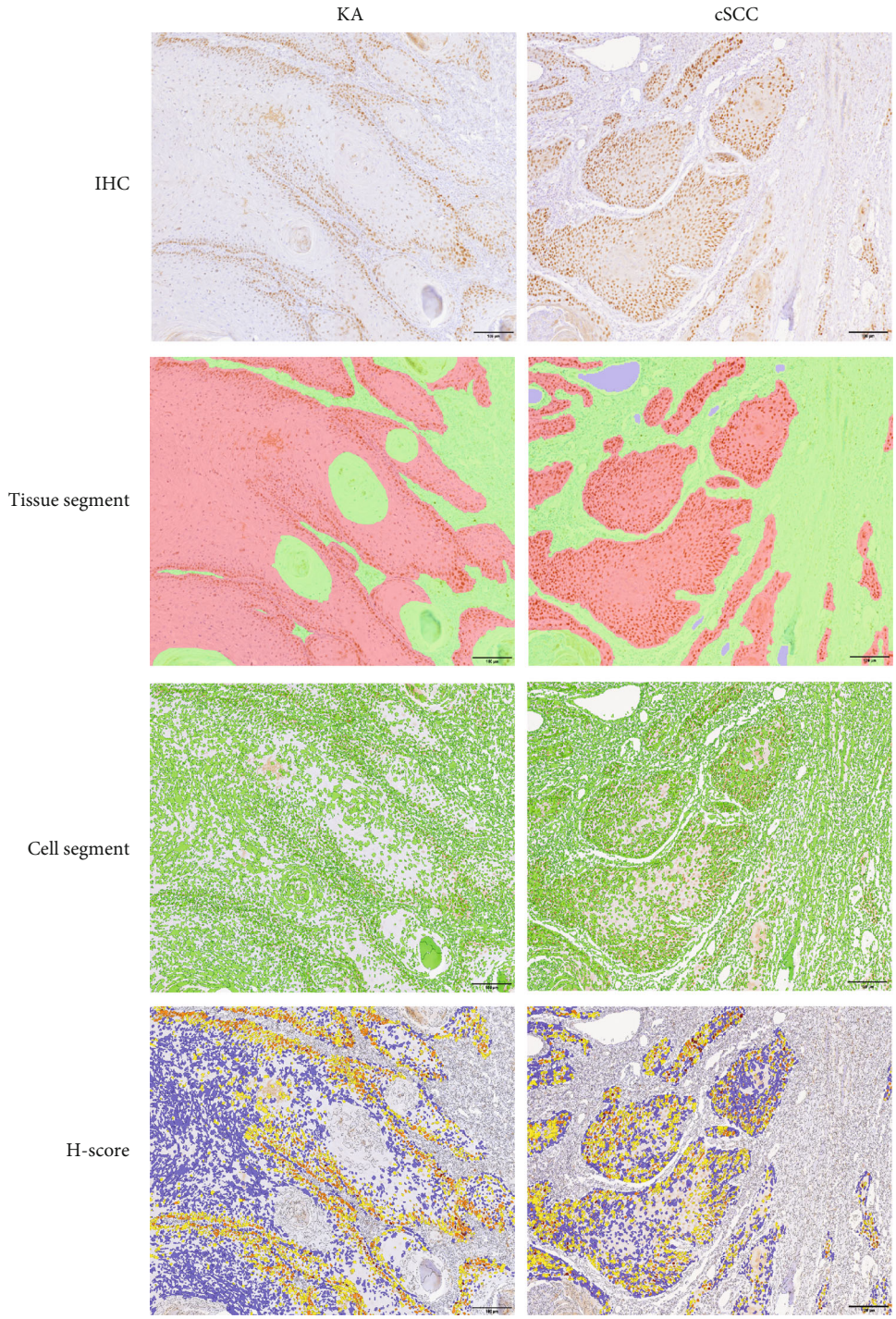
Figure 4: Expression and analysis of β-catenin in Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC). (a-1) Immunohistochemical staining. β-Catenin in KA samples was located in cell membranes and in some portions of the cell plasma, and staining in cSCC samples revealed that it was located in cell membranes and in relatively greater portions of cell plasma (20x). (a-2) Tissue area delineation (red: tumor target area; green: mesenchyme; and blue: blank). (a-3) Cytoarchitectonic delineation (green). (a-4) β-catenin DAB Score delineation (Score 0: blue; Score 1+: yellow; Score 2+: orange; and Score 3+: deep red). (b) Percent of Scores 0, 1+, 2+, and 3+ in classes of KA and cSCC samples based on the intensity of β-catenin DAB positivity. β-Catenin Score 2+ ($P = 0.026$). (c) β-catenin H-Scores in KA versus cSCC samples based on intensities of β-catenin DAB positivity ($P = 0.062$). (d, e) Cross-comparisons of β-catenin Score in the three subtypes of KA (KA-well developed, KA-cSCC, KA-regressing; β-catenin Score 3+ $P = 0.035$); β-catenin Score 3+ in KA vs. KA regressing ($P = 0.011$); KA-cSCC vs. KA regressing ($P = 0.019$).

Myc obtained in the KA and cSCC patients in this report, a diagnosis of KA with a specificity of 66.70% and sensitivity of 78.6% was generated when the C-Myc expression was present within the nucleus of basal cells at the initial portion of tumor infiltration. Further analysis revealed that the parallel diagnosis specificity was 66.70% and the sensitivity was 88.0%, which further improved the sensitivity of a concatenated versus a single diagnosis (Table 3).

4. Discussion

The results of this study reveal some noteworthy differences in the pattern of C-Myc expression between the KA and cSCC patients. To our knowledge, these findings represent the first demonstration that in samples from KA patients, C-Myc was predominantly expressed in the nuclei of basal cells located at the initial portion of tumor infiltration, along with a lower amount of expression in suprabasal cell nuclei. In contrast, C-Myc was expressed in the nuclei and/or cytoplasm in most of the cells of samples from cSCC patients. Such findings suggest that C-Myc exerts distinct activities in cSCC versus KA. Ki-67, a more established immunohistochemical biomarker of cell proliferation, provides a "tool" for quantifying tumor proliferative capacity [20]. Based on these statistical analyses performed, we found that C-Myc and Ki-67 expression patterns were correlated and consistent in KA. Interestingly, in the presence of positive Ki-67 and C-Myc expression patterns, there was an increase in the sensitivity of the diagnosis as compared with that obtained when the diagnosis was generated with separate expressions of each.

C-Myc, a critical transcriptional activator in the Wnt/β-catenin signaling pathway, acts as a proto-oncogene, regulating cell proliferation, differentiation, and apoptosis. The C-Myc proto-oncogene has been shown to be amplified and overexpressed in a variety of human malignancies, including head and neck cSCC [11]. With the use of a conditional transgene expression system (C-MycER™) to study the effect of conversion of C-Myc activation in different tissues of adult animals, it was found that in the mouse epidermal basal layer, C-Myc activation was sufficient to induce mitotic keratinocytes to enter the cell cycle. Sustained activation of C-Myc for 21 days results in epidermal hyperplasia with localized dysplasia, which creates papillomatous lesions resembling human precancerous skin lesions called actinic keratoses—the forerunner to cSCC. With C-MycER™ inactivation, these precancerous lesions spontaneously regress [25]. When collating these findings, it appears that C-Myc activation in the epidermal basal layer is essential for the maintenance of papillomatosis and angiogenesis and suggests a possible role of C-Myc activation and inactivation in skin tumor "reversal." Even greater levels of C-Myc



(a)

Figure 5: Continued.

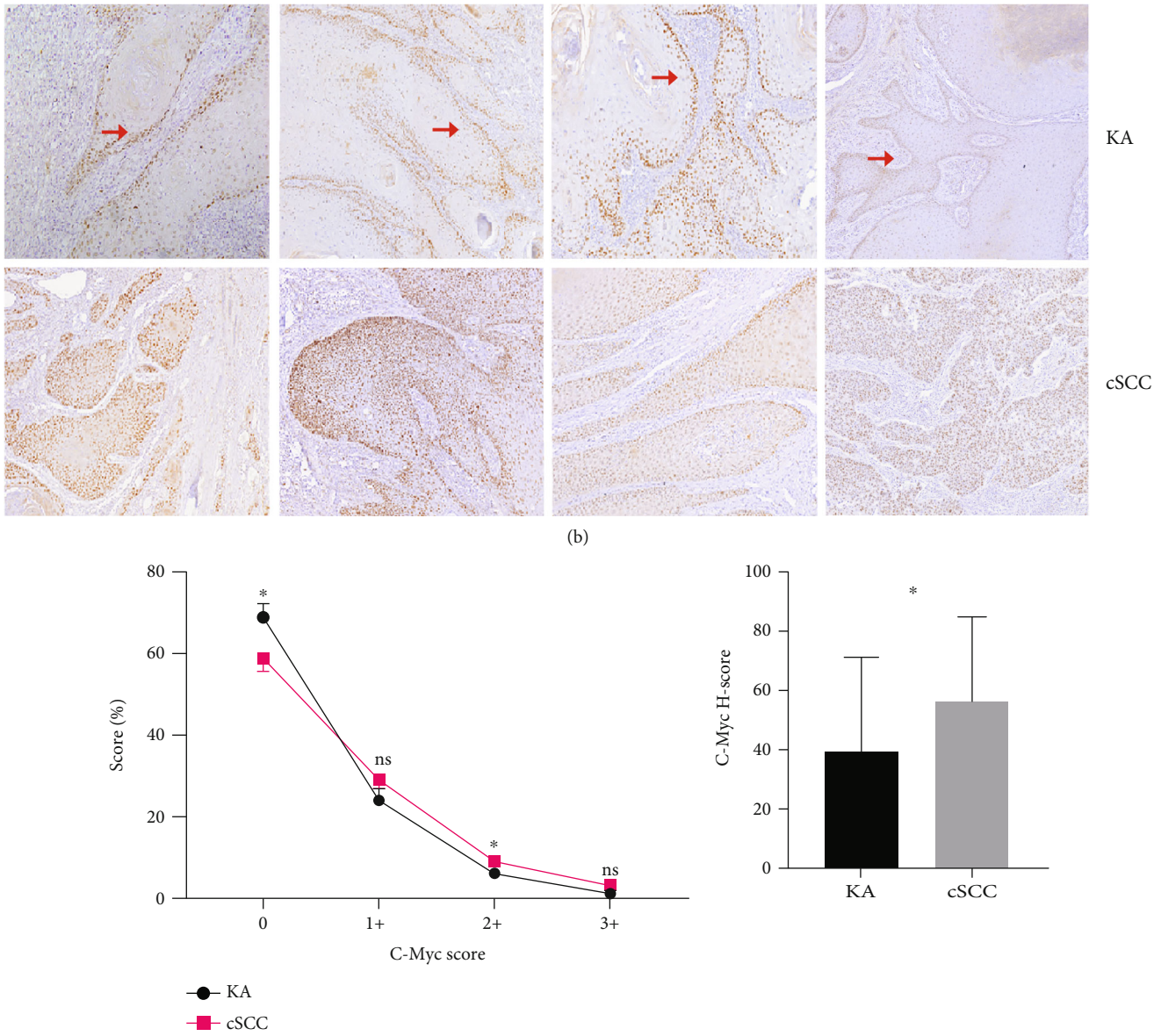
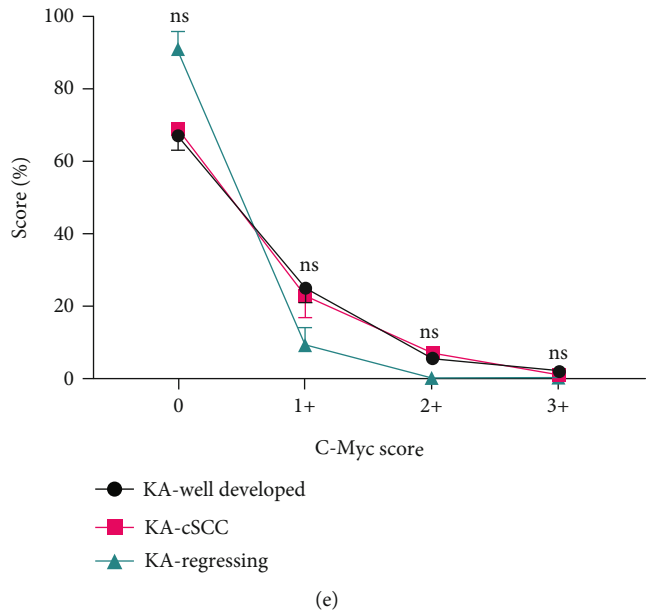


Figure 5: Continued.



(e)

Figure 5: Expression and analysis of C-Myc in Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC). (a-1) Immunohistochemical staining. C-Myc in KA samples was located in the initial portion of tumor infiltration, whereas in cSCC samples, there was diffuse nuclear staining (20x). (a-2) Tissue area delineation (red: tumor target area; green: mesenchyme; and blue: blank). (a-3) Cytoarchitectonic delineation (green). (a-4) C-Myc DAB Score delineation (Score 0: blue; Score 1+: yellow; Score 2+: orange; and Score 3+: dark red). (b) Expression of C-Myc in KA and cSCC patients with arrows indicating the basal region of KA tumors. (c) Percent Scores 0, 1+, 2+, and 3+ in grades of KA versus cSCC samples based on the intensity of C-Myc DAB positivity, with C-Myc Score 0 ($P = 0.034$) and C-Myc Score 2+ ($P = 0.018$). (d) C-Myc H-Scores in KA versus cSCC based on different intensities of C-Myc DAB positivity ($P = 0.024$). (e) Cross-comparisons of the C-Myc Score among the three subtypes of KA (KA-well developed, KA-cSCC, KA-regressing) were not statistically significant.

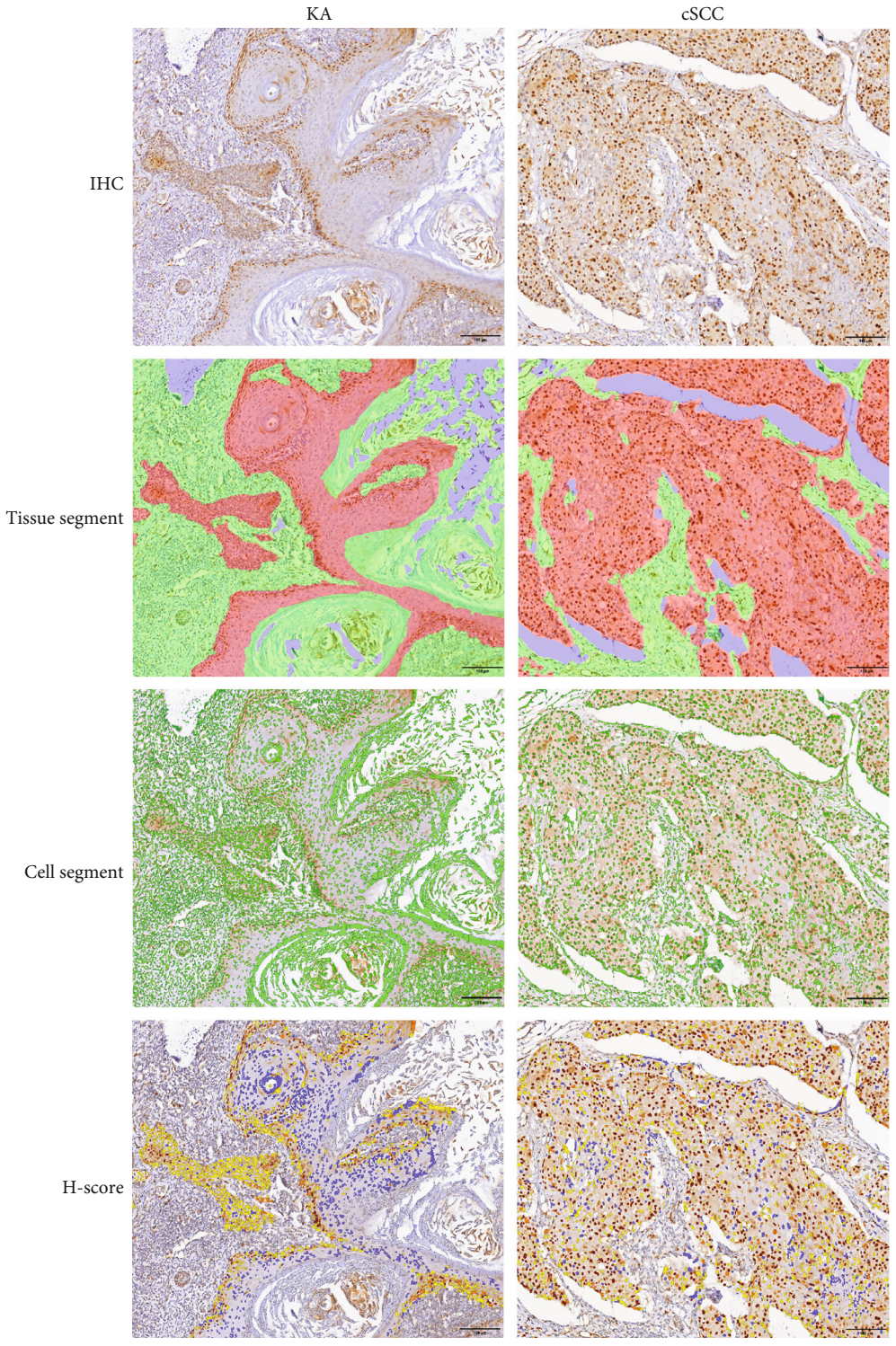
Table 2: C-Myc staining patterns in Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC) patients.

	KA (n (%))	cSCC (n (%))	P value
Diffuse staining pattern	5 (11.90)	21 (67.74)	0.000
Peripheral staining pattern	37 (88.10)	10 (32.26)	

Note. (1) Diffuse staining pattern: C-Myc expression was mainly located in the nuclei and/or in the cytosol of most cells. (2) Peripheral staining pattern: C-Myc expression was mainly located in the nuclei of basal lamina cells at the initial portion of tumor infiltration and/or showed positive staining within the nuclei of basal lamina tumor cells and some spiny lamina cells.

overexpression are observed in malignant cSCC. Moreover, there is evidence that C-Myc prevents keratinocytes from uncontrollable malignant proliferation and stimulates keratinocyte terminal differentiation in normal skin [26–29]. In this way, C-Myc may function as a suppressor of malignant C-Myc effects in KA by inhibiting the migration of keratinocytes to the stratum corneum, thus creating many "central corpuscles" in the tumor. Such an effect would provide an explanation for the cancerous effect of KA. C-Myc expression was concentrated in the nucleus of the basal layer at the initial portion of tumor infiltration, while in KA, this expression is observed in other areas, suggesting that the expression of C-Myc is negative. Hence, we speculate that the "multiple effects" of C-Myc can result in the spontaneous regression of KA and facilitate the transformation of KA to cSCC.

Pelengaris et al. reported that overexpression of C-Myc disrupted the differentiation of keratinocytes in the epidermis and that cSCC showed enhanced levels of proliferation and disruption of the differentiation process, results which are consistent with the concept of a "tumor MYC addiction." MYC activation represents one of the events required for the initiation of tumorigenesis, which implies that tumor survival is dependent on high levels of MYC [30] and the hypothesis that tumor survival is dependent on increased MYC levels. Mechanistically, it appears that in tumor cells, C-Myc boosts the output of established gene expression programs, which can lead to transcriptional amplification, resulting in increased transcription levels in gene expression programs [31]. These results provide an explanation as to the means through which oncogenic C-Myc can exert differential effects on gene expression in different tumor cells and suggest that transcriptional amplification and reduced restriction of tumor cell growth and proliferation rates, along with a progressive increase in C-Myc expression, occur in highly aggressive, but poorly differentiated, tumors. Here, we demonstrate that C-Myc expression was significantly greater in cSCC versus KA.



(a)

Figure 6: Continued.

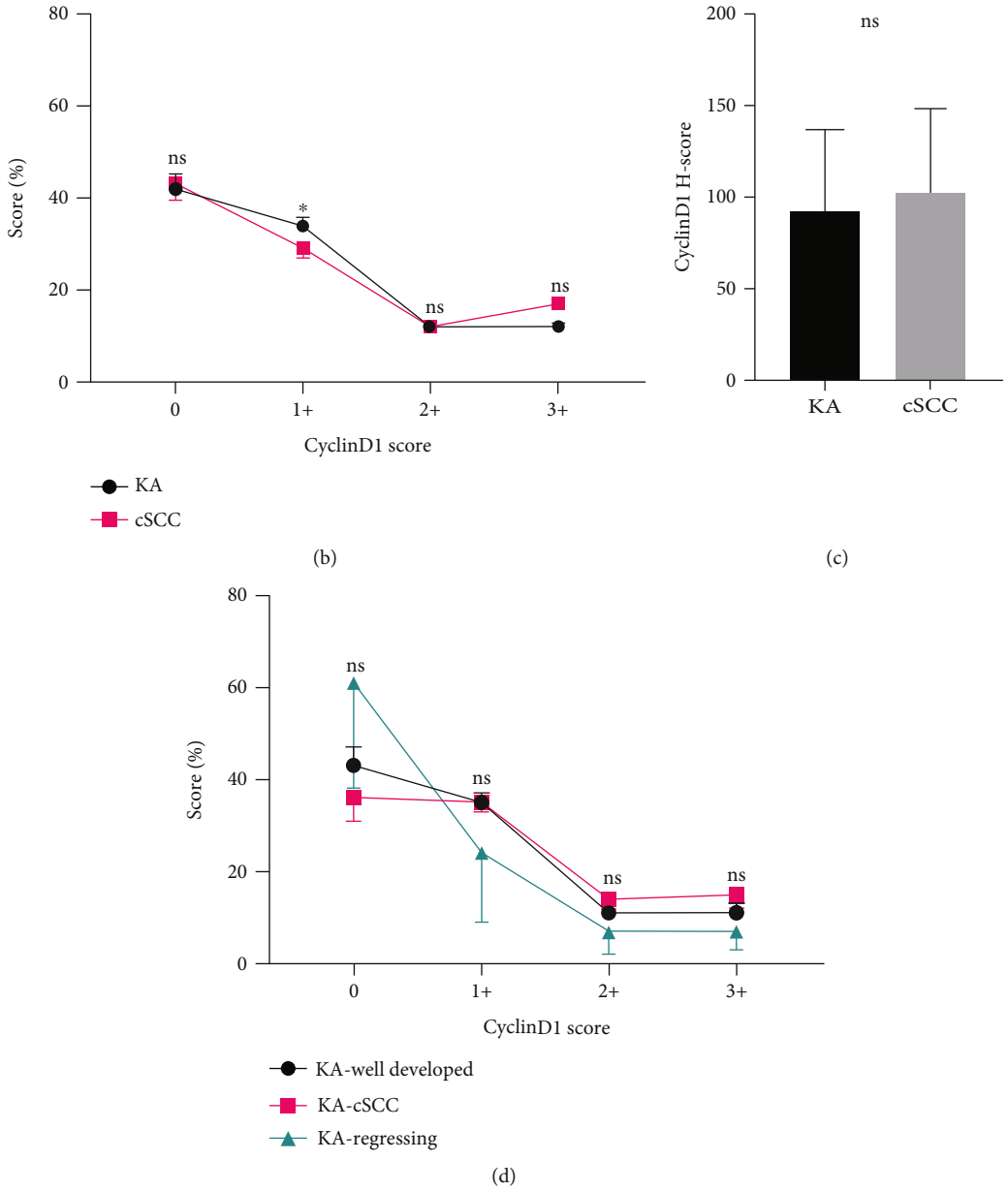
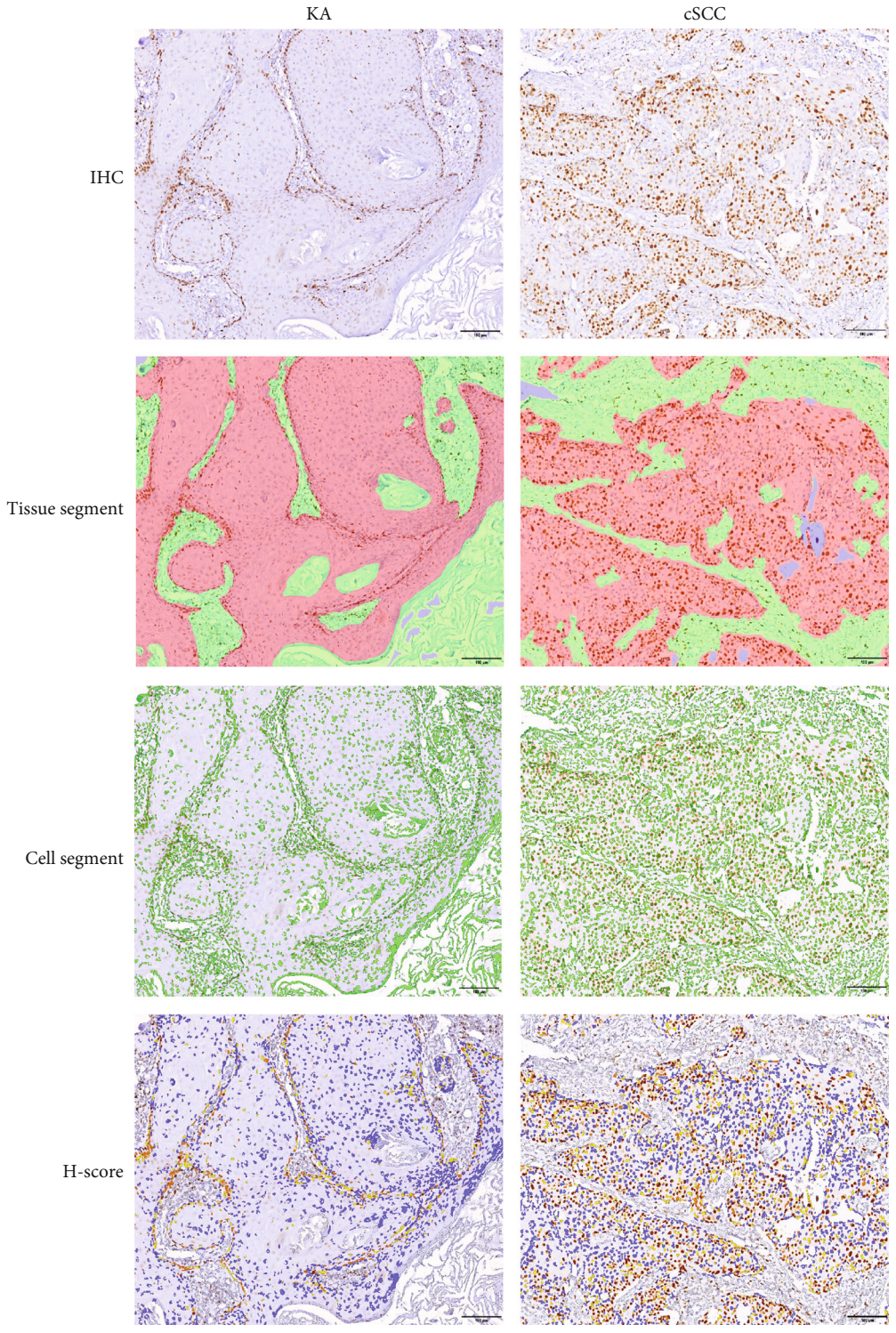


Figure 6: Expression and analysis of CyclinD1 in Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC). (a-1) Immunohistochemical staining. CyclinD1 in KA samples was located in the initial portion of tumor infiltration, whereas in cSCC samples, there was diffuse nuclear staining (20x). (a-2) Tissue area delineation (red: tumor target area; green: mesenchyme; and blue: blank). (a-3) Cytoarchitectonic delineation (green). (a-4) CyclinD1 DAB Score delineation (Score 0: blue; Score 1+: yellow; Score 2+: orange; and Score 3+: deep red). (b) Percent of Scores 0, 1+, 2+, and 3+ in classes of KA and cSCC samples based on CyclinD1 DAB positive intensity, with CyclinD1 Score 1+ ($P = 0.041$). (c) H-Scores in KA versus cSCC samples based on different intensities of CyclinD1 DAB positivity ($P = 0.360$). (d) Cross-comparisons of the CyclinD1 Score among the three subtypes of KA (KA-well developed, KA-cSCC, KA-regressing) were not statistically significant.

Interestingly, analyses of C-Myc and Ki-67 H-Scores revealed that they were positively correlated and consistent in KA but not in cSCC. In contrast, CyclinD1, a target gene activated at comparable levels as those of C-Myc in the Wnt/β-catenin pathway, does not demonstrate a positive correlation with Ki-67 ($R = 0.242$, $P = 0.095$). Such results not only suggest a potentially significant effect of C-Myc but also suggest that Ki-67 and C-Myc may play similar roles in cSCC and KA. Based on previous diagnostic recommendations, we classified a Ki-67 positive percent of >20% as cSCC and that of ≤20% as KA [7, 8], along with our summary of the C-Myc expression pattern in these patients. Moreover, there is an increase in the diagnostic sensitivity with the application of two parallel diagnoses versus that obtained with a single diagnosis. In this way, the varied expression patterns of C-Myc and Ki-67 can be used to cross-reference KA and cSCC during the clinical diagnosis process, although further research with increased sample sizes will be required



(a)

Figure 7: Continued.

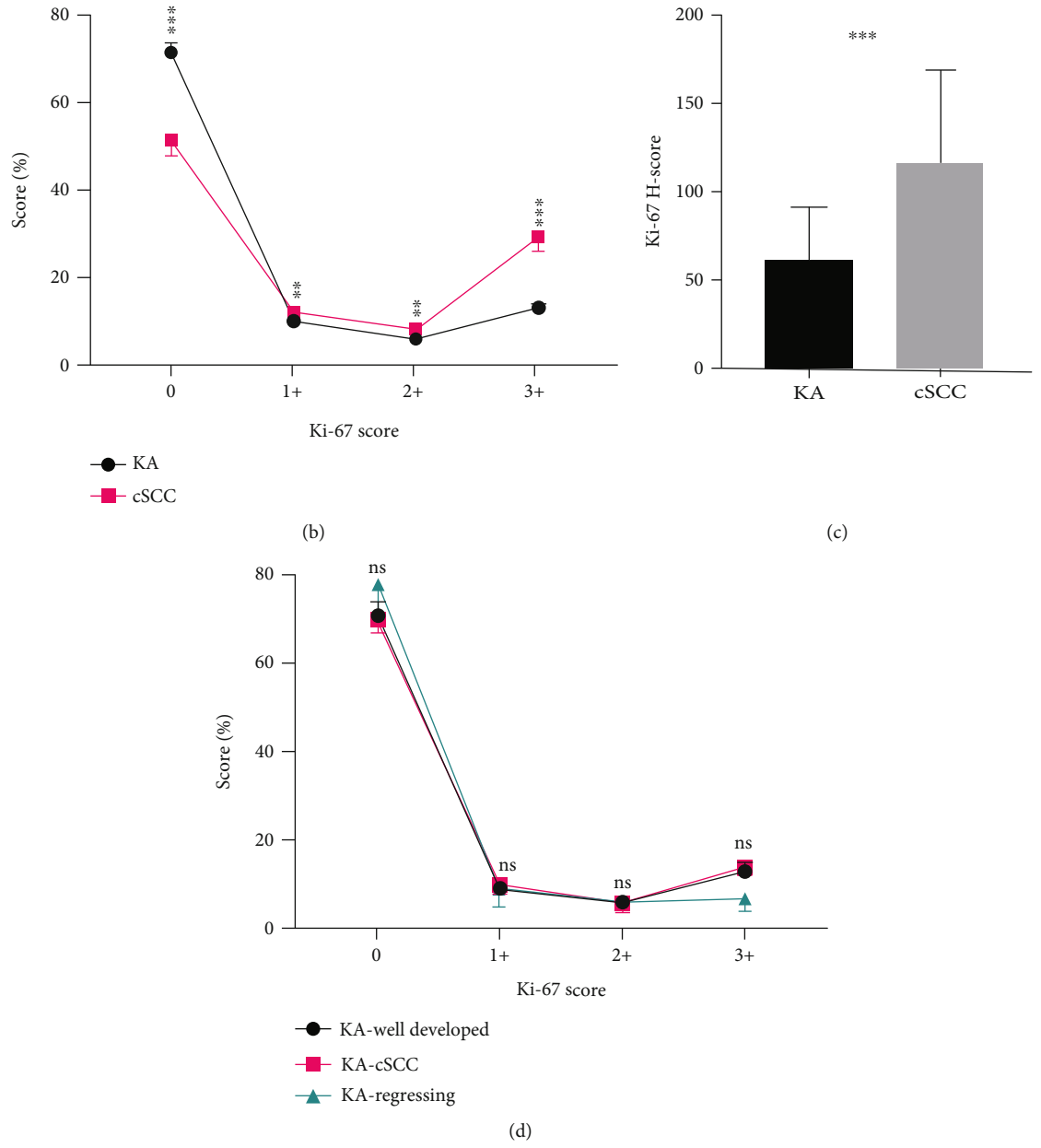


Figure 7: Expression and analysis of Ki-67 in Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC). (a-1) Immunohistochemical staining. Ki-67 in KA samples was located in the initial portion of tumor infiltration, whereas in cSCC samples, there was diffuse nuclear staining (20x). (a-2) Tissue area delineation (red: tumor target area; green: mesenchyme; and blue: blank). (a-3) Cytoarchitectonic delineation (green). (a-4) Ki-67 DAB Score delineation (Score 0: blue; Score 1+: yellow; Score 2+: orange; and Score 3+: dark red). (b) Percent of Scores 0, 1+, 2+, and 3+ in grades of KA and cSCC samples based on differences in Ki-67 DAB positive intensities, with Ki-67 Score 0 ($P = 0.000$), Ki-67 Score 1+ ($P = 0.011$), Ki-67 Score 2+ ($P = 0.005$), and Ki-67 Score 3+ ($P = 0.000$). (c) *H* -Scores in KA and cSCC samples based on differences in intensities of Ki-67 DAB positivity ($P = 0.000$). (d) Cross-comparisons of the Ki-67 Score among the three subtypes of KA (KA-well developed, KA-cSCC, KA-regressing) were not statistically significant.

to substantiate these speculations. As it currently stands, Ki-67 remains one of the primary biomarkers for distinguishing KA from cSCC.

In addition to C-Myc, CyclinD1 represents another major downstream target gene of the Wnt/β-catenin signaling pathway. CyclinD1 regulates the cell cycle primarily by boosting the G1/S phase transition, which then permits the initiation of cell division. Under normal conditions, CyclinD1 protein is only transiently present in the cell [32], however, overexpression of CyclinD1 induces excessive skin cell proliferation and is considered necessary for the development of skin carcinogenesis, as demonstrated in a transgenic mouse model of skin carcinogenesis. CyclinD1, as one of the critical oncogenes regulating the cell cycle, disrupts cell cycle regulation in KA and cSCC, and a gradual increase in the expression of CyclinD1 can contribute to the development of a more aggressive form of the tumor [33–35]. While we found that the results of CyclinD1 expression in KA were consistent

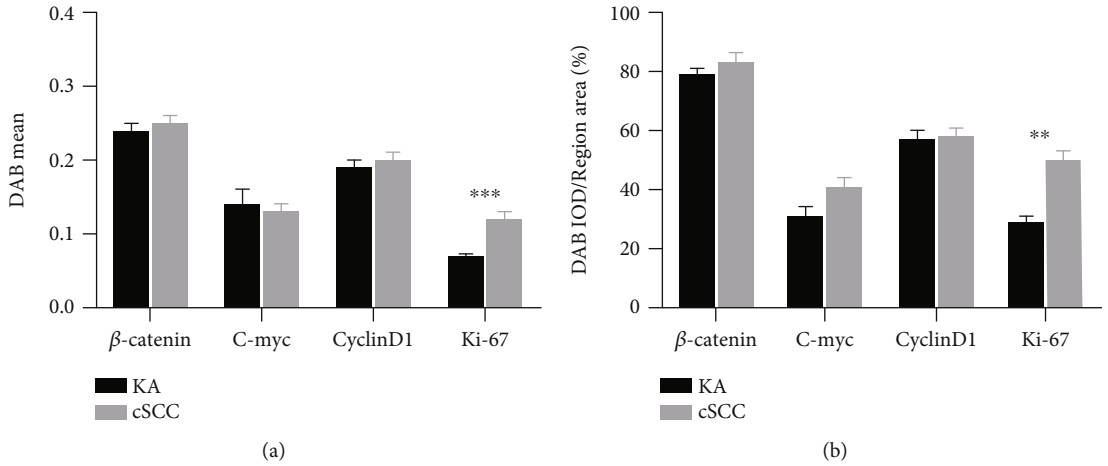


Figure 8: Mean and percent of DAB positivity in Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC). (a) Mean of Ki-67 DAB positivity in KA and cSCC ($P = 0.000$). (b) Percent of Ki-67 positivity in KA and cSCC (DAB$^+$% = Ki-67 DAB IOD/region area) ($P = 0.010$).

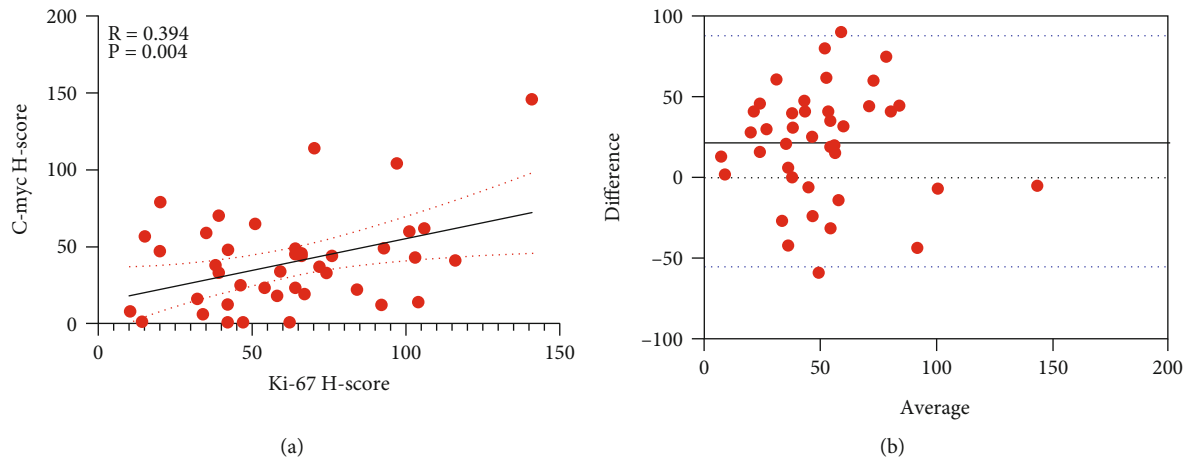


Figure 9: Correlation and consistency analysis of C-Myc and Ki-67 H-Scores in Keratoacanthoma (KA). (a) C-Myc and Ki-67 H-Scores showed a positive correlation ($R = 0.394$, $P = 0.004$). (b) Consistency analysis of C-Myc and Ki-67 H-Scores in KA (Bland-Altman plot).

Table 3: Comparisons of the diagnostic value of Ki-67 and C-Myc in KA patients.

	Sensitivity (%)	Specificity (%)
Ki-67 ≤ 20%	23.80	96.70
C-Myc expression pattern 1	78.60	66.70
Tandem diagnosis	14.30	96.70
Parallel diagnosis	88.00	66.70

Note. Diagnosis 1: a positive Ki-67 fraction of ≤20% is suggestive of KA. Diagnosis 2: a positive C-Myc expression pattern in the nuclei of cells located in the initial portion of tumor infiltration is suggestive of KA. Tandem diagnosis: simultaneous combination of Diagnosis 1 and Diagnosis 2 was required for the diagnosis of KA (Diagnosis 1&Diagnosis 2 = KA). Parallel diagnosis: either Diagnosis 1 or Diagnosis 2 was required for the diagnosis of KA (Diagnosis 1 or Diagnosis 2 = KA).

with previously published findings, the pattern of CyclinD1 expression was not significantly different between the KA and cSCC samples in the current experiment [34]. Additionally, we found that CyclinD1 expression in KA was more likely to be colocalized in the nucleus and cytoplasm of basal and multilayered spiny layer cells. Despite these findings, we believe that this does not represent a discriminatory expression pattern, suggesting that CyclinD1 expression does not play a discriminatory role in KA and cSCC.

β-Catenin was expressed in cell membranes and/or cytoplasm within both the KA and cSCC samples, with this expression in KA samples exhibiting a preference for cell membranes and some portions of the cell cytoplasm, while in cSCC samples, this expression was also found in the cell membrane but in relatively greater concentrations within the cytoplasm than that in KA, findings which were consistent with previous experimental results [36, 37]. The intensity of β-catenin expression was slightly greater in cSCC versus KA, although these differences failed to achieve

statistical significance. When KA subgroups with β-catenin Score 3+ were compared, levels of β-catenin expression were significantly greater in KA-regressing as compared with the other two subtypes. We should note that these findings may be somewhat questionable due to the limited sample size of KA-regressing specimens and a dearth of research available which could explain the morphological anomalies and underlying mechanisms.

At rest, the cell has only trace levels of free β-catenin, while activation of the Wnt/β-catenin signaling pathway increases the quantity of β-catenin in the "cytoplasmic pool" and its translocation to the nucleus, where it initiates and regulates transcription of downstream target genes [12]. As there was an activation of the Wnt/β-catenin signaling pathway in both KA and cSCC in this study, the significantly increased accumulation of free-form β-catenin within the cytoplasm of cSCC versus KA samples implies that more β-catenin would be available for entry into the nucleus. As compared with KA, cSCC demonstrates an increased ability to initiate and regulate transcription of downstream target genes and promote cell proliferation and differentiation. Moreover, cSCC has a greater capacity to control and activate the transcription of downstream target genes and enhance cell proliferation and differentiation.

5. Conclusions

In this report, we investigated the expressions of β-catenin, C-Myc, and CyclinD1 in tissue samples from KA and cSCC patients. These markers represent the major molecules within the Wnt/β-catenin signaling pathway that are believed to be involved in the development of KA and cSCC. In addition, Ki-67, a more established biomarker, was also evaluated in these samples. To evaluate these biomarkers, we used a computerized image platform, which provides a more objective, consistent, and reproducible method for analyzing these data than that achieved with a manual review. Based on results obtained with this method, we provide the first evidence indicating that C-Myc expression in KA was predominantly located in the nuclei of basal cells, whereas this expression in cSCC samples was predominantly in the nuclei of diffuse cells. In addition, we observed that distinct expression patterns of C-Myc and Ki-67, but not CyclinD1, may serve as a means to discriminate KA from cSCC during the clinical diagnosis process. Our findings also support the notion that KA's distinctive spontaneous regression is closely tied to the Wnt/β-catenin pathway in mediating hair follicle development. Based on these results, we propose that KA should be considered a separate mode of skin tumor development and carcinogenesis from that of cSCC.

Data Availability

The data and clinical information used to support the findings of our study are available from the corresponding authors upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Authors' Contributions

Bo Huang, Rui-Qun Qi, and Tao Zhang contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81673070, 82173401, and U1908206).

Supplementary Materials

Supplementary Table 1: descriptive statistics of clinical data for the three subtypes of Keratoacanthoma (KA). Supplementary Table 2: statistical description of Keratoacanthoma (KA) and Cutaneous Squamous Cell Carcinoma (cSCC). Supplementary Table 3: statistical description for the three subtypes of Keratoacanthoma (KA). *(Supplementary Materials)*

References

[1] R. A. Schwartz, "Keratoacanthoma," *Journal of the American Academy of Dermatology*, vol. 30, no. 1, pp. 1–19, 1994.

[2] A. Tisack, A. Fotouhi, C. Fidai, B. J. Friedman, D. Ozog, and J. Veenstra, "A clinical and biological review of keratoacanthoma," *The British Journal of Dermatology*, vol. 185, no. 3, pp. 487–498, 2021.

[3] A. L. Ji, A. J. Rubin, K. Thrane et al., "Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma," *Cell*, vol. 182, no. 2, pp. 497–514.e22, 2020.

[4] M. Vesely, S. Martinez-Morales, J. R. Gehlhausen et al., "Not all well-differentiated cutaneous squamous cell carcinomas are equal: tumors with disparate biologic behavior have differences in protein expression via digital spatial profiling," *Journal of the American Academy of Dermatology*, 2022.

[5] S. K. T. Que, F. O. Zwald, and C. D. Schmults, "Cutaneous squamous cell carcinoma: incidence, risk factors, diagnosis, and staging," *Journal of the American Academy of Dermatology*, vol. 78, no. 2, pp. 237–247, 2018.

[6] A. Kolmodin, N. P. Pandeya, C. M. Olsen, J. C. Dusingize, D. C. Whiteman, and M. Claeson, "Patient and tumour characteristics of keratoacanthoma in a large, community-based cohort study from Queensland, Australia," *Acta Dermato-Venereologica*, vol. 101, no. 6, 2021.

[7] N. Scola, H. M. Segert, M. Stücker, P. Altmeyer, T. Gambichler, and A. Kreuter, "Ki-67 may be useful in differentiating between keratoacanthoma and cutaneous squamous cell carcinoma," *Clinical and Experimental Dermatology*, vol. 39, no. 2, pp. 216–218, 2014.

[8] T. Shimizu, M. Muto, T. Murakami, H. Furumoto, S. Mogami, and C. Asagami, "Overexpression of p53 protein associated with Proliterative activity as evaluated by Ki-67 immunostaining in well-differentiated squamous cell carcinoma of the skin," *Dermatology*, vol. 195, no. 3, pp. 224–227, 2004.

[9] A. Toll, R. Salgado, M. Yébenes et al., "MYC gene numerical aberrations in actinic keratosis and cutaneous squamous cell

carcinoma," *The British Journal of Dermatology*, vol. 161, no. 5, pp. 1112–1118, 2009.

[10] M. S. Jacobs, D. L. Persons, and G. R. Fraga, "EGFR and MYC gene copy number aberrations are more common in squamous cell carcinoma than keratoacanthoma: a FISH study," *Journal of Cutaneous Pathology*, vol. 40, no. 5, pp. 447–454, 2013.

[11] N. Meyer and L. Z. Penn, "Reflecting on 25 years with MYC," *Nature Reviews. Cancer*, vol. 8, no. 12, pp. 976–990, 2008.

[12] R. Nusse and H. Clevers, "Wnt/β-catenin signaling, disease, and emerging therapeutic modalities," *Cell*, vol. 169, no. 6, pp. 985–999, 2017.

[13] K. Sobel, M. Tham, H. J. Stark et al., "Wnt-3a-activated human fibroblasts promote human keratinocyte proliferation and matrix destruction," *International Journal of Cancer*, vol. 136, no. 12, pp. 2786–2798, 2015.

[14] A. M. Hassanein and S. M. Glanz, "Beta-catenin expression in benign and malignant pilomatrix neoplasms," *The British Journal of Dermatology*, vol. 150, no. 3, pp. 511–516, 2004.

[15] J. Fu and W. Hsu, "Epidermal Wnt controls hair follicle induction by orchestrating dynamic signaling crosstalk between the epidermis and dermis," *The Journal of Investigative Dermatology*, vol. 133, no. 4, pp. 890–898, 2013.

[16] J. J. Bull, S. Pelengaris, S. Hendrix, C. M. Chronnell, M. Khan, and M. P. Philpott, "Ectopic expression of c-Myc in the skin affects the hair growth cycle and causes an enlargement of the sebaceous gland," *The British Journal of Dermatology*, vol. 152, no. 6, pp. 1125–1133, 2005.

[17] C. G. Ramselaar, E. J. Ruitenberg, and W. Kruizinga, "Regression of induced keratoacanthomas in anagen (hair growth phase) skin grafts in mice," *Cancer Research*, vol. 40, no. 5, pp. 1668–1673, 1980.

[18] K. Yoshikawa, Y. Katagata, and S. Kondo, "Relative amounts of keratin 17 are higher than those of keratin 16 in hair- follicle-derived tumors in comparison with nonfollicular epithelial skin tumors," *The Journal of Investigative Dermatology*, vol. 104, no. 3, pp. 396–400, 1995.

[19] G. Zito, I. Saotome, Z. Liu et al., "Spontaneous tumour regression in keratoacanthomas is driven by Wnt/retinoic acid signalling cross-talk," *Nature Communications*, vol. 5, no. 1, p. 3543, 2014.

[20] T. Scholzen and J. Gerdes, "The Ki-67 protein: from the known and the unknown," *Journal of Cellular Physiology*, vol. 182, no. 3, pp. 311–322, 2000.

[21] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: challenges and opportunities," *Medical Image Analysis*, vol. 33, pp. 170–175, 2016.

[22] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[23] J. Zhang, C. Li, S. Kosov et al., "LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, article 107885, 2021.

[24] Y. Li, C. Li, X. Li et al., "A comprehensive review of Markov random field and conditional random field approaches in pathology image analysis," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 609–639, 2022.

[25] S. Pelengaris, T. Littlewood, M. Khan, G. Elia, and G. Evan, "Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion," *Molecular Cell*, vol. 3, no. 5, pp. 565–577, 1999.

[26] G. I. Evan, A. H. Wyllie, C. S. Gilbert et al., "Induction of apoptosis in fibroblasts by c-myc protein," *Cell*, vol. 69, no. 1, pp. 119–128, 1992.

[27] E. B. Thompson, "The many roles of c-Myc in apoptosis," *Annual Review of Physiology*, vol. 60, no. 1, pp. 575–600, 1998.

[28] S. Pelengaris, S. Abouna, L. Cheung, V. Ifandi, S. Zervou, and M. Khan, "Brief inactivation of c-Myc is not sufficient for sustained regression of c-Myc-induced tumours of pancreatic islets and skin epidermis," *BMC Biology*, vol. 2, no. 1, p. 26, 2004.

[29] Y. Li, S. C. Casey, and D. W. Felsher, "Inactivation of MYC reverses tumorigenesis," *Journal of Internal Medicine*, vol. 276, no. 1, pp. 52–60, 2014.

[30] M. Gabay, Y. Li, and D. W. Felsher, "MYC activation is a hallmark of cancer initiation and maintenance," *Cold Spring Harbor Perspectives in Medicine*, vol. 4, no. 6, 2014.

[31] C. Y. Lin, J. Lovén, P. B. Rahl et al., "Transcriptional amplification in tumor cells with elevated c-Myc," *Cell*, vol. 151, no. 1, pp. 56–67, 2012.

[32] E. Tashiro, A. Tsuchiya, and M. Imoto, "Functions of cyclin D1 as an oncogene and regulation of cyclin D1 expression," *Cancer Science*, vol. 98, no. 5, pp. 629–635, 2007.

[33] S. I. Mitsunaga, S. Y. Zhang, B. A. Ruggeri et al., "Positive immunohistochemical staining of p53 and cyclin D in advanced mouse skin tumors, but not in precancerous lesions produced by benzo[a]pyrene," *Carcinogenesis*, vol. 16, no. 7, pp. 1629–1635, 1995.

[34] B. Burnworth, S. Popp, H. J. Stark et al., "Gain of 11q/cyclin D1 overexpression is an essential early step in skin cancer development and causes abnormal tissue organization and differentiation," *Oncogene*, vol. 25, no. 32, pp. 4399–4412, 2006.

[35] A. A. Mohamed, A. A. A. Omar, R. R. El-Awady et al., "miR-155 and miR-665 role as potential non-invasive biomarkers for hepatocellular carcinoma in Egyptian patients with chronic hepatitis C virus infection," *Journal of Translational Internal Medicine*, vol. 8, no. 1, pp. 32–40, 2020.

[36] K. Fukumaru, N. Yoshii, T. Kanzaki, and T. Kanekura, "Immunohistochemical comparison of β-catenin expression by human normal epidermis and epidermal tumors," *The Journal of Dermatology*, vol. 34, no. 11, pp. 746–753, 2007.

[37] E. Papadavid, M. Pignatelli, S. Zakynthinos, T. Krausz, and A. C. Chu, "The potential role of abnormal E-cadherin and alpha-, beta- and gamma-catenin immunoreactivity in the determination of the biological behaviour of keratoacanthoma," *The British Journal of Dermatology*, vol. 145, no. 4, pp. 582–589, 2001.

Hindawi
BioMed Research International
Volume 2022, Article ID 7966553, 10 pages
https://doi.org/10.1155/2022/7966553

Research Article

Histopathological Tissue Segmentation of Lung Cancer with Bilinear CNN and Soft Attention

Rui Xu,[1] Zhizhen Wang,[1] Zhenbing Liu,[1] Chu Han,[2,3,4] Lixu Yan,[5] Huan Lin,[2,4] Zeyan Xu,[2,4] Zhengyun Feng,[1] Changhong Liang,[2,4] Xin Chen,[6] Xipeng Pan,[1,2,3,4] and Zaiyi Liu[2,4]

[1]*School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, China*
[2]*Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China*
[3]*Guangdong Cardiovascular Institute, Guangzhou, China*
[4]*Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China*
[5]*Department of Pathology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China*
[6]*Department of Radiology, Guangzhou First People's Hospital, The Second Affiliated Hospital of South China University of Technology, Guangzhou, China*

Correspondence should be addressed to Xin Chen; wolfchenxin@163.com, Xipeng Pan; pxp201@guet.edu.cn, and Zaiyi Liu; liuzaiyi@gdph.org.cn

Received 21 February 2022; Revised 15 May 2022; Accepted 10 June 2022; Published 7 July 2022

Academic Editor: Cheng Lu

Copyright © 2022 Rui Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Automatic tissue segmentation in whole-slide images (WSIs) is a critical task in hematoxylin and eosin- (H&E-) stained histopathological images for accurate diagnosis and risk stratification of lung cancer. Patch classification and stitching the classification results can fast conduct tissue segmentation of WSIs. However, due to the tumour heterogeneity, large intraclass variability and small interclass variability make the classification task challenging. In this paper, we propose a novel bilinear convolutional neural network- (Bilinear-CNN-) based model with a bilinear convolutional module and a soft attention module to tackle this problem. This method investigates the intraclass semantic correspondence and focuses on the more distinguishable features that make feature output variations relatively large between interclass. The performance of the Bilinear-CNN-based model is compared with other state-of-the-art methods on the histopathological classification dataset, which consists of 107.7 k patches of lung cancer. We further evaluate our proposed algorithm on an additional dataset from colorectal cancer. Extensive experiments show that the performance of our proposed method is superior to that of previous state-of-the-art ones and the interpretability of our proposed method is demonstrated by Grad-CAM.

1. Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide [1, 2]. Precise diagnosis is crucial for treatment planning. Histological assessment of hematoxylin and eosin- (H&E-) stained tissue specimens remains the gold standard for lung cancer diagnosis [3, 4]. In clinical practice, pathologists use their domain knowledge and experience to assess the complex morphological and cytological features of tissue samples under a light microscope to diagnose [5, 6]. However, the process is time-consuming, subjective, with considerable inter- and intraobserver variability [3, 7]. Recently, digital pathology, converting conventional glass slides into digital resources known as whole-slide images (WSIs), rises the development of automatic diagnosis [8–10]. One of the most needs for automatic disease diagnosis is to distinguish different tissue components (tumour epithelium, stroma, necrosis, tumour-infiltrating lymphocytes, etc.) in H&E-

stained WSIs [11–13]. Therefore, an initial step in automatic diagnosis is to develop a robust automatic tissue segmentation algorithm.

Deep learning models have demonstrated a strong segmentation ability in histopathological images [14–16]. Various DL models have been proposed for patch-level tissue segmentations in WSIs. Xu et al. [15] proposed a DCNN model to extract the convolutional features for classifying epithelial and stromal in histopathological images. Zhao et al. [17] presented a VGG19-based model for automatic tissue segmentation and automated TSR quantification in WSIs of colorectal cancer. Kather et al. [18] compared recently deep learning models in histopathological images in colon cancer and concluded that the VGG19 model worked best at tissue classification. Chan et al. [8] applied a CAM-based method with a fully connected conditional random field for patch-level tissue segmentation. Xu et al. [14] proposed a DenseNet-based approach with focal loss to deal with class imbalance in histopathological images. Anklin et al. [19] proposed a weakly supervised method based on tissue graphs to utilize inexact and incomplete annotations to segment whole-slide images. Yang et al. [20] found that the multimodel method was relatively better than single model-based ones for the automatic diagnosis of lung cancer. Li et al. [21] proposed an EfficientNet-based model to identify tissue in histopathological images. However, they did not take into account the high heterogeneity of tissue types (as shown in Figure 1). Even homogeneous tissue types differ in color, shape, and texture, which provides a further challenge for automatic segmentation.

Bilinear convolutional neural network (Bilinear-CNN) is an effective architecture for fine-grained visual recognition tasks [22–24]. The original bilinear pooling can be generalized to all convolutional neural networks [25]. The bilinear pooling provides an advantage for Bilinear-CNN in that computational layers in networks can have a strong capacity with pairwise interactions [25, 26].

In this paper, we propose a novel Bilinear-CNN-based model to handle the issue of large intraclass variability and small interclass variability in histopathological images. The Bilinear-CNN-based model combines a bilinear convolutional module and soft attention module to perform multitissue classification of histopathological images in lung cancer. It investigates the correct semantic correspondence of intraclass and focuses on the more distinguishable features that make feature output variations relatively large between interclass.

2. Materials and Methods

2.1. Datasets. In this work, lung cancer and colorectal cancer multitissue histopathological image datasets are used for experiments.

The lung cancer multitissue histopathological image dataset is introduced in this work. It contains 107.7 k patches from 67 slides of lung cancer, which were scanned by an Aperio-AT2 scanner in the Department of Pathology at Guangdong Provincial People's Hospital, China. Each slide corresponds to an independent patient. The training set includes 78 k image patches from 57 slides. The independent test set includes 29.7 k image patches from 10 slides. The image patches are extracted partially overlapping tiles from H&E-WSI images by a sliding window with the resolution of 224 × 224 (20x magnification). The step size of the sliding window is 56 pixels. Within this dataset, tumour epithelium (TUM), stroma (STR), tumour-infiltrating lymphocytes (LYM), necrosis (NEC), bronchus (BRO), vessel (VES), normal (NOR), background (BAC), areas polluted by carbon dust (APC), and others (OTH) can be observed. The dataset is validated by two experienced pathologists and judged by a senior pathologist if there are differences in classification.

The colorectal cancer multitissue histopathological image dataset was published by Zhao et al. [17]. Tissue types were grouped into nine classes, including TUM, STR, LYM, BAC, NOR, debris (DEB), mucus (MUS), smooth muscle (MUC), and adipose (ADP). The training set included 283.1 k image patches from 191 slides. The independent test set included 28.8 k image patches from 48 slides. Then, image patches with the size 224 × 224 (20x magnification) were extracted partially overlapping tiles from H&E-WSI images. The step size of the sliding window was 84 pixels.

2.2. Methodology. In this part, we describe our proposed two-stage multitissue segmentation algorithm. First, we introduce a novel Bilinear-CNN-based model to discriminate multiclass tissue types. Second, patches are predicted by our model, then stitched back to get the prediction map. The entire algorithm for automatic tissue segmentation is shown in Figure 2, and an overview of the proposed classification network is shown in Figure 3.

2.2.1. Classification Network. To make feature output variations relatively large between interclass and within the intraclass, we propose a simple but effective method that combined Bilinear-CNN and a soft attention module (Figure 3). The portion of the network to extract the features is ResNet50, because of its outstanding performance in recent computer vision tasks. We remove the global average pooling layer compared to the standard pretrained ResNet50 implementation. Instead, the features extracted by the convolution layer are fed into a bilinear pooling module. Then, a soft attention module is added after the bilinear pooling module and used to receive the features that represented the biological significance of the tissue components, which is the output of the bilinear pooling module. Finally, the softmax layer is used for prediction.

(1) Bilinear Pooling Module. The bilinear pooling module [24] is used to investigate the correct semantic correspondence between the intraclass. When given an input feature vector $x \in R_n$ of a sample, the general linear transformation can be expressed as

$$y = b + w^T x, \tag{1}$$

where y is the output of a node, b is the bias, $w \in R_n$ is the corresponding transformation weight matrix, and the dimension of the input features is n.

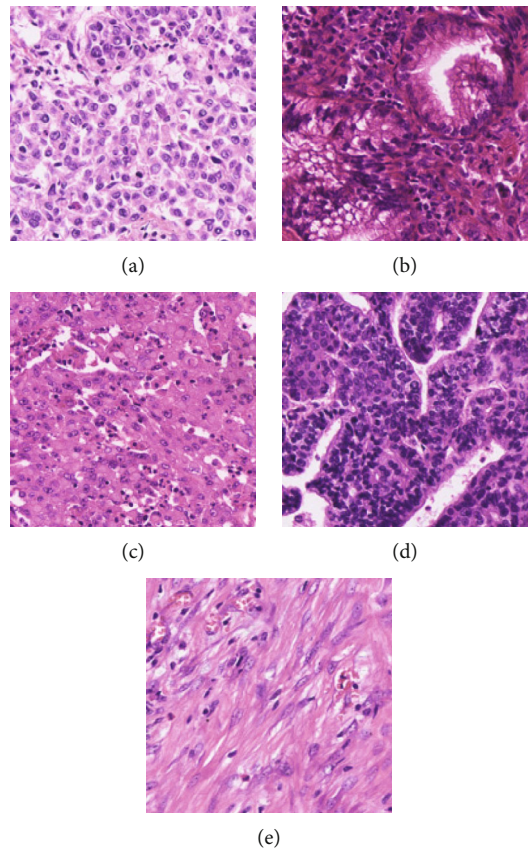


Figure 1: Example images of the (a–d) tumour epithelium and (e) stroma. Tumour epithelium (a–d) has large intraclass variability. (e) Stroma has small interclass variability with (c) tumour epithelium.

To investigate the correct semantic correspondence between the intraclass, we use the bilinear pooling module as follows:

$$y = b + w^T x + x^T F^T F x, \tag{2}$$

where $F \in R_{k \times n}$ is the corresponding reciprocity weight matrix with $k \in N^+$ factors.

To illustrate the bilinear pool module more clearly, the expression of Equation (2) can be expatiated as

$$y = b + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle f_i, f_j \right\rangle x_i x_j. \tag{3}$$

The ith value of the input x is x_i. The ith variable of the first-order weight is w_i, and the ith column of F is f_i. $\langle f_i, f_j \rangle$ is the inner product of f_i and f_j, which explains the interaction between the ith and jth values of the input feature vector.

Compared to other deep learning models, we not only use first-order features but also use second-order features to achieve better classification. The bilinear pooling module can help the convolution layer and full connection layer to break through linear transformation, capture nonlinear features, improve the richness of extracted features, and thus obtain bilinear features of the same subclass, which we use as input into the attention module.

(2) Soft Attention Module. The soft attention module [27] receives the bilinear features of the same subclass and increases the feature output variations between different subclass. The attention module provides an attention weight for features that can be participated in backpropagation. First, matrix multiplication between attention weights with feature vector is performed. Then, we get the scalar by using the softmax function, which can be learned with training iterations. Finally, we take the corresponding scalar and matrix multiply each neuron, and sum to get the distinguishable features as follows:

$$c = \sum_{i=1}^{n} a_i f_i, \tag{4}$$

where c is the distinguishable feature, a_i is the ith variable of the attention weight a, and f_i is the ith value of the input feature y from Equation (3).

The expression of the differentiable a can be explained as

$$a_i = \frac{\exp(e_i)}{\sum_{i=1}^{n} \exp(e_i)}, \tag{5}$$

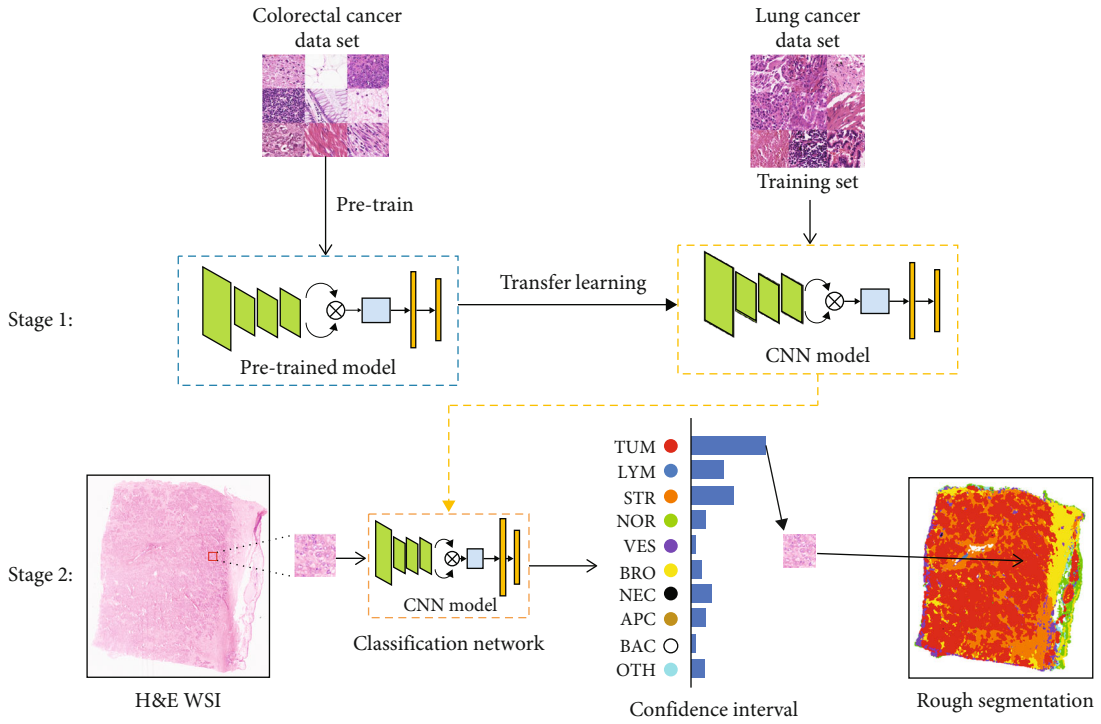


Figure 2: An overview of the proposed multitissue segmentation algorithm. Stage 1: a classification network is pretrained on the colorectal cancer dataset, and transfer learning is used to train the classification network with the training set of the lung cancer dataset. The independent image dataset is used to evaluate the classification accuracy of the network. Stage 2: H&E-WSI image (20x magnification) is segmented through stitching the classification results tile by tile. H&E: hematoxylin and eosin; WSI: whole-slide image; TUM: tumour epithelium; LYM: tumour-infiltrating lymphocytes; STR: stroma; NOR: normal; VES: vessel; BRO: bronchus; NEC: necrosis; APC: areas polluted by carbon dust; BAC: background; OTH: others.

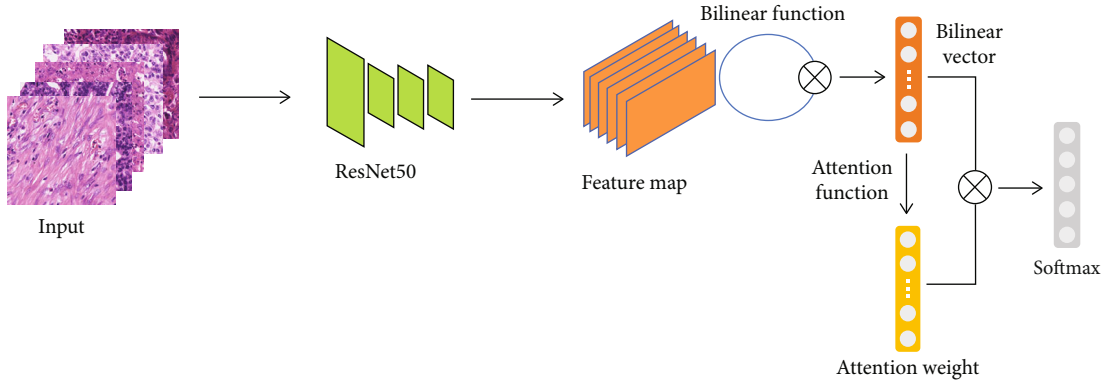


Figure 3: An overview of the proposed classification network. First, the images are fed into ResNet50 to get the feature maps, and feature maps are input to the bilinear function to obtain the bilinear vector; then, the attention weight is got from the attention function, and finally, the bilinear vector multiplies the attention weight to obtain the attention feature flowed into the softmax layer for classification.

where e_i is ith value of the scalar and e is expatiated as

$$e = \sum_{i=1}^{n} w_i f_i, \tag{6}$$

where w_i is ith variable of the attention weight, which can be learned with training iterations.

The soft attention module can improve the ability of the model to learn distinguishable features of different subclass in histopathological images. The scalar is used to make the model focus on the more distinguishable features, and with the learning, important distinguishable features become more prominent in the model. And we have demonstrated the effectiveness of this method through experiments. The details of the experiments are shown in Section 2.3.

2.2.2. Transfer Learning. In the lung multitissue histopathological image classification task, the classification network pretrained with transfer learning. This strategy can make

Table 1: Results on lung cancer dataset (all models pretrained on colorectal cancer dataset).

Model	Average precision	Average recall	Average F1
ResNet50 [29]	0.9239	0.9279	0.9259
VGG19 [30]	0.9185	0.9238	0.9211
EfficientNet [31]	0.9184	0.9295	0.9239
DeepTissue Net [14]	0.9218	0.9250	0.9234
ResNet50+bilinear pooling module	0.9253	0.9286	0.9269
ResNet50+attention module	0.9268	0.9291	0.9279
ResNet50+bilinear pooling module+attention module	0.9394	0.9415	0.9404

Table 2: The classification F1 score of tissue types on lung cancer dataset.

Model	TUM	LYM	STR	NOR	VES	BRO	NEC	APC	BAC	OTH
ResNet50 [29]	0.9686	0.9914	0.8687	0.8532	0.8512	0.9615	0.9668	0.9962	0.9959	0.7734
VGG19 [30]	0.9639	0.9913	0.8720	0.8354	0.8821	0.9545	0.9628	0.9939	0.9944	0.7219
EfficientNet [31]	0.9514	0.9789	0.8854	0.8414	0.8768	0.9533	0.9689	0.9955	0.9957	0.7326
DeepTissue Net [14]	0.9331	0.9640	0.8669	0.8635	0.8973	0.9542	0.9589	0.9967	0.9915	0.7852
ResNet50+bilinear pooling module	0.9698	0.9911	0.8753	0.8658	0.8736	0.9538	0.9638	0.9969	0.9936	0.7857
ResNet50+attention module	0.9712	0.9916	0.8862	0.8732	0.8954	0.9582	0.9615	0.9972	0.9931	0.7516
ResNet50+bilinear pooling module+attention module	0.9739	0.9911	0.9056	0.8788	0.9186	0.9586	0.9766	0.9952	0.9935	0.8025

better use of existing public multitissue histopathological image datasets and achieve better results on the multitissue classification task. We used the Bilinear-CNN-based model trained on colorectal cancer multitissue histopathological image dataset as the pretrained model. And then, we fine-tuned the model in the lung multitissue histopathological images dataset based on the pretrained model.

2.2.3. Visual Interpretability of the Classification Network. To demonstrate that the Bilinear-CNN-based model can identify the tissue types, we utilized the Grad-CAM to generate the visual interpretability of the classification network. Grad-CAM [28] describes a visual explanation of models based on object gradients. A localization map of important image regions is highlighted by Grad-CAM. In the decision-making process, the gradient information flowing into the last convolutional layer of the CNN is utilized by Grad-CAM to assess the importance of features.

2.2.4. Segmentation Map. The trained Bilinear-CNN-based model is used for patch-level multitissue segmentation on histopathological images, and a predictive segmentation map is generated. The detailed operations are as follows. First, we use OpenSlide software to downsample the WSI of lung cancer to get the tissue mask. To distinguish tissue from the background, the tissue region is obtained by threshold segmentation algorithm from the tissue mask. Overlapping tiles are extracted partially by a sliding window with a size of 224×224 from the tissue region. The step size of the sliding window is set at 128 pixels. Then, each image tile is input into the trained multitissue classification model to generate a prediction probability. Finally, the tissue class with the highest prediction probability is selected as the classification result of the image tile.

2.3. Implementation and Training Details. The study was implemented with the open-source software library PyTorch version 1.6.0 on a workstation with Intel(R) Core(TM) i5-10600KF CPU, 32 GB memory, and equipped with NVIDIA GeForce 3090 GPU. During training, the augmentation techniques were applied for the training dataset, including rotations, normalized color appearance, and horizontal flipping. All models in this implementation received input patches of size 224×224. All models were trained with a batch size of 32, weight decay of $1e-4$, and momentum of 0.9 for 80 epochs. Adam optimization with a learning rate of $3e-4$ was used on the colorectal cancer multitissue histopathological image dataset. We used Adam optimization with an initial learning rate of $3e-4$, and then, it reduced to one-tenth if the loss stopped reducing for 30 epochs on the lung cancer multitissue histopathological image dataset.

3. Results

We made independent comparisons to the evaluated model on the lung cancer dataset and an additional dataset from colorectal cancer. Our proposed model was compared to state-of-the-art approaches recently used in computer vision and models specifically designed for the task of tissue classification.

3.1. Comparison on the Lung Cancer Dataset. For comparison of existing models on the lung cancer dataset, all models were pretrained on the colorectal cancer dataset. The results of the comparison are shown in Tables 1 and 2. Figure 4(a) shows the comparison between the ResNet50 model with a bilinear pooling module and attention module and other models concerning the loss on the lung cancer dataset. The loss of the proposed model decreases much faster and

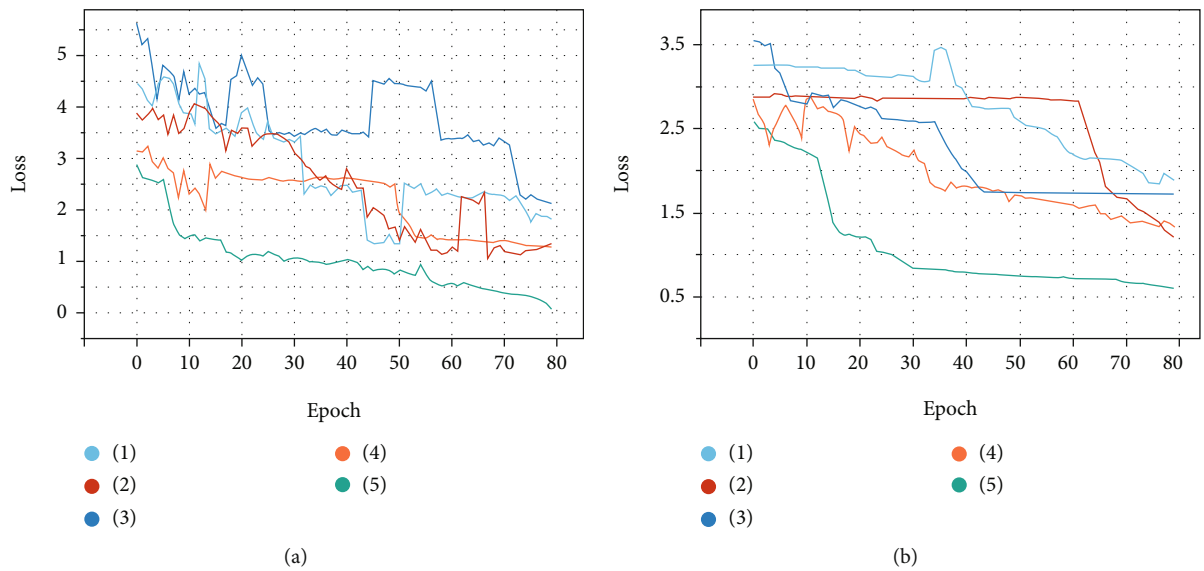


Figure 4: Convergence analysis of models. Implementation and training details of all models are consistent with Section 2.3. (a) Loss on lung cancer dataset. (b) Loss on colorectal cancer dataset. (1) EfficientNet, (2) DeepTissue Net, (3) VGG19, (4) ResNet50, and (5) ResNet50 model with bilinear pooling module and attention module.

smoother than that of other models, demonstrating its superior convergence speed. The ResNet50 model with bilinear pooling module and attention module achieves the best performance.

To analyze which image regions the proposed model focused on, heatmaps of different models are shown in Figure 5. A localization map of important image regions in heatmaps is highlighted by Grad-CAM. Heatmaps of the four most common tissue types are shown, and the slides of lung cancer scanned in 20x magnification factor are selected at random. Figure 5 shows the heatmaps of different models by Grad-CAM which highlights the importance of regions for classification and demonstrates a better focus of the ResNet50 model with bilinear pooling module and attention module on histopathological regions than classic CNNs.

3.2. Comparison on the Colorectal Cancer Dataset. To further evaluate our proposed algorithm, the comparative experiments on an additional dataset from colorectal cancer were implemented. On the colorectal cancer dataset, the pretrained model with ImageNet was used. ImageNet is a large image dataset containing hundreds and thousands of images. In transfer learning tasks, ImageNet is usually used for pretrained models. This public dataset was used to assess the generalization ability and robustness of our multitissue classification model. The results of the comparison are shown in Tables 3 and 4. Figure 4(b) shows the comparison between the ResNet50 model with a bilinear pooling module and attention module and other models concerning the loss on the colorectal cancer dataset. The loss of the proposed model converges faster than in other models. The results illustrate that the proposed model combines the bilinear pooling module and soft attention module to make feature output variations relatively large between interclass and within the intraclass, learn the distinguishable features, and improve the accuracy of the multitissue classification task.

Several conclusions can be drawn: (1) The result of the proposed method is superior to state-of-the-arts recently. (2) the ResNet50 model with bilinear pooling module and attention module achieves the highest classification accuracy in the test, and it is shown that the Bilinear-CNN-based model works well on the multitissue task and effectively alleviates the problem of large intraclass variability and small interclass variability. (3) The model combined Bilinear-CNN, and soft attention module is suitable for the multitissue task.

3.3. Visualizing the Segmentation Results of WSIs. The segmentation result of H&E-stained WSIs in lung cancer is drawn as a map covered on the tiles with various colours representing the output tissue types. In Figure 6, colour standing for each tissue type is randomly selected. The predictions of tissue types are observed and mapped to the in situ tissues. Our method obtains the tissue mask of the downscaled WSI. A threshold segmentation algorithm is used to distinguish tissue from the background and then get the tissue region from the tissue mask. Figure 6 also shows that the predicted regions by our classifier are highly consistent with the distribution of tissue types in histopathological images.

4. Discussion

Automatic tissue segmentation is faced with a challenge in that whole-slide images usually have a large resolution and cannot be directly fed into CNNs. This challenge cannot be alleviated by resizing the image size, which causes the loss of much information. Moreover, in the study of lung cancer histopathological images, the existing works are basically based on the backbone of natural image classification to identify the tissue types. In addition, compared with natural images, histopathological images are high heterogeneity in

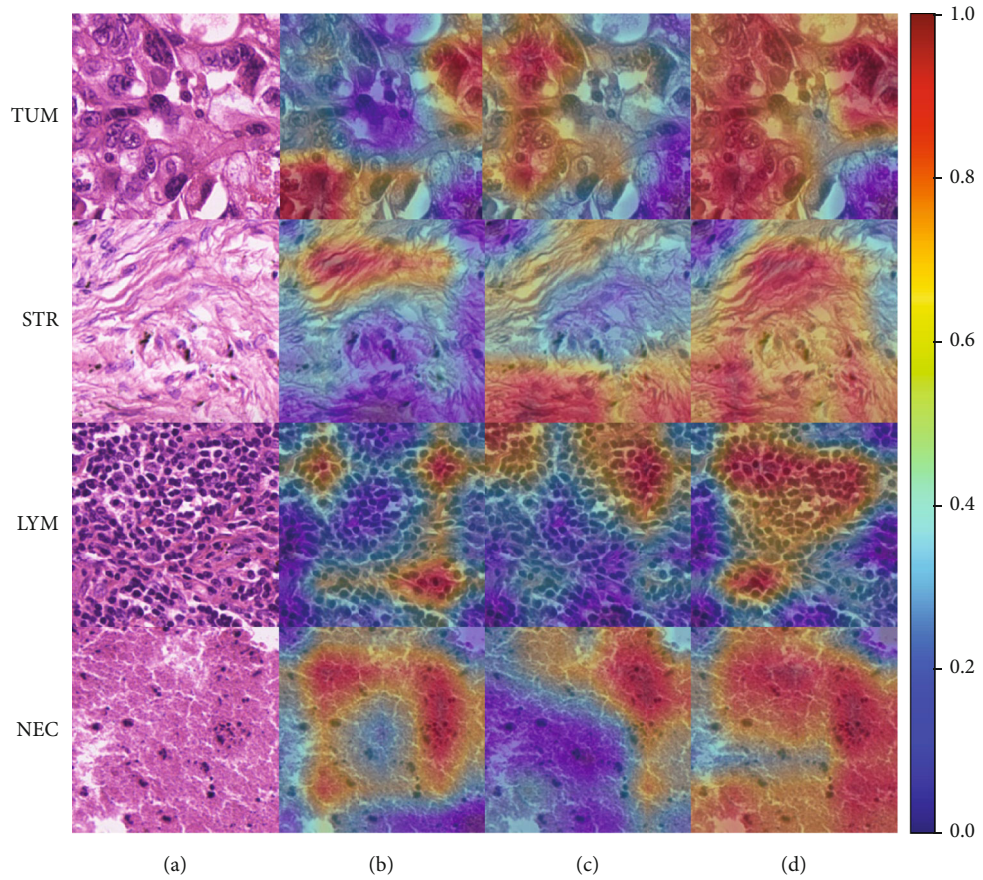


Figure 5: Heatmap of different models generated by Grad-CAM. (a) Slides of lung cancer scanned in 20x magnification factor, (b) heatmap of DeepTissue Net, (c) heatmap of ResNet50, (d) heatmap of the ResNet50 model with bilinear pooling module and attention module. It shows that the ResNet50 model with bilinear pooling module and attention module can detect the largest histopathological region. TUM: tumour epithelium; STR: stroma; LYM: tumour-infiltrating lymphocytes; NEC: necrosis.

Table 3: Results on colorectal cancer dataset (all models were pretrained on ImageNet).

Model	Average precision	Average recall	Average F1
VGG19 [30]	0.9650	0.9660	0.9655
DeepTissue Net [14]	0.9770	0.9775	0.9772
EfficientNet [31]	0.9779	0.9784	0.9781
ResNet50 [29]	0.9736	0.9746	0.9741
ResNet50+bilinear pooling module	0.9764	0.9771	0.9767
ResNet50+attention module	0.9789	0.9794	0.9791
ResNet50+bilinear pooling module+attention module	0.9823	0.9826	0.9824

Table 4: The classification F1 score of tissue types on colorectal cancer dataset.

Model	TUM	STR	LYM	MUC	MUS	NOR	BAC	DEB	ADI
VGG19 [30]	0.9870	0.9266	0.9773	0.9748	0.9662	0.9720	0.9586	0.9642	0.9580
DeepTissue Net [14]	0.9806	0.9413	0.9722	0.9784	0.9742	0.9806	0.9988	0.9711	0.9958
EfficientNet [31]	0.9814	0.9569	0.9827	0.9665	0.9671	0.9818	0.9982	0.9718	0.9942
ResNet50 [29]	0.9756	0.9243	0.9673	0.9852	0.9568	0.9866	0.9978	0.9729	0.9972
ResNet50+bilinear pooling module	0.9787	0.9502	0.9808	0.9614	0.9602	0.9869	0.9978	0.9771	0.9940
ResNet50+attention module	0.9803	0.9554	0.9810	0.9667	0.9826	0.9851	0.9974	0.9673	0.9940
ResNet50+bilinear pooling module+attention module	0.9860	0.9591	0.9845	0.9763	0.9693	0.9879	0.9982	0.9830	0.9966

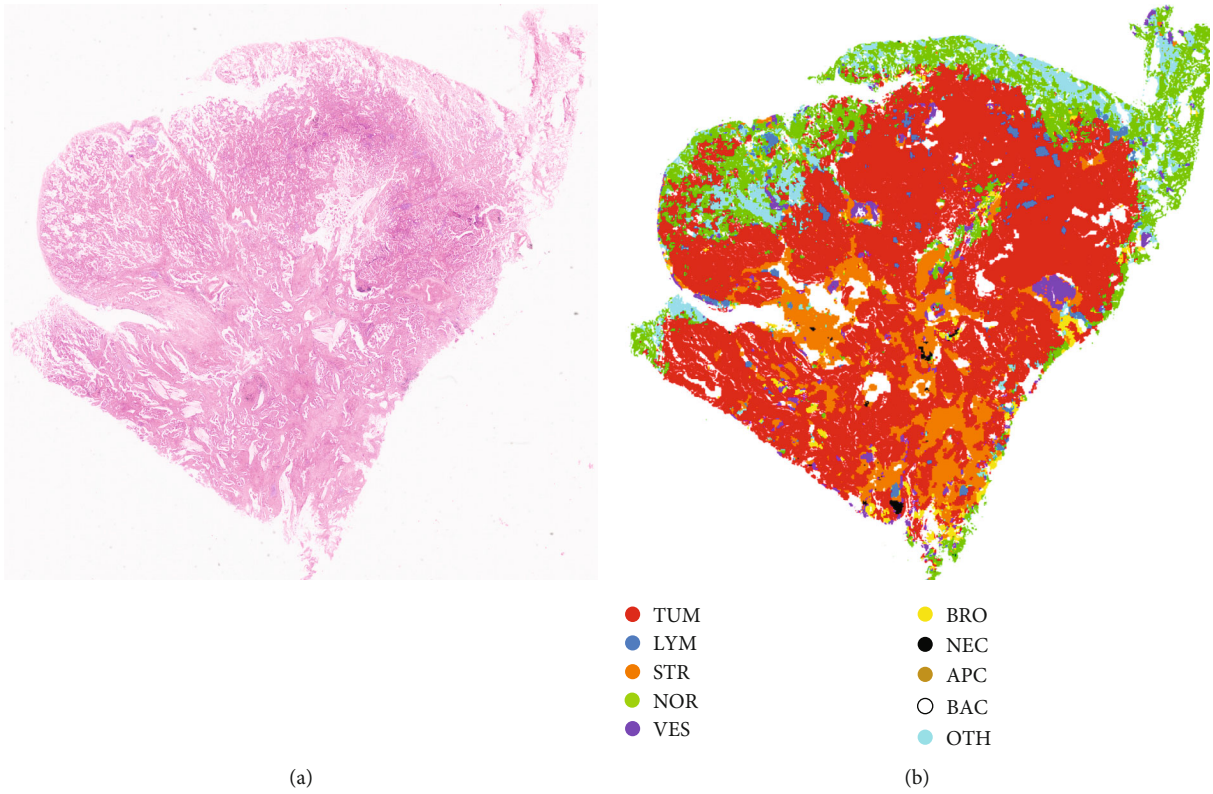


Figure 6: (a) Examples of H&E-stained WSIs in lung cancer and (b) corresponding segmented results.

different tissue types with large intraclass variability and small interclass variability.

To address these issues, we propose a two-stage automated tissue segmentation framework. In the first stage, large resolution WSIs are cut into small patches and then feed into the proposed classification model to predict separately. To alleviate the issue of large intraclass variability and small interclass variability, we introduce a Bilinear-CNN-based classification model. In previous studies, Kather et al. [18] used the VGG model to extract deep learning feature to identify tissue types. In Results, experimental results show that the VGG model does not perform well in the face of high heterogeneity of multiple tissue types. It is guessed that the model is designed for natural images and does not take into account the subtle features of pathological images. Chan et al. [8] improved the CNN model combined with Grad-CAM for the segmentation and classification of histological images. It relies on the task-specific postprocessing steps and generalizes poorly. Different from the previous approaches, this Bilinear-CNN-based classification network investigates the correct semantic correspondence between the intraclass by bilinear convolutional module and focuses on the distinguishable features of the interclass by soft attention module. The classification network can capture subtle features of pathological images well. The classification result of the proposed method is superior to state-of-the-arts recently. In addition, Figure 4 shows the proposed model has the fastest convergence speed than that of other models. Heatmaps generated by Grad-CAM provide visual interpretability of the classification results. It shows that this network is more sensitive to histopathological regions than classic CNNs. In the second stage, the classification results are stitched tile by tile to implement automatic tissue segmentation. Zhao et al. [17] input each image tile of entire WSI into the CNN model, including the background. In our method, the threshold segmentation algorithm is introduced to distinguish the tissue region from the background, and then, the category prediction of the tissue region is carried out. Compared with directly traversing the whole WSI, a large number of redundant computing overhead is reduced and the efficiency of WSI segmentation is improved.

Although our method is effective for lung cancer segmentation, some limitations remain. Our method uses the histopathological dataset for pretrained models to accelerate the training convergence. However, the histopathological dataset is not as convenient as ImageNet for different models because there are no prepared pretrained models like ImageNet. Moreover, the result of segmentation is patch-level, which is roughly compared with semantic segmentation. But the proposed framework uses image-level annotations to complete the segmentation tasks. Image-level annotations are easier to obtain than pixel-level annotations. Therefore, bridging the gap between image-level annotations and pixel-level segmentation will be the focus of the future investigation.

5. Conclusions

In this paper, we propose an automated tissue segmentation framework with two stages. In the first stage, the classification

model combines a bilinear convolutional module and soft attention module to improve the accuracy of tissue classification. In the second stage, the threshold segmentation algorithm distinguishes tissue from the background to avoid redundant computing of the background. The framework completes the tissue segmentation task via utilizing the image-level annotations.

Data Availability

Previously reported colorectal cancer data was used to support this study and is available at doi:10.5281/zenodo.4024676. This prior study (and dataset) is cited at a relevant place within the text as Reference [17]. The lung cancer data used to support the findings of this study have not been made available because of third-party rights. The code will be available at https://github.com/Hellowmyname/bcnn_attention_lung.

Conflicts of Interest

The authors declare no competing interest.

Authors' Contributions

RX and ZZW designed this study and prepared the first draft of the manuscript. ZZW, XPP, and ZBL designed the algorithm. LXY, HL, and ZYX contributed to data collection. CH, ZYF, and XPP contributed to the experimental design. CHL, XC, XPP, and ZYL revised the manuscript. All authors read and approved the final manuscript. Rui Xu, Zhizhen Wang, and Zhenbing Liu contributed equally to this work.

Acknowledgments

This research was supported in part by the Key R&D Program of Guangdong Province, China (Grant No. 2021B0101420006), the National Key R&D Program of China (Grant No. 2021YFF1201003), the National Science Fund for Distinguished Young Scholars, China (Grant No. 81925023), the National Natural Science Foundation of China (Grant Nos. 62002082, 82072090, 62102103, and 81771912), the Natural Science Foundation of Guangxi Province (Grant No. 2020GXNSFBA238014), the China Postdoctoral Science Foundation (Grant No. 2021M690753), Guangxi University Young and Middle-Aged Teachers' Research Ability Improvement Project (Grant No. 2020KY05034), the Major Achievement Transformation Foundation of Guilin (Grant No. 20192013-1), the Guangxi Key Research and Development Program (Grant No. GuikeAB21196063), the Innovation Project of Guangxi Graduate Education (Grant No. YCSW2021172), the Innovation Project of GUET Graduate Education (Grant No. 2021YCXS060), the High-Level Hospital Construction Project (Grant Nos. DFJHBF202105 and DFJH201805), and the Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (Grant No. 2022B1212010011).

References

[1] J. A. Barta, C. A. Powell, and J. P. Wisnivesky, "Global epidemiology of lung cancer," *Annals of Global Health*, vol. 85, no. 1, p. 8, 2019.

[2] J. Tang, D. Ramis-Cabrer, V. Curull et al., "B cells and tertiary lymphoid structures influence survival in lung cancer patients with resectable tumors," *Cancers*, vol. 12, no. 9, p. 2644, 2020.

[3] S. Wang, D. M. Yang, R. Rong et al., "Artificial intelligence in lung cancer pathology image analysis," *Cancers*, vol. 11, no. 11, p. 1673, 2019.

[4] C. Lu, C. Koyuncu, G. Corredor et al., "Feature-driven local cell graph (FLocK): new computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers," *Medical Image Analysis*, vol. 68, article 101903, 2021.

[5] C. Lu, K. Bera, X. Wang et al., "A prognostic model for overall survival of patients with early-stage non-small cell lung cancer: a multicentre, retrospective study," *The Lancet Digital Health*, vol. 2, no. 11, pp. e594–e606, 2020.

[6] C. Yan, K. Nakane, X. Wang et al., "Automated Gleason grading on prostate biopsy slides by statistical representations of homology profile," *Computer Methods and Programs in Biomedicine*, vol. 194, article 105528, 2020.

[7] S. Graham, Q. D. Vu, S. E. A. Raza et al., "Hover-Net: simultaneous segmentation and classification of nuclei in multi- tissue histology images," *Medical Image Analysis*, vol. 58, article 101563, 2019.

[8] L. Chan, M. Hosseini, C. Rowsell, K. Plataniotis, and S. Damaskinos, "HistoSegNet: semantic segmentation of histological tissue type in whole slide images," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10661–10670, Seoul, Korea (South), 2019.

[9] W. Lu, S. Graham, M. Bilal, N. Rajpoot, and F. Minhas, "Capturing cellular topology in multi-gigapixel pathology images," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1049–1058, Seattle, WA, USA, 2020.

[10] X. Pan, L. Li, H. Yang et al., "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing*, vol. 229, pp. 88–99, 2017.

[11] W. Liu, M. Juhas, and Y. Zhang, "Fine-grained breast cancer classification with bilinear convolutional neural networks (BCNNs)," *Frontiers in Genetics*, vol. 11, article 547327, 2020.

[12] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2424–2433, 2016.

[13] L. Li, X. Pan, H. Yang et al., "Multi-task deep learning for fine-grained classification and grading in breast cancer histopathological images," *Multimedia Tools and Applications*, vol. 79, no. 21-22, pp. 14509–14528, 2020.

[14] J. Xu, C. Cai, Y. Zhou et al., "Multi-tissue partitioning for whole slide images of colorectal cancer histopathology images with Deeptissue Net," in *Digital Pathology. ECDP 2019. Lecture Notes in Computer Science*, C. Reyes-Aldasoro, A. Janowczyk, M. Veta, P. Bankhead, and K. Sirinukunwattana, Eds., vol. 11435, pp. 100–108, Springer, Cham, 2019.

[15] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016.

[16] I. Hirra, M. Ahmad, A. Hussain et al., "Breast cancer classification from histopathological images using patch-based deep learning modeling," *IEEE Access*, vol. 9, pp. 24273–24287, 2021.

[17] K. Zhao, Z. Li, S. Yao et al., "Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer," *eBioMedicine*, vol. 61, article 103054, 2020.

[18] J. N. Kather, J. Krisam, P. Charoentong et al., "Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study," *PLoS Medicine*, vol. 16, no. 1, article e1002730, 2019.

[19] V. Anklin, P. Pati, G. Jaume et al., "Learning whole-slide segmentation from inexact and incomplete labels using tissue graphs," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. MICCAI 2021*, vol. 12902 of Lecture Notes in Computer Science, Springer, Cham, 2021.

[20] H. Yang, L. Chen, Z. Cheng et al., "Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study," *BMC Medicine*, vol. 19, no. 1, p. 80, 2021.

[21] Z. Li, J. Zhang, T. Tan et al., "Deep learning methods for lung cancer segmentation in whole-slide histopathology images—the ACDC@LungHP challenge 2019," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 429–440, 2021.

[22] X.-S. Wei, J. Wu, and Q. Cui, "Deep learning for fine-grained image analysis: a survey," 2019, https://arxiv.org/abs/1907.03069.

[23] Y. Wang and Z. Wang, "A survey of recent work on fine-grained image classification techniques," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 210–214, 2019.

[24] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN Models for Fine-Grained Visual Recognition," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015.

[25] Y. Li, N. Wang, J. Liu, and X. Hou, "Factorized bilinear models for image recognition," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2098–2106, Venice, Italy, 2017.

[26] C. Wang, J. Shi, Q. Zhang, and S. Ying, "Histopathological image classification with bilinear convolutional neural networks," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4050–4053, Seogwipo, 2017.

[27] K. Xu, J. Ba, R. Kiros et al., "Show, attend and tell: neural image caption generation with visual attention," http://arxiv.org/abs/1502.03044.

[28] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, 2016.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[31] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Convolutional Neural Networks with Swift for Tensorflow*, pp. 6105–6114, Apress, Berkeley, CA, 2019.

Hindawi
BioMed Research International
Volume 2022, Article ID 3341924, 10 pages
https://doi.org/10.1155/2022/3341924

Research Article

A Pulmonary Nodule Spiculation Recognition Algorithm Based on Generative Adversarial Networks

Jing Zhang[1], **Shi Qiu**[2], **Xiaohai Cui**[1] and **Ting Liang**[3,4]

[1]*Department of Thoracic Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China*
[2]*Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China*
[3]*Department of Radiology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China*
[4]*Department of Biomedical Engineering, The Key Laboratory of Biomedical Information Engineering of the Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China*

Correspondence should be addressed to Ting Liang; liangting31500@126.com

Received 13 April 2022; Revised 28 April 2022; Accepted 8 June 2022; Published 24 June 2022

Academic Editor: Chen Li

Copyright © 2022 Jing Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pulmonary nodules have been found as the main pathological change in the lung. Signs of pulmonary nodule lay the major basis for the recognition of the benign and malignant of pulmonary nodules. The spiculation of pulmonary nodules is one of the main signs. Pulmonary nodules are small in volume, so they are difficult to extract accurately. Moreover, the number of spiculation samples is limited, so it is difficult to build a stable network structure. Thus, a novel pulmonary nodule spiculation recognition algorithm is proposed. MCA (morphological component analysis) model is built to segment pulmonary nodules in accordance with the composition of pulmonary CT images. Subsequently, the maximum density projection mechanism is introduced to characterize the boundary features of pulmonary nodules to the maximum extent. Inspired by time series dynamic programming, this paper proposes DTW (dynamic time warping) distance to measure data similarity. Lastly, a semisupervised generative adversarial network is built to solve the problem of insufficient positive samples, and it is capable of recognizing pulmonary nodule spiculation. As revealed by the experimental result, the proposed algorithm exhibited strong robustness.

1. Introduction

Pulmonary nodule has been found as one type of main diseases in the lung. Malignant pulmonary nodule can be transformed into lung cancer, thus seriously affecting human health [1]. Spiculation lays a solid basis for judging the benign or malignant of pulmonary nodules. The spiculation signs in patients are classified as short or long spiculations. Short spiculation signs take on a critical clinical significance in the diagnosis of peripheral pulmonary cancer [2]. CT has been used as a main method to detect pulmonary nodules [3]. However, the proportion of pulmonary nodules is small in the lung, thus resulting in only limited pixels in CT images, so it is difficult for the detection.

Extensive researches have been conducted on computer-aided diagnosis, and some achievements have been made in the detection of pulmonary nodules, of which several representatives are presented below. Alilou et al. [4] proposed a complete framework for pulmonary nodule detection. Mehta et al. [5] analyzed the correlation between pulmonary nodule volume and benign or malignant. Van Ginneken et al. [6] analyzed the imaging features of pulmonary nodules with a computer. Elsayed et al. [7] proposed an automatic pulmonary nodule detection algorithm. Qiu et al. [8] built a Gestalt model for the detection of pulmonary nodules based on the principle of vision. Dou et al. [9] investigated pulmonary nodules from a 3D perspective. Feng et al. [10] built a CNN network to segment pulmonary nodules. Wilson and Devaraj [11] studied the radiology of pulmonary nodules. Gruetzemacher et al. [12] built a 3D deep learning framework to detect pulmonary nodules. Choi et al. [13] formulated the spiculation quantitative standard of lung cancer. Shen et al. [14] built an international deep hierarchical semantic dynamic network to classify pulmonary nodules.

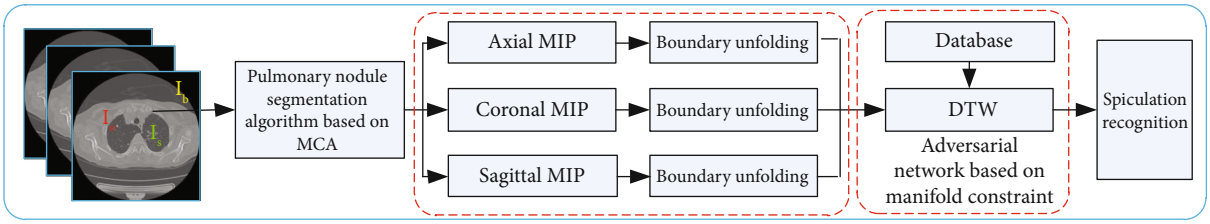


Figure 1: Recognition process of pulmonary nodule spiculation.

Qiu et al. distinguished between the types of pulmonary nodules [15] by constructing a recurrence plot. Qiu et al. [16] predicted potential burr features through machine learning. Li et al. [17] predicted cancer incidence rate based on burr sign. Marques et al. [18] built multitask CNN to achieve benign and malignant classification of pulmonary nodules. Zhou et al. [19] used a statistical model to predict tumor signs. He et al. [20] proposed an ISHAP-based interpretation model-guided classification method to predict tumor signs. Erdogdu et al. [21] used morphology-isolated pulmonary nodules to predict malignant tumors.

In brief, main detection and sign recognition problems of pulmonary nodules are presented as follows. (1) The boundary of pulmonary nodule is fuzzy, and the membership degree of boundary is not considered in the model, thus causing inaccurate boundary segmentation. (2) The limited feature of pulmonary nodules is observed from one dimension, thus causing inaccurate identification of signs. (3) Due to the limited number of spiculations, the positive and negative samples are not uniform, thus causing poor classification effect.

In this paper, a novel pulmonary nodule spiculation recognition algorithm is proposed based on the above three defects. (1) The model is established based on morphological component analysis to segment pulmonary nodules accurately. (2) A three-dimensional maximum density projection algorithm is proposed to integrate local three-dimensional information into two-dimensional image for the boundary representation of pulmonary nodules. (3) The generative adversarial network based on manifold constraints is built to enhance the stability of sign discrimination.

2. Algorithm

The recognition process of pulmonary nodule spiculation is built, as presented in Figure 1. First, MCA (morphological component analysis) model is established to segment pulmonary nodules. Then, the maximum density projection model is established from axial, coronal, and sagittal planes to characterize the boundary of pulmonary nodules. The boundary is expanded into time series, and DTW (dynamic time warping) algorithm is adopted to measure the similarity between the sample and the database for the spiculation recognition. To solve the problem of imbalance between positive and negative samples, an adversarial network based on manifold constraint is built. A complete recognition process of pulmonary nodule spiculation is established.

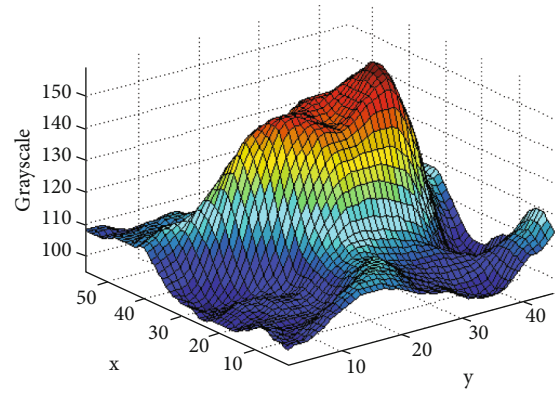


Figure 2: Gray scale image of a pulmonary nodule.

2.1. Segmentation Algorithm Based on Morphological Component Analysis. One pulmonary CT image $\mathbf{I}$ comprises pulmonary nodule $\mathbf{I}_o$, tissue area $\mathbf{I}_s$, and background $\mathbf{I}_b$.

$$\mathbf{I} = \mathbf{I}_o + \mathbf{I}_s + \mathbf{I}_b. \tag{1}$$

Pulmonary nodules are locally highlighted with limited pixels in CT images, as presented in Figure 2. The complete segmentation of pulmonary nodules is the premise of pulmonary nodule sign recognition.

In accordance with the analysis of CT composition of lung, morphological component analysis (MCA) is introduced to segment $\mathbf{I}_s$ [22]. MCA suggests that the signal comprises a series of component signals. The respective morphological component exhibits a unique dictionary sparsity corresponding to it. Accordingly, different signal components can be identified by dictionary.

Then, $\mathbf{I}$ can be represented by K morphological components:

$$\mathbf{I} = \sum_{k=1}^{3} \mathbf{I}_k = \sum_{k=1}^{3} \mathbf{D}_k \alpha_k, \tag{2}$$

where $\mathbf{D}_k$ denotes a super complete dictionary and α_k represents the sparse coefficient. In the solving process, the solution of the equation is not unique since the limit that the number of unknown parameters is greater than the number of equations. To ensure the sparsity maximization, MCA

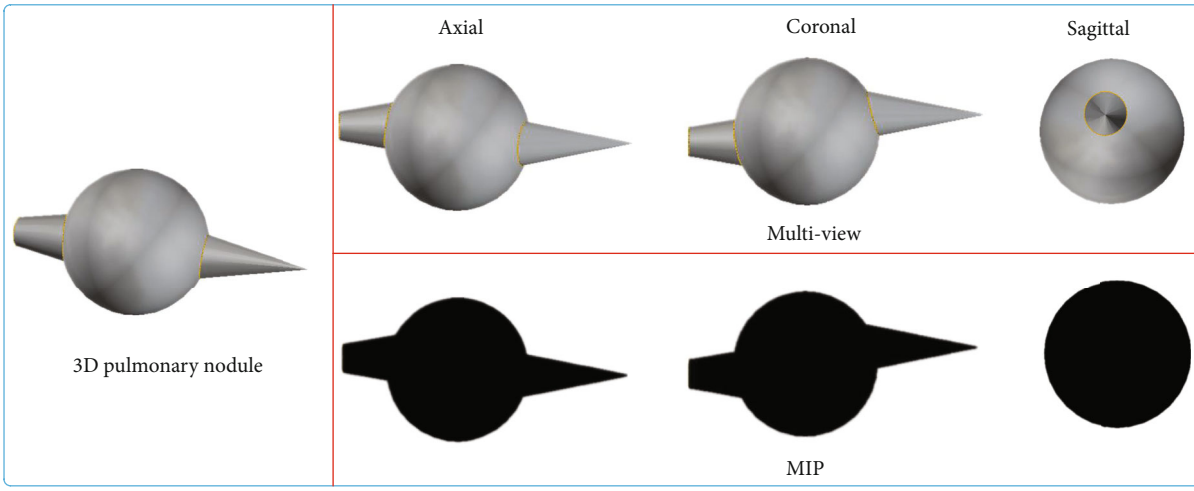


Figure 3: Multidirectional MIP effect.

model is established:

$$\min_{\boldsymbol{\alpha}_1,..\boldsymbol{\alpha}_3} \sum_{k=1}^{3} \|\boldsymbol{\alpha}_k\|_p \quad s.t. \left\| \mathbf{I} - \sum_{k=1}^{3} \mathbf{D}_k \boldsymbol{\alpha}_k \right\|_2 \le \Gamma. \tag{3}$$

where Γ represents the error between the original signal and the combined classification and $||.||_p$ is p norm.

Since there is a significant difference among $\mathbf{I}_o$, $\mathbf{I}_s$, and $\mathbf{I}_b$, and pulmonary nodules cannot be in the range of $\mathbf{I}_b$, only $\mathbf{I}_o$ and $\mathbf{I}_s$ are calculated by MCA model:

$$\begin{cases} \boldsymbol{\alpha}_o = \arg\min \|\boldsymbol{\alpha}_o\|_0 & s.t. \mathbf{I}_o = \mathbf{D}_o \boldsymbol{\alpha}_o, \\ \boldsymbol{\alpha}_s = \arg\min \|\boldsymbol{\alpha}_s\|_0 & s.t. \mathbf{I}_s = \mathbf{D}_s \boldsymbol{\alpha}_s. \end{cases} \tag{4}$$

The optimization problem is then transformed into

$$\left\{\boldsymbol{\alpha}_o^{opt}, \boldsymbol{\alpha}_s^{opt}\right\} = \arg\min \left(\|\boldsymbol{\alpha}_o\|_0 + \|\boldsymbol{\alpha}_s\|_0\right), \quad s.t. \|\mathbf{I} - \mathbf{D}_o \boldsymbol{\alpha}_o - \mathbf{D}_s \boldsymbol{\alpha}_s\|_2 \le \delta. \tag{5}$$

It is to find the sparse representation of the signal from the super complete dictionary, and its computational complexity increases with the increase in the number of dictionary columns. According to the BP algorithm, norm $L0$ in the equation can be replaced by norm $L1$, and then, Equation (5) becomes a linear programming problem:

$$\left\{\boldsymbol{\alpha}_o^{\text{opt}}, \boldsymbol{\alpha}_s^{\text{opt}}\right\} = \arg\min \left(\|\boldsymbol{\alpha}_o\|_1 + \|\boldsymbol{\alpha}_s\|_1\right) + \lambda \|\mathbf{I} - \mathbf{D}_o \boldsymbol{\alpha}_o - \mathbf{D}_s \boldsymbol{\alpha}_s\|_2 + \gamma V(\mathbf{D}_o \boldsymbol{\alpha}_o). \tag{6}$$

where λ and γ are the weights and V is the total change of pulmonary nodules.

As revealed by the above analysis, the extraction of dictionary is of great significance. Curvelet transform is adopted to sparsely represent the smooth part of the image, and local discrete cosine transform is employed to represent the texture part of the image.

Curvelet transform [23] represents the anisotropic elements in the image and better expresses the image edge, and the definition is as follows:

$$C(j, l, k) = \int_{R^2} I(x, y) \varphi_{j,l,k} dx dy, \tag{7}$$

where $C(j, l, k)$ is the curvelet transform coefficient and $\varphi_{j,l,k}$ is the wavelet basis.

Local discrete cosine transform (LDCT) can represent the component of texture in image, as defined below:

$$C(u, v) = a(u)a(v)b(u, v),$$
$$b(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} I(x, y) \cos\left[\frac{(2x+1)u\pi}{2M}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right], \tag{8}$$

where $C(\mu, v)$ denotes the discrete cosine transform coefficient.

Based on the decomposition of MCA algorithm, the threshold segmentation algorithm is employed to extract the pulmonary nodule region.

$$\left\{\mathbf{D}_o^{opt}, \mathbf{D}_s^{opt}\right\} = \arg\min \left(\|\mathbf{D}_o^* \mathbf{I}_o\|_1 + \|\mathbf{D}_o^* \mathbf{I}_o\|_1\right) + \lambda \|\mathbf{I} - \mathbf{I}_o - \mathbf{I}_s\|_2 + \gamma V(\mathbf{I}_o), \tag{9}$$

$$\mathbf{D}^* = \mathbf{D}^T \left(\mathbf{D}\mathbf{D}^T\right)^{-1}. \tag{10}$$

$\boldsymbol{\alpha}_o^*$ is selected from $\boldsymbol{\alpha}_o$ to reconstruct the image

$$\mathbf{I}_o = \mathbf{D}_o^{\text{opt}} \boldsymbol{\alpha}_o^*. \tag{11}$$

The watershed algorithm is applied. Set $C(M_i)$ as the catchment basin in the minimal area M_i related, and $T[n]$ denote the set of points satisfying the condition $g(x, y) < n$.

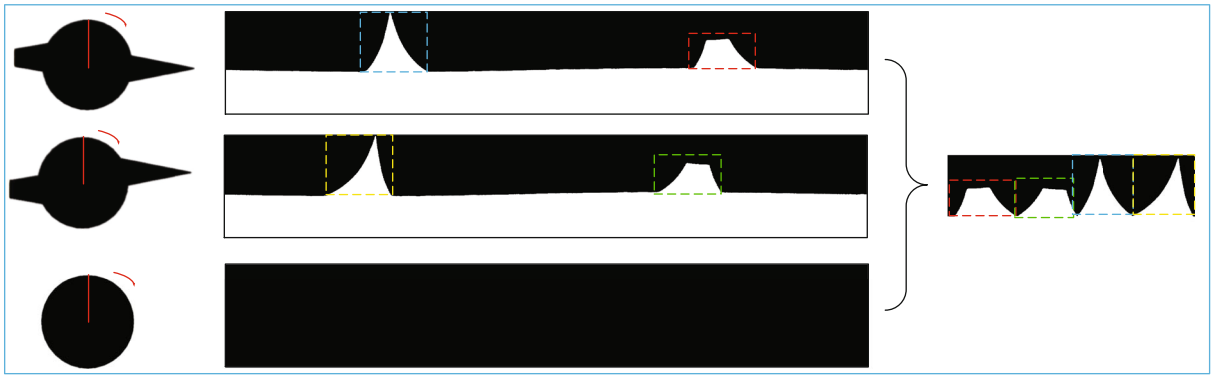


Figure 4: Boundary unfolding of pulmonary nodules.

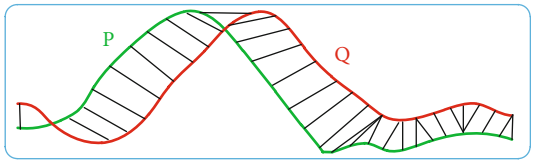


Figure 5: DTW distance.

According to

$$C_n(M_i) = \begin{cases} 1, & C(M_i) \cap T[n], \\ 0, & \text{others}, \end{cases} \qquad T[n] = \{(x,y) | g(x,y) \le n\}. \tag{12}$$

The image after segmentation can be obtained. To enhance the boundary signal intensity of pulmonary nodules, a rectangular region is selected for the region where pulmonary nodules are located. The gradient transformation is used as input:

$$G_o(x,y) = \sqrt{\left(\frac{\partial I_o(x,y)}{\partial x}\right)^2 + \left(\frac{\partial I_o(x,y)}{\partial y}\right)^2}. \tag{13}$$

Morphological open and close operation is applied to the gradient image to remove some redundant local minimum areas. The next step is to locate the local minimum pixel and its adjacent region. Lastly, the gradient image is applied twice by the watershed algorithm to obtain the region of pulmonary nodules.

2.2. Multidirection Projection Sign Analysis Algorithm. Pulmonary nodules have 3D tissue structure. However, when extracting from the conventional axial direction, pulmonary nodules can lead to incomplete burr information. However, 3D pulmonary nodule model has a large amount of calculation, which cannot meet the demand of real-time detection [24]. Thus, the concept of maximum density projection is introduced:

$$M(x,y) = \max\left(I_0(x,y) \cdots I_N(x,y)\right). \tag{14}$$

The local 3D information of pulmonary nodules can be displayed from one location. However, only from the MIP projection image of a single position, the spiculation perpendicular to the projection plane can submerge into the MIP image, as presented in Figure 3. Accordingly, we build a multidirectional MIP density projection algorithm to ensure that any spiculation is not missed.

According to the definition of spiculation, physicians mainly distinguish by observing the boundary characteristics of pulmonary nodules. It is necessary to choose an appropriate algorithm to analyze the boundary of pulmonary nodules. Conventional statistical analysis methods are based on the assumption that the data sequence has independence. Time series analysis [25] is different from conventional statistical methods, which focuses on the study and analysis of the interdependence of data series. Thus, in this paper, we take the center of the pulmonary nodule as the center of the circle (x_0, y_0) and the boundary closest to (x_0, y_0) as the starting point to expand the MIP images in different directions clockwise to form a time series for analysis, as presented in Figure 4.

$$\begin{cases} \rho = \sqrt{(x-x_o)^2 + (y-y_o)^2}, \\ \theta = \arctan\dfrac{y-y_o}{x-x_o}. \end{cases} \tag{15}$$

According to the definition of spiculation, the spicule area of protuberance is only considered, instead of the flat area [26]. For this reason, we extract the protuberance regions from the unfolded image and sort them into a new time series according to the degree of protuberance for analysis.

The sequence is compared with the trained data in the database to determine whether it is the spiculation. The DTW distance of dynamic programming is introduced based on the scalability of time series to measure the degree

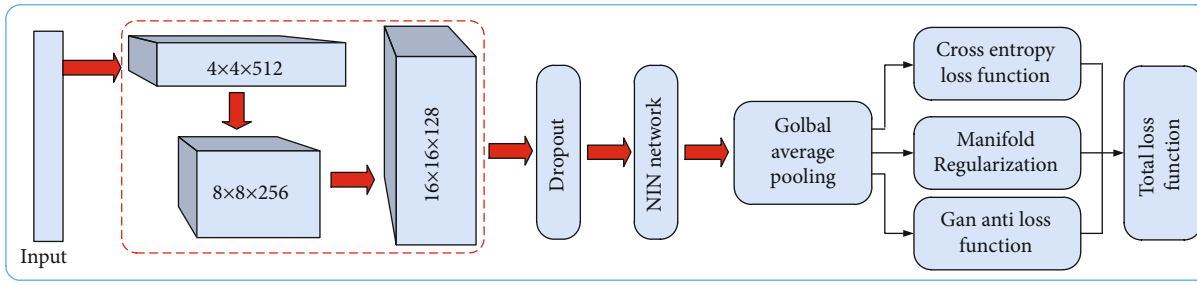


Figure 6: Network structure.

Table 1: Comparisons of image segmentation algorithms.

Algorithm	AOM	AVM	AUM	CM
3D	0.78	0.31	0.29	0.73
CNN	0.71	0.35	0.33	0.68
Watershed	0.75	0.33	0.31	0.7
DBRN	0.80	0.26	0.27	0.76
Ours	0.83	0.22	0.20	0.80

Table 2: Comparison of algorithm effect.

Algorithm	F	
	Normal	MIP
Axial	80	91
Coronal	82	90
Sagittal	80	90
Multidirectional	92	98

of data similarity [27].

$$\begin{cases} D(P,Q) = f(n,m), \\ f(n,m) = d(P_i,Q_i) + \min\left(f(i,j-1), f(i-1,j), f(i-1,j-1)\right), \\ f(0,0) = 0, f(0,i) = f(j,0) = \infty, \end{cases} \tag{16}$$

where $d()$ denotes Euclidean distance. The schematic diagram of DTW distance is illustrated in Figure 5. Signs of pulmonary nodules are distinguished according to the degree of similarity.

2.3. Semisupervised Generative Adversarial Network. During pulmonary nodule classification, the amount of spiculation data is less than the total number of pulmonary nodules. To solve the problem of uneven distribution of samples, scholars have proposed an unsupervised generative model based on Game Theory (GAN), as presented in Figure 6.

The model consists of generator A and discriminator B. The optimization process is the training process of binary minimax confrontation. The objective function is

$$\min_A \max_B V(A,B) = E(\log A) + E(\log(1-B)). \tag{17}$$

To apply GAN to semisupervised learning, it is assumed that classifier H_θ output k-dimensional vector $l = \{l_1, l_2, \cdots l_K\}$, and generate output probability through softmax function

$$P(y = i|x;\theta) = \frac{\exp l_i(x)}{\sum_{k=1}^{K} \exp l_k(x)}, \tag{18}$$

where $P(y = i|x;\theta)$ denotes the probability that the real sample x belongs to class I. Discriminator B_θ is adopted to distinguish the real sample x from the generated sample x_g:

$$B_\theta(x) = \frac{\sum_{k=1}^{K} \exp l_k(x)}{1 + \sum_{k=1}^{K} \exp l_k(x)}. \tag{19}$$

The loss function is yielded as

$$L = E\left[\log\left(1 + \sum_{i=1}^{K} \exp l_i(x_g)\right)\right] - E\left[\log \sum_{i=1}^{K} \exp l_i(x) - \log\left(1 + \sum_{i=1}^{K} \exp l_i(x)\right)\right]. \tag{20}$$

The derivation of L is as follows:

$$\nabla L = E\left[\sum_{i=1}^{K} P(y = i|x_g)\nabla l_i(x_g)\right] - E\left[\sum_{i=1}^{K} \{P(y = i|x, y \le K) - P(y = i|x)\}\nabla l_i(x)\right]. \tag{21}$$

To ensure L minimization, the updating process will promote the network to further enhance the current prediction $P(y = i|x;\theta)$. At the same time, the activation value of the predicted class connected neurons will be updated to the maximum, so the false prediction will be enhanced by map updating to get worse enhancement.

The classifier exhibiting better classification performance is obtained by considerable editing training samples at the early stage or by fully learning at the later stage. If the classifier is not mature and the resolution is not strong, it will produce wrong prediction. This problem will be reflected in the task of using a small number of labeled samples. Especially at the early stage of learning, the classifier is too poor to show enough discriminative power on the above samples.

Accordingly, the discriminator B is optimized by variable loss rather than the original counter loss. The update of the network at the early stage of training does not depend on the prediction of the classifier, and the standard deviation

matching loss in the generator is introduced to stabilize the GAN training. To optimize the performance of the classifier, manifold regularization term which is convenient for GAN calculation is added to the variable loss to further increase the robustness of the classifier to manifold local disturbance.

First, we improve the discriminator to alleviate the adverse effects of poor classifiers at the early stage of training on a small number of labeled samples:

$$B_\theta^*(x) = \frac{\exp\left((1/K)\sum_{k=1}^{K} l_k(x)\right)}{1 + \exp\left((1/K)\sum_{k=1}^{K} l_k(x)\right)}, \tag{22}$$

where B_θ^* is to distinguish the real image and generate the image based on the mean value of the activation value of the last layer K neurons. B_θ^* is updated to

$$\begin{cases} -\dfrac{\partial[-\log\{B_\theta^*(x)\}]}{\partial l_i(x)} = \dfrac{1}{K} P(fake|x; B_\theta^*), \\ -\dfrac{\partial\left[-\log\left\{1 - B_\theta^*\left(x_g\right)\right\}\right]}{\partial l_i\left(x_g\right)} = -\dfrac{1}{K} P\left(real\middle|x_g; B_\theta^*\right). \end{cases} \tag{23}$$

The above neurons are only activated on the real image x to form the subspace of the feature $l_i(x)$ of the real image. Since the generator provides a more realistic generated image x_g, the subspace becomes more compact to facilitate the classifier learning, and the focus is placed on the labeled samples in the limited subspace to facilitate the generalization of the classifier on it.

Discriminator B_θ^* is capable of effectively solving the problem of poor performance of the previous classifiers by not relying on the current prediction. However, for the classifiers with good classification performance in the middle and later learning stages, B_θ can be employed to replace them.

Manifold regularization term is introduced to increase the robustness of the classifier to local disturbances. For classifiers φ_θ, $\nabla_z \varphi_\theta(B(z^{(i)}))$ expresses the change of classification decision on manifold formed by $B(z^{(i)})$, and the change of φ_θ restricted by $B(z^{(i)})$ is written as

$$\left\|\varphi_\theta\left(G\left(z^{(i)} + \delta\right)\right) - \varphi_\theta\left(G\left(z^{(i)}\right)\right)\right\|^2 \approx \left\|\nabla_z^g \varphi_\theta\left(G\left(z^{(i)}\right)\right)\right\|^2 \delta z, \tag{24}$$

where $\nabla_z^g \varphi_\theta(G(z^{(i)}))$ is capable of training the classifier to resist the small disturbance on the manifold, $\|\nabla_z^g \varphi_\theta(G(z^{(i)}))\|^2 \delta z$ is adopted to approximate the regular term of the manifold, and ξ is employed to adjust the step size of updating the gradient direction of the manifold:

$$\Psi(\varphi_\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\|\varphi_\theta\left(G\left(z^{(i)} + \frac{\varepsilon\delta}{\|\delta\|}\right)\right) - \varphi_\theta\left(G\left(z^{(i)}\right)\right)\right\|^2. \tag{25}$$

The loss function of discriminator is written as

$$\begin{cases} \min\limits_\theta L = \alpha L_1 + \beta L_2 + \gamma\Psi(\varphi_\theta): \quad \alpha + \beta + \gamma = 1, \\ L_1 = E[-\log\{P(y = i|x; \theta)\}], \\ L_2 = E[-\log\{B_\theta^*(x; w)\}] + E\left[-\log\left\{1 - B_\theta^*\left(x_g; w\right)\right\}\right], \end{cases} \tag{26}$$

where α represents the weight coefficient of supervision loss, B represents the weight coefficient of unsupervised loss, and γ represents the manifold regular term weight coefficient.

To improve the stability of training GAN, the standard deviation matching loss is introduced:

$$\min_\eta L = \left\|\mu_\theta^l - \mu_{\theta,\eta}^l\right\|^2 + \left\|\sigma_\theta^l - \sigma_{\theta,\eta}^l\right\|^2, \tag{27}$$

where μ_θ^l, $\mu_{\theta,\eta}^l$, σ_θ^l, and $\sigma_{\theta,\eta}^l$ denote the mean and standard deviation of the real and generated images in the classifier output vector, respectively.

3. Experiment and Result Analysis

The experimental data were acquired from the international early lung cancer action project and the American lung imaging association database [28]. A total of 514 pulmonary nodule spiculation and 501 nonpulmonary nodule spiculation were labeled by two professional doctors as the basis of algorithm detection. CT data were 16 bits. The window width was adjusted to 1250~1600 HU to more effectively observe pulmonary nodules. It was normalized to [0,1] within the above range. The input image was normalized to 512×512.

3.1. Image Segmentation. To verify the performance of the proposed algorithm, the following indicators are introduced for measurement [15]:

$$\begin{cases} \text{AOM} = \dfrac{R_s \cap R_g}{R_s \cup R_g}, \\ \text{AVM} = \dfrac{R_s - R_g}{R_s}, \\ \text{AUM} = \dfrac{R_g - R_s}{R_g}, \\ CM = \dfrac{1}{3}\{\text{AOM} + (1 - \text{AVM}) + (1 - \text{AUM})\}, \end{cases} \tag{28}$$

where R_g denotes gold standard, R_s represents segmentation results, AOM and CM are proportional to the segmentation result, and AVM and AUM are inversely proportional to the segmentation results.

The segmentation effect of different algorithms was compared. As depicted in Table 1, Reference [9] built the model from 3D perspective to extract the overall structure of pulmonary nodules, and the built model is capable of achieving better segmentation of pulmonary nodules. Reference [10]

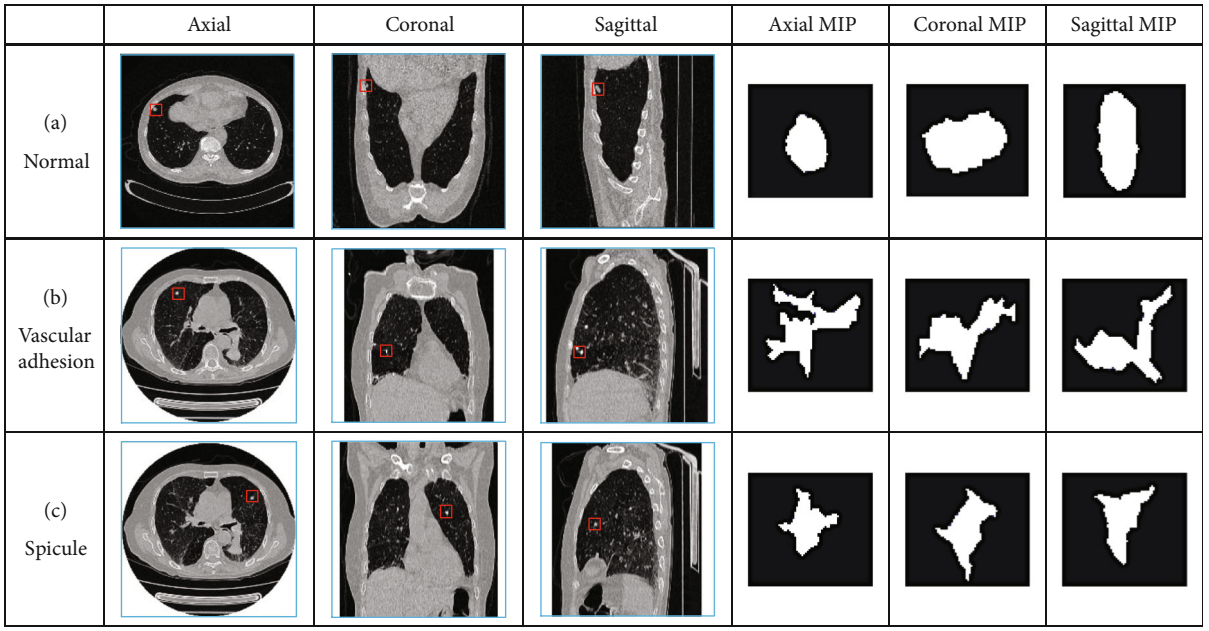


Figure 7: Multidirection MIP images of pulmonary nodule.

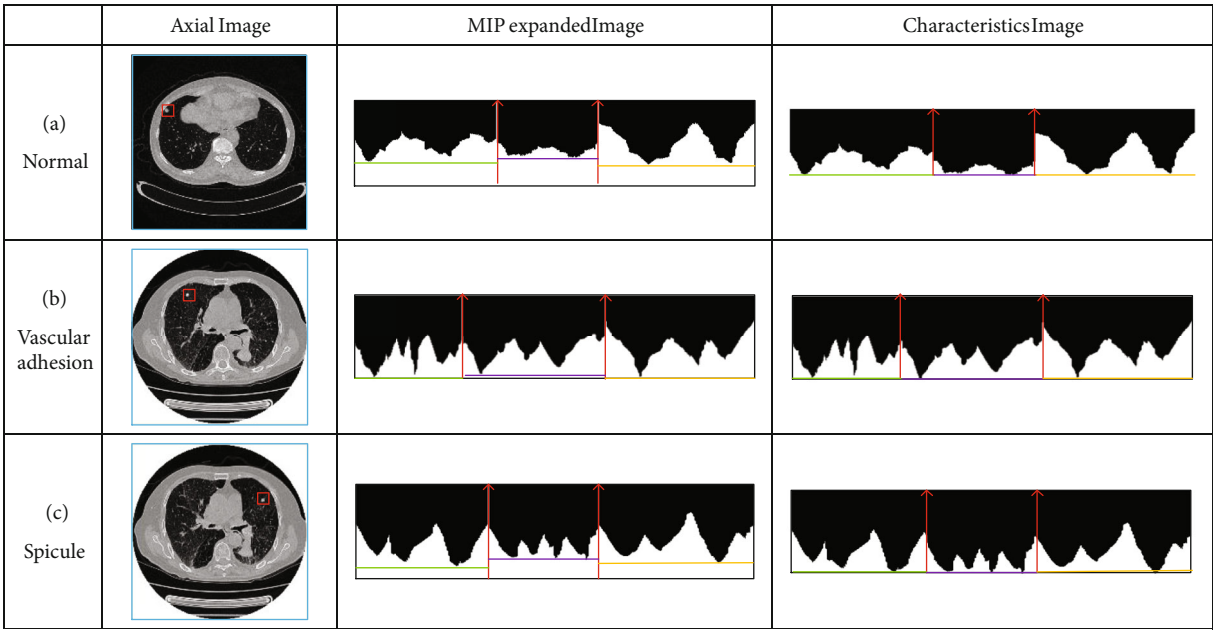


Figure 8: Computer features of spiculation.

built CNN network using a deep learning method to segment single-layer pulmonary nodules. Reference [29] combined watershed with pixel attribution probability to develop a network for the segmentation of pulmonary nodules. Reference [30] built the dual-branch residual network by fusing the whole and local information of the image for the segmentation of pulmonary nodules. The proposed algorithm is adopted to build the MIP model in accordance with three positions. It fuses the three-dimensional information into the local image and then uses the segmentation algorithm based on morphological component analysis to achieve effective results.

3.2. Multidirection MIP Algorithm Effect. To verify the effectiveness of the algorithm, the detection rate for measurement is presented.

$$F = \frac{A}{B}. \tag{29}$$

Table 3: Detection effect.

Algorithm	SEN	SPE	ACC	FPF
SVM	81	12	85	15
DTW	86	10	89	11
CNN	90	6	91	9
Ours	94	4	93	7

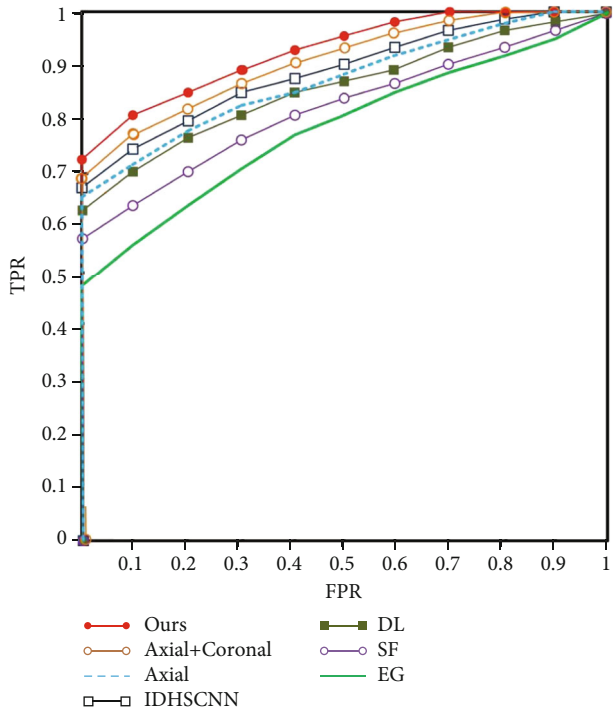


Figure 9: ROC curve.

The coronal and sagittal images were reconstructed in accordance with the CT axial images, and the axial, coronal, and sagittal images with the most obvious characteristics of pulmonary nodules were selected, respectively, for the comparison with the MIP algorithm, as listed in Table 2. It is not comprehensive to observe the signs of pulmonary nodules from a single location, which cannot indicate the overall characteristics of pulmonary nodules. With the increase in observation angles, the manifestation of pulmonary nodules increases to 92%. The MIP algorithm fuses local information into MIP image, which can show local 3D information, and significantly increases the detection rate of single position in a single frame. The detection rate of MIP algorithm can reach 98%, whereas there are still a few pulmonary nodule spiculation syndrome that are not detected, which is primarily because the segmentation is missed.

As depicted in Figure 7, the common pulmonary nodule (Figure 7(a)) is round in axial, coronal, and sagittal images. The boundary is smoother. In Figure 7(b), the type of pulmonary nodule with vascular adhesion is round in axial view and radial in coronal and sagittal view. If only from the axis detection, there will be the risk of missing detection. The internal structure of MIP can be displayed more intuitively through the three positions of MIP. In Figure 7(c), the features of spiculation are not obvious in axial, coronal, and sagittal view. After MIP, the local 3D information is fused into the image to get the intuitive spiculation structure. It lays a foundation for the recognition of spiculation feature.

3.3. Recognition Algorithm of Spiculation. The features extracted in this paper are intuitively presented in Figure 8. First, the MIP image of pulmonary nodules is expanded from the axial, coronal, and sagittal planes to the same coordinate system. Subsequently, the lowest point of the respective position is determined as the benchmark for removal to obtain the spiculation features. As depicted in Figure 8(a), the depression of common pulmonary nodules is small, and the boundary is relatively smooth, so more region is removed. Boundary of spiculation pulmonary nodule in Figures 8(b) and 8(c) is relatively steep, so the removal is relatively less. As a result, the accurate recognition of pulmonary nodules is achieved.

Since there are fewer positive samples and more negative samples in the identification process of pulmonary nodules, the following indicators are introduced to evaluate the performance of the algorithm:

$$\begin{cases} \text{SEN} = \dfrac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{SPE} = \dfrac{\text{TN}}{\text{TN} + \text{FP}}, \\ \text{ACC} = \dfrac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \\ \text{FPF} = 1\text{-ACC}, \end{cases} \tag{30}$$

where TP denotes true positive, FP represents false positive, FN is false negative, and TN is true negative. The results are listed in Table 3. SVM [4] algorithm is adopted to build linear classifiers for classification. DTW [27] algorithm uses the sequence correlation to set thresholds to achieve classification. CNN [10] algorithm is adopted to build a convolutional neural network for developing the correlation between pixel value and target and realize classification. The semisupervised generative adversarial network proposed in this paper comprehensively considers the correlation between positive samples and negative samples, thus improving the loss function and leads to good results.

To verify the proposed algorithm for spiculation recognition, ROC curves are introduced to measure its effect. As depicted in Figure 9, ROC curves indicate the results of axial, coronal, and sagittal positions, suggesting that the ability to extract pulmonary nodules from axial position is limited. With the increase in observation locations, the extraction performance of pulmonary nodules tends to increase, and the effect is optimal at three positions.

Reference [14] proposed an interpretable deep hierarchical semantic convolutional neural network (IDHSCNN) to fuse semantic information into images for the recognition of different signs. In Reference [16], deep learning (DL) is adopted to analyze the single-layer pulmonary nodules and identify the signs. In Reference [31], the shape feature is

employed to classify pulmonary nodules. In Reference [32], pulmonary nodule signs are identified based on edge gradient (EG) information. The above algorithms are adopted to build models for recognizing pulmonary nodules from different perspectives. However, the problem of limited feature samples is not considered, so the improvement is not significant in large samples. The semisupervised generative adversarial network proposed reduces the imbalance of positive and negative samples while achieving a better recognition effect

4. Conclusion

Pulmonary nodule spiculation is difficult to segment and recognize. To solve this problem, a segmentation algorithm based on morphological component analysis is proposed by focusing on the intrinsic features of pulmonary nodules. A multidirection MIP algorithm is proposed to solve the problems of inaccurate representation of single position and complex 3D modeling, which fuses multidirection information to accurately represent pulmonary nodules. To solve the problem of poor network stability caused by the imbalance of positive and negative samples, a semisupervised generative adversarial network is proposed to increase the recognition rate of pulmonary nodule spiculation. In future research, research on computer-aided detection of pulmonary nodules should be conducted to enhance the ability of automatic detection of pulmonary nodules. The characteristics exhibited by COVID-19 will be explored, and novel pneumonia detection and recognition systems will be developed.

Data Availability

Data are available from the International Early Lung Cancer Action Project (http://www.via.cornell.edu/lungdb.html).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jing Zhang and Xiaohai Cui contributed equally to this work.

Acknowledgments

This work is supported by the Clinical Research Award of the First Affiliated Hospital of Xi'an Jiaotong University, China, under Grant No. XJTU1AF2021CRF-015; the Shanxi National Science Foundation under Grant No. 2020JQ-518; the Open Project Program of the State Key Lab of CAD&CG, Zhejiang University, under Grant No. A2206; and the Science and Technology Rising Star of Shaanxi Youth under Grant No. 2021KJXX-61.

References

[1] S. G. Maldonado, S. Delorme, A. Hüsing et al., "Evaluation of prediction models for identifying malignancy in pulmonary nodules detected via low-dose computed tomography," *JAMA Network Open*, vol. 3, no. 2, article e1921221, 2020.

[2] C. A. de Pinho Pinheiro, N. Nedjah, and L. de Macedo Mourelle, "Detection and classification of pulmonary nodules using deep learning and swarm intelligence," *Multimedia Tools and Applications*, vol. 79, no. 21-22, pp. 15437–15465, 2020.

[3] J. Gu, Z. Tian, and Y. Qi, "Pulmonary nodules detection based on deformable convolution," *IEEE Access*, vol. 8, pp. 16302–16309, 2020.

[4] M. Alilou, V. Kovalev, E. Snezhko, and V. Taimouri, "A comprehensive framework for automatic detection of pulmonary nodules in lung CT images," *Image Analysis & Stereology*, vol. 33, no. 1, pp. 13–27, 2014.

[5] H. J. Mehta, J. G. Ravenel, S. R. Shaftman et al., "The utility of nodule volume in the context of malignancy prediction for small pulmonary nodules," *Chest*, vol. 145, no. 3, pp. 464–472, 2014.

[6] B. Van Ginneken, A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 286–289, Brooklyn, NY, USA, 2015, April.

[7] O. Elsayed, K. Mahar, M. Kholief, and H. A. Khater, "Automatic detection of the pulmonary nodules from CT images," in *2015 SAI Intelligent Systems Conference (IntelliSys)*, vol. 742-746, London, UK, 2015, November.

[8] S. Qiu, D. Wen, Y. Cui, and J. Feng, "Lung nodules detection in CT images using Gestalt-based algorithm," *Chinese Journal of Electronics*, vol. 25, no. 4, pp. 711–718, 2016.

[9] Q. Dou, H. Chen, L. Yu, J. Qin, and P. A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2016.

[10] X. Feng, J. Yang, A. F. Laine, and E. D. Angelini, "Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 568–576, Quebec City, QC, Canada, 2017.

[11] R. Wilson and A. Devaraj, "Radiomics of pulmonary nodules and lung cancer," *Translational Lung Cancer Research*, vol. 6, no. 1, pp. 86–91, 2017.

[12] R. Gruetzemacher, A. Gupta, and D. Paradice, "3D deep learning for detecting pulmonary nodules in CT scans," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1301–1310, 2018.

[13] W. Choi, S. Nadeem, S. Riyahi, J. O. Deasy, A. Tannenbaum, and W. Lu, "Interpretable spiculation quantification for lung cancer screening," in *International Workshop on Shape in Medical Imaging*, vol. 11167, pp. 38–48, Springer, 2018.

[14] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Systems with Applications*, vol. 128, pp. 84–95, 2019.

[15] S. Qiu, Q. Guo, D. Zhou, Y. Jin, T. Zhou, and Z.'. He, "Isolated pulmonary nodules characteristics detection based on CT images," *IEEE Access*, vol. 7, pp. 165597–165606, 2019.

[16] B. Qiu, J. Furst, A. Rasin, R. Tchoua, and D. Raicu, "Learning latent spiculated features for lung nodule characterization," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1254–1257, Montreal, QC, Canada, 2020.

[17] Y. Li, T. Wang, Y. F. Fu, and Y. B. Shi, "Computed tomography-based spiculated sign for prediction of malignancy in lung nodules: a meta-analysis," *The Clinical Respiratory Journal*, vol. 14, no. 12, pp. 1113–1121, 2020.

[18] S. Marques, F. Schiavo, C. A. Ferreira, J. Pedrosa, A. Cunha, and A. Campilho, "A multi-task CNN approach for lung nodule malignancy classification and characterization," *Expert Systems with Applications*, vol. 184, article 115469, 2021.

[19] C. Zhou, X. B. Liu, X. J. Gan, and X. Li, "Calcification sign for prediction of benignity in pulmonary nodules: a meta- analysis," *The Clinical Respiratory Journal*, vol. 15, no. 10, pp. 1073–1080, 2021.

[20] W. He, B. Li, R. Liao, H. Mo, and L. Tian, "An ISHAP-based interpretation-model-guided classification method for malignant pulmonary nodule," *Knowledge-Based Systems*, vol. 237, article 107778, 2022.

[21] E. Erdogdu, B. Ozkan, S. Duman et al., "Predictors of malignancy in patients with solitary pulmonary nodules undergoing pulmonary resection," *The Clinical Respiratory Journal*, vol. 16, no. 5, pp. 361–368, 2022.

[22] S. Wagner-Carena, M. Hopkins, A. Diaz Rivero, and C. Dvorkin, "A novel CMB component separation method: hierarchical generalized morphological component analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 494, no. 1, pp. 1507–1529, 2020.

[23] M. Arif and G. Wang, "Fast curvelet transform through genetic algorithm for multimodal medical image fusion," *Soft Computing*, vol. 24, no. 3, pp. 1815–1836, 2020.

[24] Y. Ren, M. Y. Tsai, L. Chen et al., "A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 2, pp. 287–295, 2020.

[25] E. L. Bullock, C. E. Woodcock, and P. Olofsson, "Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis," *Remote Sensing of Environment*, vol. 238, article 110968, 2020.

[26] P. Macauley, M. V. Cheing, A. Poor, and T. Newman, "All that is spiculated is not lung cancer: foreign body granuloma masquerading as lung cancer," in *C68. Thoracic oncology case reports II*, pp. A5777–A5777, American Thoracic Society, 2020.

[27] A. Rutkowska and M. Szyszko, "New DTW windows type for forward- and backward-lookingness examination. Application for inflation expectation," *Computational Economics*, vol. 59, no. 2, pp. 701–718, 2022.

[28] S. Qiu, J. Li, M. Cong, C. Wu, Y. Qin, and T. Liang, "Detection of solitary pulmonary nodules based on brain-computer interface," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 4930972, 10 pages, 2020.

[29] P. M. Shakeel, M. I. Desa, and M. A. Burhanuddin, "Improved watershed histogram thresholding with probabilistic neural networks for lung cancer diagnosis for CBMIR systems," *Multimedia Tools and Applications*, vol. 79, no. 23-24, pp. 17115–17133, 2020.

[30] H. Cao, H. Liu, E. Song et al., "Dual-branch residual network for lung nodule segmentation," *Applied Soft Computing*, vol. 86, article 105934, 2020.

[31] R. Niehaus, D. S. Raicu, J. Furst, and S. Armato, "Toward understanding the size dependence of shape features for predicting spiculation in lung nodules for computer-aided diagnosis," *Journal of Digital Imaging*, vol. 28, no. 6, pp. 704–717, 2015.

[32] G. Zhang, N. Xiao, and W. Guo, "Spiculation quantification method based on edge gradient orientation histogram," in *2014 International Conference on Virtual Reality and Visualization*, pp. 86–91, Shenyang, China, 2014.

[33] J. Chi, S. Zhang, X. Yu, C. Wu, and Y. Jiang, "A novel pulmonary nodule detection model based on multi-step cascaded networks," *Sensors*, vol. 20, no. 15, p. 4301, 2020.

Hindawi
BioMed Research International
Volume 2022, Article ID 4368928, 10 pages
https://doi.org/10.1155/2022/4368928

Research Article

Application of Surface-Enhanced Raman Spectroscopy in the Screening of Pulmonary Adenocarcinoma Nodules

Bowen Peng,[1] Huan Yan,[2] Runrui Lin,[2] and Gang Yin[3]

[1]*Nanjing University, School of Electronic Science and Engineering, Nanjing 210023, China*
[2]*School of Medicine, University of Electronic Science and Technology of China, Chengdu 611731, China*
[3]*Sichuan Cancer Hospital & Institute, Radiation Oncology Key Laboratory of Sichuan Province, Chengdu 610041, China*

Correspondence should be addressed to Gang Yin; cxqyguestc@163.com

Received 29 April 2022; Revised 3 June 2022; Accepted 9 June 2022; Published 23 June 2022

Academic Editor: Chen Li

Copyright © 2022 Bowen Peng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study is aimed at evaluating the feasibility of a screening method for the pulmonary adenocarcinoma nodules through surface-enhanced Raman spectroscopy (SERS). *Objective*. Using SERS to measure serum from pulmonary nodules and healthy subjects, intraoperative biopsy pathological diagnosis was regarded as the gold standard for labeling serum samples. To explore the application value of SERS in the differential diagnosis of pulmonary adenocarcinoma nodules, benign nodules, and healthy, we build a machine learning model. *Method*. We collected 116 serum samples from patients. Radiographically confirmed nodules less than 3 cm in maximum diameter in all patients, including 58 cancer (pathologic diagnosis: adenocarcinoma nodules, labeled as cancer) patients, 58 pathologic diagnoses as benign nodule (labeled as benign) patients, and 63 healthy (labeled as normal) people from the clinical laboratory of Sichuan Cancer Hospital. Gold nanorods were employed as SERS substrates. Support vector machine (SVM) was used to classify the normal, benign, and cancer sample groups, and SVM model evaluated using cross-validation. *Results*. The average SERS spectra of serum were significantly different between the normal group and the cancer/benign group. While the average SERS spectra of the cancer group and the benign group differed slightly, for the cancer, benign, and normal groups, SVM models can predict with 93.33% accuracy. *Conclusion*. This exploratory study demonstrates that the SERS technique based on nanoparticles in conjunction with SVM has great potential as a clinical auxiliary diagnosis and screening for pulmonary adenocarcinoma nodules.

1. Introduction

Lung cancer is the leading cause of cancer deaths globally, and the mortality rates of lung cancer occupy the first position among all cancers. It is estimated that there are 2.09 million new cases and 1.76 million deaths from lung cancer worldwide each year [1]. And the overall 5-year survival rate in lung cancer patients was only 18.7% in China [2]. However, the corresponding operation of radiotherapy and chemotherapy intervention for lung cancer of diagnosis in the early stage can improve patients' survival rate [3]. Currently, the primary screening method for early lung cancer is low-dose CT (LDCT) [4]. Early stage of lung cancer is mainly manifested in pulmonary nodules; CT may detect the early small pulmonary nodules, but it contains radiation risk [5]. Moreover, exposure to ionizing radiation should be avoided in certain groups, such as the elderly, women planning pregnancy, pregnant and lactating mothers, and people with thyroid problems. In addition, there are other methods for early-stage lung cancer screening using circulating or non-circulating markers [6]. The marker detection method is a body fluid-based analysis method, including circulating tumor cell detection and biomarker detection [7]. However, the low concentration and low correlation of specific markers also make it difficult for these liquid biopsies to obtain significant indicative results, and the price is relatively high. Therefore, there is a great need to explore simple, reliable, and low-cost methods to diagnose benign and malignant pulmonary nodules.

Through the continuous development of imaging examination methods, the necessary interventions such as surgical resection and radiofrequency ablation in the early stage of

lung cancer patients can improve their survival rate and quality of life significantly and even achieve an effectiveness of the cure. Bronchoscopy and CT-guided biopsy can provide an accurate pathological diagnosis. However, adenocarcinoma frequently occurs in subsegmental bronchi or alveoli, which are often inaccessible by bronchoscopy. The pathological results are always difficult to obtain. Also, CT-guided puncture biopsy is invasive, and some patients cannot tolerate it. CT is also affected by the degree of cooperation of patients and the operator's technical level. Presently, the discovery and diagnosis of pulmonary nodules mainly rely on CT examination. For the patient who cannot take a puncture biopsy, the diagnosis of pulmonary nodules still needs an intraoperative pathological biopsy. Adenocarcinoma is the primary pathological type of malignant pulmonary nodules. To improve the accuracy of preoperative diagnosis and early diagnosis, it is necessary to reduce unnecessary intervention for patients with benign nodules, particularly invasive surgery, and to develop radiation-free, low-injury, low-cost, and easily popularized approaches [8, 9].

Raman spectroscopy is an analytical method for studying molecular structure by analyzing the scattering spectra of target molecules by extracting information on molecular vibration and rotation. And the Raman characteristic spectrum of early cancer tissue and body fluid is different from that of normal tissue, making Raman spectroscopy a promising candidate for early cancer screening.

On the other hand, both markers and metabolites may face deficient levels in the early stage of tumor formation [10, 11]; it is necessary to effectively amplify the Raman signal for effective observation in the practical phase. Surface-enhanced Raman spectroscopy (SERS) is an enhancement effect on the surface of nanosized metal structures, which can escalate the intensity of ordinary Raman scattering by several orders of magnitude, even to the level of single-molecule detection. SERS has the advantages of high sensitivity, fast detection speed, low cost, and strong versatility and has been widely used in many fields such as biomedicine [12], drug detection, and bacterial detection [13]. The serum is composed of water, carbohydrates, proteins, and other components. During the cell carcinogenesis process, the types and contents of substances in serum will change. The Raman signal of many substances is relatively weak, and sometimes, the clinic-pathological symptoms of cancer patients may not be obvious enough to diagnose. By measuring the SERS spectra of biofluids, the characteristics of cancer could be reflected in the spectral difference to provide a reference for the diagnosis of cancer. Pichardo-Molina et al. studied Raman spectroscopy for identifying serum from healthy and breast cancer patients. The multivariate statistical analysis constructed the diagnostic model, obtaining a sensitivity of 97% and a specificity of 78% [14]. Chowdary et al. compared the Raman spectra of normal colon samples and cancerous colon samples ($N = 11$ cases, 102 Raman spectra). They found that the Raman spectra of cancerous tissues were relatively higher than those of normal tissues. It has a larger spectrum, indicating that cancerous tissue has more biological macromolecules than normal tissue, which provides the primary basis for the early diagnosis of colon cancer [15]. The characteristic peaks in the Raman spectrum are referred to as "fingerprints" and represent the biochemical composition of the sample. Raman spectroscopy can be used as an in vitro, minimally invasive, or even noninvasive early cancer detection technology.

This project plans to use precious metal nanoparticles as the SERS substrate. Serum from pulmonary adenocarcinoma patients, benign pulmonary nodules, and healthy patients (control group) was tested and classified by SVM to evaluate their detection ability.

In this study, we used a retrospective research method, with pathological diagnosis as the gold standard to label serum, and collected the surface-enhanced Raman spectroscopy of serum. SVM was used to establish a classification model for analyzing benign and malignant pulmonary nodules, which may be an additional tool for surgery or diagnosing pulmonary adenocarcinoma nodules.

2. Methods

2.1. Research Object. The Ethics Committee of Sichuan Cancer Hospital approved this study, and all serum data were collected at Sichuan Cancer Hospital from July 1, 2020, to November 31, 2021. Inclusion and exclusion criteria are as follows:

(1) Patients with malignant lung tumors: ① the diagnosis of lung cancer (adenocarcinoma) was confirmed by pathology, pulmonary nodules ≤ 3 cm; ② no surgery, radiotherapy, and chemotherapy were received before sample collection; ③ no combined systemic tumors; ④ no other diseases; ⑤ those who have no heart, liver, kidney, and other organ dysfunction; and ⑥ blood routine, liver, and kidney functions are normal

(2) Patients with benign lung tumors: ① the pathological diagnosis is benign (inflammation, pulmonary tuberculosis, lung cyst, etc.), pulmonary nodules ≤ 3 cm; ② no history of malignant tumor; ③ no other diseases; ④ no heart, liver, kidney, and other organ dysfunction; and ⑤ blood routine, liver, and kidney function are normal

(3) Normal control group: ① no history of malignant tumor; ② no heart, liver, kidney, and other organ dysfunction; ③ normal blood routine, liver, and kidney function; ④ body mass index (BMI) range of 18.5-23.9; and ⑤ no lung nodules nodes and inflammatory lesions

Serum collection: after fasting for 8 hours, the subjects take 2 mL of fasting venous blood at 6-7 o'clock the following day without adding anticoagulants. The upper serum was taken as a sample and frozen in a -80°C refrigerator after centrifugation at 4000 R/min for 10 minutes.

They were selected from 116 patients with pulmonary nodules in Sichuan Cancer Hospital, including 58 cancer (adenocarcinoma nodules, labeled as cancer) patients, 58 benign patients (labeled as benign), and 63 healthy (labeled

as normal) people. The mean age of the cancer group is 57.5 ± 9.6, including 34 males and 24 females. The mean age of the benign patient group is 52.4 ± 11.8, including 23 males and 35 females. A total of 63 healthy people (26 males, 34 females, age 60.4 ± 13.7 years old) were selected from the health examination of our hospital as the control group. General information of all investigated individuals, such as gender, age, clinical stage, tumor type, and tumor growth location, is shown in Table 1. All individuals participating in the study gave informed consent and signed informed consent for the study purpose.

2.2. Substrate Production. We prepared a freshly mixed solution with 0.01 M sodium borohydride dissolved in 0.01 M sodium hydroxide. Take an HAuCl4 solution (10 mL, 0.5 mM) in 0.1 M CTAB added to 460 μL of the sodium borohydride aqueous solution, and rapidly stir the mix solution, leading to a change in color from greenish to light brown. Silver nitrate (70 μL, 0.1 M) solution was added to an HAuCl4 solution (10 mL, 0.5 mM) in 0.1 M CTAB, followed by the addition of hydroquinone aqueous solution (500 μL, 0.1 M), and the resulting mixture was hand-stirred until it became clear. Next, 160 μL of seed solution was added, and the growth solution was mixed thoroughly and allowed to age overnight [16]. Figure 1 is the transmission electron microscope image of the synthesized gold nanorods.

2.3. Test Method. The instrument used is the Horiba iHR550 remote fiber-coupled Raman microscope system. It carries a fiber-coupled microscope that provides complete microscopic analysis capabilities simple and cost-effective manner. The serum of the investigated individuals obtained from the Laboratory Department of the Provincial Cancer Hospital will be stored in a -80°C freezer until subsequent testing. Before starting the test, concentrated sulfuric acid was added to a 30% hydrogen peroxide solution (volume ratio is 7 : 3), and glass slides with a size of 1 cm × 1 cm were soaked in the solution for 1 minute. The slides were then washed with deionized water and dried to ensure cleanness. The serum to be tested and the gold nanorods were mixed at a ratio of 1 : 1, mixed evenly with a pipette, and dried at room temperature for 30 min. Then, a drop of the hybrid solution was placed on the cleaned glass slide for SERS detection. Each sample measures 8 spectra at different positions. The experiment was carried out in a dark room, all operated by the same person. Spectral calibration of the Horiba iHR550 using a single crystal silicon wafer (which continuously emits a characteristic peak at 520.7 cm^{-1}) is required before sampling. After calibration, the SERS measurement was performed with an incident laser at 785 nm, whose power was 5 mW. The accumulation time of each measurement was 10 seconds. Multiple spectra were collected from each sample to avoid an abnormal sample composition.

2.4. Data Processing and Analysis. The 1217 spectra of 179 individuals were preprocessed, and all Raman spectral lines were intercepted in the spectral range of 500 cm^{-1}-1700 cm^{-1}. Due to the instrument, the carrier substrate, the acquired spectra had some background shifts, so noise reduction and debaselining are required. After obtaining the spectrum, first debaseline and denoise each original spectrum using the BEADS algorithm [17], preserving the main peak information and relative intensities. Then, standardize the dispersion of each spectrum (subtract the minimum value of the spectral data and divide it by the form of range), and normalize the spectral data to the range of 0~1 so that all the spectra of each sample occupy the proportion of the following average processing of the same ratio. All spectra for each sample are summed, and dispersion normalized again. That becomes the normalized data for training.

2.5. Model Establishment and Verification. A support vector machine (SVM) is a linear binary classifier [18]. Because of the multivariate data set, a one-vs-all multiclass implementation was used to transform this binary classifier into a multiple class discrimination model [19, 20]. In brief, a binary SVM can be extended to distinguish between multiple classes by training an SVM model that all samples in a particular class y are labeled as positive, i.e., $c = +1$. Similarly, all other samples are labeled as negative, i.e., $c = -1$; then, finding the class that maximizes the decision function can determine the decision boundary.

Then, a dual-layer cross-validation scheme was used to avoid overestimating and overfitting [21]. In the dual-layer cross-validation, loop data were split into $N = 5$ folds, and each fold as a completely independent test set was used only once. Then, the external training set uses the remaining $N - 1$ folds and splits into $M = 5$ folds in the internal cross-validation loop. Furthermore, data were classified into the internal training data set ($M - 1$ fold) and the validation data set (1 fold) within this inner loop. The internal training set was only used to build the hyperparameter optimization method, and the validation data set was only used to validate, resulting in $M = 5$ validation accuracies. Then, choose the optimal model parameters that show the highest validation accuracy to build the SVM model with the external training set. Lastly, the external training set was used to predict the optimal model. The process is repeated until all the samples are independently tested. Afterward, the testing accuracy is computed using the average of $N = 5$ folds.

3. Results

3.1. Clinical Features. In the malignant lung cancer group, female patients accounted for 60.3% of the total number of lung cancer patients, the median age was 56, and the range was 32 to 82 years old; pathological stage I accounted for 61.8% of the total lung cancer group. In the benign lung nodule group, female patients accounted for 41.3% of the total benign group, the median age was 53 years, and the range was 23 to 77 years; among normal people, female patients accounted for 53.9% of the normal population, with a median age of 60 years and a range of 25 to 84 years. There was no significant difference in gender ($p = 0.574$) and size ($p = 0.735$) of benign and malignant tumors between groups, and there was no significant difference in tumor location ($p = 0.707$).

Table 1: Clinical characteristics of individuals under investigation.

	Cancer ($n = 58$)	Benign ($n = 58$)	Normal ($n = 63$)	p value
Type of lung cancer	Adenocarcinoma			
Age, y	57.5 ± 9.6 (32-82)	52.4 ± 11.8 (23-77)	60.4 ± 13.7 (25-84)	0.486
Sex				
Male	23	34	26	0.574
Female	35	24	34	
Pathological stage				
c-stage IA	28	Nan	Nan	
c-stage IB	15	Nan	Nan	
c-stage IIA	10	Nan	Nan	
c-stage IIB	5	Nan	Nan	
Maximum nodules size, cm	1.55 ± 0.73	1.43 ± 0.6		0.735
Location				0.707
Upper lobe of a lung	35	31	Nan	
Middle lobe of a lung	7	7	Nan	
An inferior lobe of a lung	16	20	Nan	

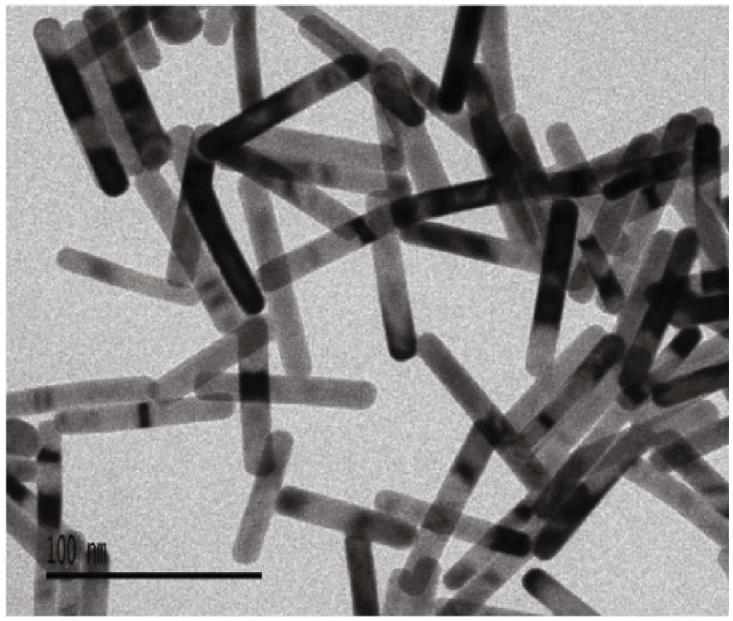


Figure 1: Transmission electron microscope image of the synthesized gold nanorods.

3.2. Raman Spectroscopy and Statistical Analysis. As shown in Figure 2, there is a certain difference in the mean Raman spectrum of serum between cancer patients and healthy people. Because cancer, benign, and normal people contain a large number of biological macromolecules in the serum, but the content of each biological macromolecule in the serum is different, there is a lot of similar vibration information in the spectrum, but the peak intensity is different. Several representative peaks may involve differences between the sample and control groups. The peaks at 650 cm^{-1} and 857 cm^{-1} represent vibrations of tyrosine, the amount of which increases with cancer progression, while the peak at 734 cm^{-1} may represent vibrations of hypoxanthine. The mean spectra of the benign and malignant groups were almost indistinguishable at the 734 cm^{-1} peaks, indicating that the experimental conditions were practically identical in the two groups. Phenylalanine, a lung cancer biomarker, is located at the peak of 1004 cm^{-1} in the Raman spectrum [22]. Metabolic disorders may cause abnormal content changes of phenylalanine in the lung cancer group. The amide I peak at 1640 cm^{-1}, the amide III peak at 1270 cm^{-1}, and the tryptophan peak at 1370 cm^{-1} (all of which are higher in cancer patients) have a relatively high average spectral intensity. Raman peaks, such as the uric acid peak at 637 cm^{-1} (higher intensity in cancer patients), also have a certain degree of difference between cancer and benign, indicating that the serum uric acid content is higher in lung cancer patients. This phenomenon may be related to the

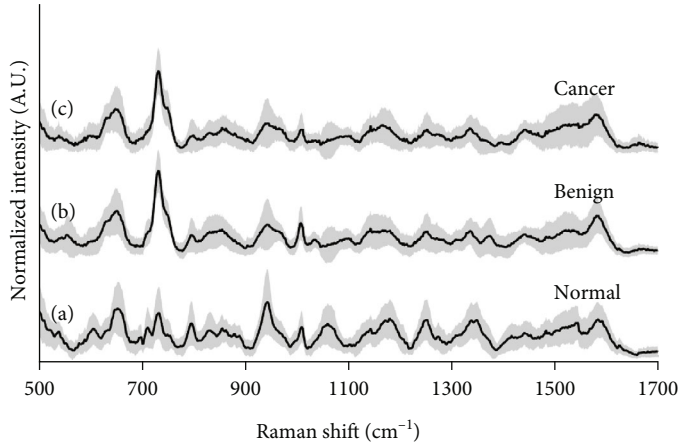


Figure 2: The average normalized intensity SERS spectra of three groups. The three colored spectra are averaged and normalized spectra with a standard deviation of (a) normal (healthy) group, (b) benign group, and (c) cancer group.

disorder of purine metabolism in lung cancer patients [23]. Therefore, this shows that it is feasible to use Raman spectroscopy to distinguish benign from malignant.

Figure 2 depicts the mean normalized spectra of the cancer, benign, and normal groups. Aid to appreciate the variability of the data, shaded error bars were meant to indicate one standard deviation from the mean. Figure 3 depicts the difference spectra, and we can see that spectra from the cancer and benign groups have the most resembling molecular fingerprint. Hence, we hope the classifier mislabels some samples from these groups. However, normal spectra show evident diversity among cancer and benign groups. Figure 3 shows the 3 groups have an enormous difference in the 734 cm^{-1} bands. The peak at 734 cm^{-1} was assigned to the existence of hypoxanthine. It was found that the SERS intensity of hypoxanthine gradually increases with exposure time because of metabolic degradation in the blood complex [24].

Figure 4 shows the classification results of the support vector machine (SVM) model proposed by us, and the overall accuracy is 93.33%. We can see that only 5 were misclassified and labeled as benign in the 58 cancer samples. And 6 samples were mislabeled, 5 cases were classified as cancer, and 1 case was classified as normal among the 58 benign cases. And there is one sample that was not classified correctly and was given a cancer label in the 63 normal controls. Different spectra show a high error rate between cancer and benign diseases because their molecular fingerprints are highly similar. The ROC curve of the SVM model is shown in Figure 5. The AUC of the three groups (cancer, benign, and normal) is 0.98, 0.97, and 0.99, respectively. These results show that our model has high distinguishing power for cancer.

Table 2 contains the sensitivity, accuracy, and specific values for each category. Our model can specifically distinguish the three groups (i.e., cancer, benign, and the normal). The significant advantage of our classifier is its high sensitivity to normal samples (Se = 0.99). Our two-tier cross-validation and testing of the samples are entirely independent of the samples used in the optimization, ensuring a reliable and repeatable approach that can be extended to real-world clinical settings.

4. Discussion

In the development of lung cancer, with the change in metabolic levels in patients, the structure and quantity of many biomacromolecules will also change. SERS can detect the changes in chemical bonds corresponding to these biomacromolecules, so expected to be an effective means of early cancer screening. Early lung cancer screening is still mainly based on low-dose CT. To some extent, a biopsy can help with early diagnosis. However, obtaining bronchoscopy results appeared to be difficult because adenocarcinoma arises in the subsegmental bronchi or alveoli. Such patients generally need an intraoperative biopsy to determine whether the nodules are benign or malignant. Because adenocarcinoma accounts for the majority of pulmonary nodules, the preoperative auxiliary judgment of benign and malignant pulmonary nodules plays an essential role in the clinic that may reduce the risk of excessive medical treatment. From the results of this study, we can see a significant difference between the serum of normal people and patients with benign and malignant pulmonary nodules. The SVM model can predict malignant nodules, benign nodules, and normal with 93.33% accuracy. So the surface-enhanced Raman scattering of serum has great potential in the auxiliary diagnosis of lung adenocarcinoma nodules.

Serum SERS's most crucial detection mechanism is based on enhancing the electromagnetic field caused by the strong localized surface plasmon resonance between the aggregated gold nanoparticles. When the gold colloids are mixed with the serum samples, the biochemical substances in the serum are nonspecifically adsorbed on the surface of gold nanoparticles, resulting in aggregation. And another mechanism involves chemical enhancement caused by charge transfer between the surface of nanoparticles and serum molecules [25, 26]. Due to these two main mechanisms, the Raman signal of the serum can be greatly enhanced, which provides a unique opportunity to explore

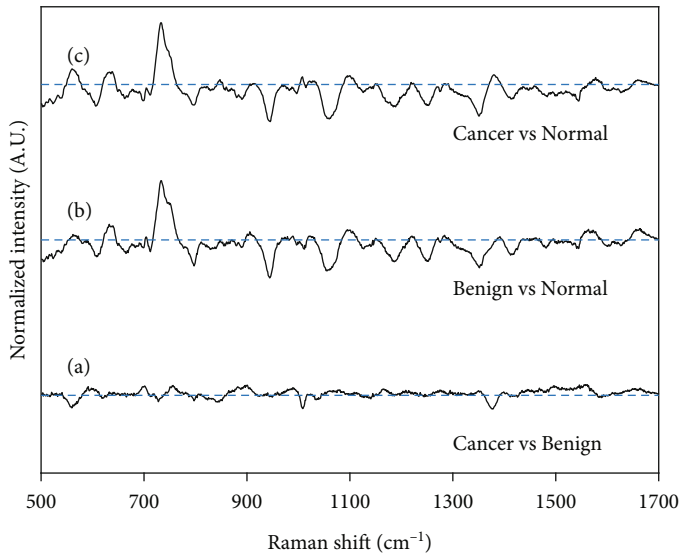


Figure 3: The difference between the three groups, (a) cancer vs. benign groups, (b) benign vs. normal groups, and (c) cancer vs. normal groups.

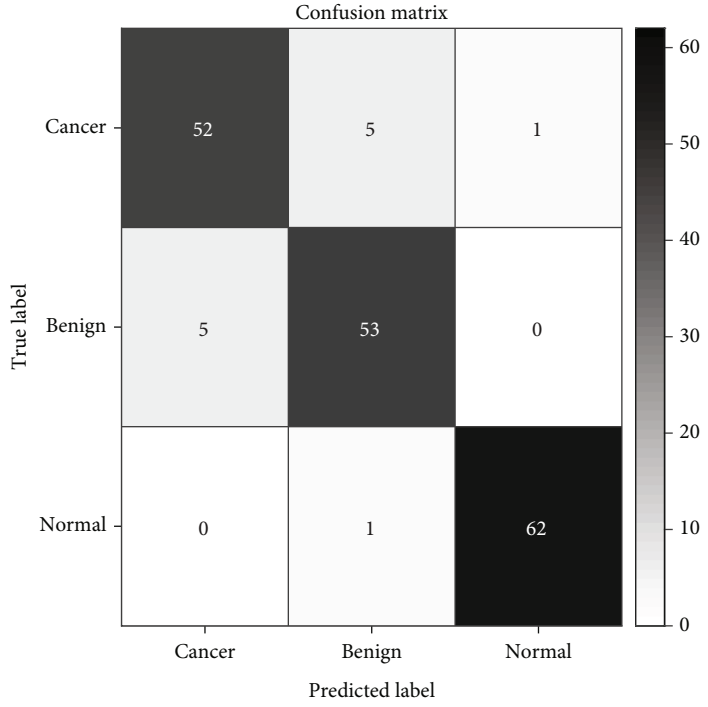


Figure 4: The classification results from our proposed SVM model.

subtle changes in the serum of patients with early pulmonary nodules at the molecular level.

Compared with normal, the serum SERS signal of the cancer and benign group has some similar changes. The serum samples of malignant and benign pulmonary nodules still contain identical components. The changes in protein-related SERS peaks at 1004 cm^{-1} and 1665 cm^{-1} are usually related to cancer progression and development, which may reflect unique content and conformational changes [27]. And compared with normal controls, a decrease in SERS signal was found at 1004 cm in malignant nodule samples, indicating a reduction in the percentage of phenylalanine and tryptophan relative to the total SERS active components in the blood of cancer patients. The Raman peak at 1004 cm^{-1} is a significant and stable Raman signal that reflects changes in phenylalanine in tissues, cells, blood, and other samples [27].

We know that the peaks at 650 cm^{-1} and 857 cm^{-1} represented the vibration of tyrosine [28], whose content increases as the cancer progresses [29]. And the peak at 734 cm^{-1} represents hypoxanthine [30]. The Raman band near 734 cm^{-1} is the most profound feature among these

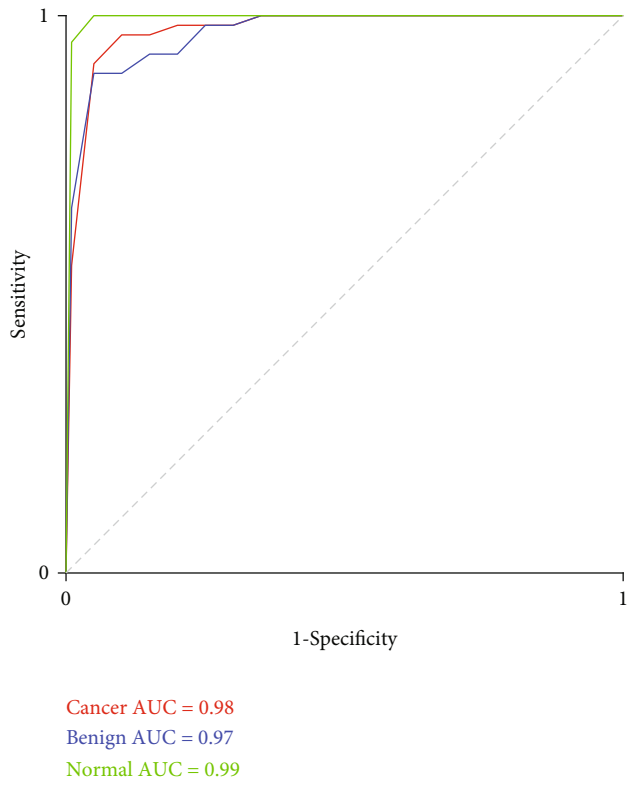


Figure 5: Receiver-operating characteristic (ROC) curve of SVM model.

Table 2: Performance parameters of the SVM.

Class	Performance parameter	Value ± Std	95% CI
Cancer	Sensitivity	0.92 ± 0.05	0.85-0.98
	Specificity	0.95 ± 0.05	0.89-1.00
	Accuracy	0.94 ± 0.02	0.91-0.96
Benign	Sensitivity	0.92 ± 0.11	0.77-1.00
	Specificity	0.96 ± 0.03	0.92-0.99
	Accuracy	0.94 ± 0.04	0.89-0.98
Normal	Sensitivity	0.98 ± 0.03	0.94-1.00
	Specificity	0.99 ± 0.01	0.96-1.00
	Accuracy	0.98 ± 0.02	0.97-1.00

peaks. However, it was found that the SERS intensity of hypoxanthine gradually diminished with exposure time [31]. Although the dynamic factor will hinder the significance of this band, the approximately zero difference between the benign and malignant groups near 734 cm^{-1} indicates that the sampling condition between these groups was almost identical.

On the other hand, uric acid, xanthine, and hypoxanthine are oxidation products of purine degradation and metabolism in the human body. The metabolic pathway of purine refers to the conversion of hypoxanthine to xanthine and xanthine to uric acid under the action of xanthine oxidase. Therefore, xanthine is the intermediate product of purine metabolism [32]. Xanthine is also an essential biomolecule in the human body. It is a purine base widely distributed in human organs, fluids, and other organisms. It is produced by the catabolism of guanine and adenosine triphosphate in animal muscle tissue. However, its accumulation in the body can lead to death [33]. Prajda's study showed that the imbalance of purine metabolism in liver cancer could be detected. The results showed that xanthine oxidase activity decreased in all liver cancer tissues compared with normal liver tissues, regardless of the tumor's malignant degree, differentiation degree, and growth rate. This phenomenon seems to be specific to the process of tumor genesis [34, 35]. At the same time, compared with normal tissues, the activity of xanthine oxidase in liver cancer is decreased, which reflects the increase in growth rate or different degrees of differentiation. Most importantly, this phenomenon may be the characteristic of tumor proliferation. In addition, the researchers found that xanthine oxidase can inhibit a variety of malignant phenotypes of liver cancer cells and enhance the sensitivity of liver cancer cells to chemotherapeutic drugs.

In addition, because the spectral width at the position of peak 734 cm^{-1} is wide for cancer and benign, and the DNA/RNA base correlation bands (725 cm^{-1},743 cm^{-1}) are in this wide range, the intensity is also relatively higher than that of normal people, suggesting that patients with pulmonary nodules may be related to the increase in the relative number of nucleic acid bases relative to the whole active components of SERS. The appearance of these nucleic acid bases may indicate abnormal metabolism of DNA or RNA bases in the blood of patients with pulmonary nodules, which is due to apoptosis and necrosis or the release and lysis of intact cells in the bloodstream. In recent years, the study of circulating RNAs in human blood has attracted much attention because of its role as a new tumor-specific marker in tumor detection [36].

The peak of 1004 cm^{-1} was ascribed to phenylalanine as a biomarker for lung cancer, whose aberrant metabolism may produce content changes [22, 28, 37]. The peaks at 1335 cm^{-1} and 1582 cm^{-1} were found to be tryptophan [28], and its relative SERS intensity was higher in the benign group but decreased in the malignant group, as shown in the previous finding [38, 39]. Although the averaged spectra of the two patient groups differed little, there was still sufficient information which was worth for systematically analyzing.

Currently, surgical diagnosis is recommended for those with solid pulmonary nodules ≥ 8 mm. For the solid pulmonary nodules greater than or equal to 8 mm with a low probability of clinical malignant tumor, the result of functional imaging examination is negative (i.e., the lesion of PET examination does not have hypermetabolism). The test result shows that when the probability of a malignant tumor is very low, the lung nodule cannot be diagnosed by biopsy. As a result, 3-6 months, 9-12 months, or 18-24 months of continuous CT follow-up is recommended. For pulmonary nodules with 6-8 mm diameter, a CT scan is recommended within 6-12 months, and if there is no change, a CT scan is performed within 18-24 months. If the diameter of the

pulmonary nodule is 4-6 mm, a CT scan should be performed within 12 months. If there is no change, a subsequent scan is not required. And if the diameter of the pulmonary nodule is less than or equal to 4 mm, no follow-up is needed [40]. Although pulmonary nodules are currently mainly found by CT scan, CT scan cannot determine the benign or malignant nodules. And for nodules of size > 8 mm, expensive PET is needed to further determination, which is difficult for low-income families to bear. Grass-roots hospitals may not have such expensive and complex medical equipment, so patients can only go to higher level hospitals for further examination, making them unable to achieve patient triage. However, blood routine examination equipment is available in every primary medical institution. Raman spectrum screening of serum only needs 2-3 mL blood, which is harmless to the human body, let alone the radiation risk of CT and PET, so the detection based on Raman spectrum screening is a low-cost, easy to carry out screening technology.

It should be noted that the oversimplified peak intensity analysis described above only used limited SERS peak information. At the same time, there are significant changes and overlaps in the SERS spectral intensities of plasma between normal subjects and cancer patients. Therefore, we used multivariate statistical analysis to combine the entire spectrum and automatically determine the most diagnostically essential features to improve serum analysis and differentiation efficiency. Researchers have studied this powerful program to identify Raman spectroscopy in cancer tissues, cells, and blood [41–43]. And our results show that the accuracy of diagnosis of malignant nodules, benign nodules, and normal is 93.33%, which means that lung adenocarcinoma nodules whose pathology cannot be determined before operation can be diagnosed in advance. At present, the positive surgical rate of pulmonary nodules is roughly 80-90%, according to the different levels of different hospitals, which means that some benign patients may receive unnecessary surgery. Our findings suggest that using SERS to identify malignant nodules before an operation is expected to increase by about 10%. From the experimental results, the Raman spectra of benign and malignant nodules are relatively similar, which is also similar to the prediction from a clinical and imaging point of view. For patients with pulmonary nodules, the Raman spectrum is quite different from that of normal people, which indicates that the metabolism of patients with pulmonary nodules is different from that of normal people, which provides a feasible idea for the screening of pulmonary nodules. On the other hand, benign and malignant nodules have similar Raman spectra. This result reflects that the serum components of benign and malignant nodules have similar changes compared with normal, indicating that the samples of pulmonary nodules at different stages still contain some similar components. However, a tumor is a complex system formed in a complex human environment, and it is one-sided to look at a tumor in a single dimension. This study only explored the possibility of an auxiliary diagnosis of benign and malignant pulmonary nodules from the perspective of serum (metabolism/immunity). The AI of imaging has also done a lot of work and feasibility studies from the dimension of imaging. But this paper does not analyze the classification of benign and malignant tumors from the perspective of imaging. So in the next step of work, we will consider continuing in-depth research combining imaging and serum Raman dimensions. One more dimension of information is theoretically helpful for the reliability of auxiliary diagnosis results.

The difference between normal serum and cancer/benign is clear from the results. This shows the sensitivity of this aspect, and the serum can reflect the metabolism and immune status in vivo. Abnormal apoptosis or immunity will lead to certain changes in serum components, and changes in different dimensions reflect the difference between malignant and benign tumors.

This paper uses three groups of variables, normal, benign, and malignant, for statistical analysis, which can accurately classify patients with 93.33% and help detect patients with pulmonary nodules earlier and reduce excessive medical treatment.

5. Conclusion

In this work, we prepared gold nanorods as the surface-enhanced Raman spectroscopy detection substrate, which greatly enhanced the Raman signal of serum samples. With the help of the SVM prediction model, our prediction accuracy of malignant nodules, benign nodules, and normal groups reached 93.33%. Our results suggest that surface-enhanced Raman spectroscopy may be fast, effective, and convenient for screening pulmonary adenocarcinoma nodules. It may also provide auxiliary diagnostic information for whether pulmonary nodules should undergo surgery as a reference method for other tumors.

Data Availability

The Raman data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declares that they have no conflicts of interest.

Acknowledgments

This work was supported by the Sichuan Natural Science Foundation (no. 22NSFSC1429); the Radiation Oncology Key Laboratory of Sichuan Province Open Fund (no. 2020FSZLX-03); UESTC-Sichuan Cancer Hospital 2021 Medical-Engineering Oncology Innovation Fund (no. ZYGX2021YGCX013); and the Department of Science and Technology of Sichuan Province (grant/award number: 2019YFG0185).

References

[1] J. Ferlay, M. Colombet, I. Soerjomataram et al., "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *International Journal of Cancer*, vol. 144, no. 8, pp. 1941–1953, 2019.

[2] CONCORD Working Group, "Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries," *The Lancet*, vol. 391, no. 10125, pp. 1023–1075, 2018.

[3] F. C. Detterbeck, D. J. Boffa, A. W. Kim, and L. T. Tanoue, "The eighth edition lung cancer stage classification," *Chest*, vol. 151, no. 1, pp. 193–203, 2017.

[4] G. Zhang, S. Jiang, Z. Yang et al., "Automatic nodule detection for lung cancer in CT images: a review," *Computers in Biology and Medicine*, vol. 103, no. 287-300, pp. 287–300, 2018.

[5] G. Veronesi, "Lung cancer screening: the European perspective," *Thoracic Surgery Clinics*, vol. 25, no. 2, pp. 161–174, 2015.

[6] P. Hofman, "Liquid biopsy for early detection of lung cancer," *Current Opinion in Oncology*, vol. 29, no. 1, pp. 73–78, 2017.

[7] R. Wilson and A. Devaraj, "Radiomics of pulmonary nodules and lung cancer," *Translational Lung Cancer Research*, vol. 6, no. 1, pp. 86–91, 2017.

[8] E. Y. Chan, P. Gaur, Y. Ge et al., "Management of the solitary pulmonary nodule," *Archives of Pathology & Laboratory Medicine*, vol. 141, no. 7, pp. 927–931, 2017.

[9] N. T. Tanner, A. Porter, M. K. Gould, X. J. Li, A. Vachani, and G. A. Silvestri, "Physician assessment of pretest probability of malignancy and adherence with guidelines for pulmonary nodule evaluation," *Chest*, vol. 152, no. 2, pp. 263–270, 2017.

[10] B. Ghaderi, H. Moghbel, N. Daneshkhah, A. Babahajian, and F. Sheikhesmaeili, "Clinical evaluation of serum tumor markers in the diagnosis of gastric adenocarcinoma staging and grading," *Journal of Gastrointestinal Cancer*, vol. 50, no. 3, pp. 525–529, 2019.

[11] F. Feng, Y. Tian, G. Xu et al., "Diagnostic and prognostic value of CEA, CA19-9, AFP, and CA125 for early gastric cancer," *BMC Cancer*, vol. 17, no. 1, p. 737, 2017.

[12] E. Er, A. Sanchez-Iglesias, A. Silvestri et al., "Metal Nanoparticles/MoS2Surface-Enhanced Raman scattering-based sandwich immunoassay for α-fetoprotein detection," *ACS Applied Materials & Interfaces*, vol. 13, no. 7, pp. 8823–8831, 2021.

[13] P. Wang, Y. Sun, X. Li et al., "Recent advances in dual recognition based surface enhanced Raman scattering for pathogenic bacteria detection: a review," *Analytica Chimica Acta*, vol. 1157, no. 338279, p. 338279, 2021.

[14] J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García et al., "Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients," *Lasers in Medical Science*, vol. 22, no. 4, pp. 229–236, 2007.

[15] M. V. Chowdary, K. K. Kumar, K. Thakur et al., "Discrimination of normal and malignant mucosal tissues of the colon by Raman spectroscopy," *Photomedicine and Laser Surgery*, vol. 25, no. 4, pp. 269–274, 2007.

[16] L. Vigderman and E. R. Zubarev, "High-yield synthesis of gold nanorods with longitudinal SPR peak greater than 1200 nm using hydroquinone as a reducing agent," *Chemistry of Materials*, vol. 25, no. 8, pp. 1450–1457, 2013.

[17] X. Ning, I. W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometrics and Intelligent Laboratory Systems*, vol. 139, no. 156-167, pp. 156–167, 2014.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Leaming*, vol. 20, no. 3, pp. 273–297, 1995.

[19] C. C. Leon Bottou, J. S. Denker, H. Drucker et al., "Comparison of classifier methods: a case study in handwritten digit recognition," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing (Cat. No.94CH3440-5)*, Jerusalem, Israel, 1994.

[20] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," *Analyst*, vol. 135, no. 2, pp. 230–267, 2010.

[21] S. Guo, T. Bocklitz, U. Neugebauer, and J. Popp, "Common mistakes in cross-validating classification models," *Analytical Methods*, vol. 9, no. 30, pp. 4410–4417, 2017.

[22] C. Li, W. Chen, M. Zhang et al., "Modulatory effects of Xihuang pill on lung cancer treatment by an integrative approach," *Biomedicine & Pharmacotherapy*, vol. 130, no. 110533, p. 110533, 2020.

[23] H. Jin, Q. Lu, S. Jin et al., "Research on measurement conditions for obtaining significant, stable, and repeatable SERS signal of human blood serum," *IEEE Photonics Journal*, vol. 9, no. 2, pp. 1–10, 2017.

[24] J. F. Fahrmann, D. D. Grapov, K. Wanichthanarak et al., "Integrated metabolomics and proteomics highlight altered nicotinamide- and polyamine pathways in lung adenocarcinoma," *Carcinogenesis*, vol. 38, no. 3, pp. 271–280, 2017.

[25] K. Faulds, R. E. Littleford, D. Graham, G. Dent, and W. E. Smith, "Comparison of surface-enhanced resonance Raman scattering from unaggregated and aggregated nanoparticles," *Analytical Chemistry*, vol. 76, no. 3, pp. 592–598, 2004.

[26] M. Kahraman, O. Aydin, and M. Culha, "Size effect of 3D aggregates assembled from silver nanoparticles on surface-enhanced Raman scattering," *ChemPhysChem*, vol. 10, no. 3, pp. 537–542, 2009.

[27] Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, and H. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *International Journal of Cancer*, vol. 107, no. 6, pp. 1047–1052, 2003.

[28] A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk, and M. Baranska, "Raman spectroscopy of proteins: a review," *Journal of Raman Spectroscopy*, vol. 44, no. 8, pp. 1061–1076, 2013.

[29] A. Laviano, A. Cascino, M. Muscaritoli, F. Fanfarillo, and F. R. Fanelli, "Tumor-induced changes in host metabolism: a possible role for free tryptophan as a marker of neoplastic disease," *Developments in Tryptophan and Serotonin Metabolism*, vol. 527, no. 363-366, 2003.

[30] N. Kuhar, S. Sil, and S. Umapathy, "Potential of Raman spectroscopic techniques to study proteins," *Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy*, vol. 258, no. 119712, p. 119712, 2021.

[31] J. Yu, F. Du, L. Yang et al., "Identification of potential serum biomarkers for simultaneously classifying lung adenocarcinoma, squamous cell carcinoma and small cell carcinoma," *Cancer Biomarkers*, vol. 30, no. 3, pp. 331–342, 2021.

[32] R. Ojani, A. Alinezhad, and Z. Abedi, "A highly sensitive electrochemical sensor for simultaneous detection of uric acid, xanthine and hypoxanthine based on poly(l-methionine) modified glassy carbon electrode," *Sensors and Actuators B: Chemical*, vol. 188, no. 621-630, pp. 621–630, 2013.

[33] M. Dervisevic, E. Dervisevic, E. Çevik, and M. Şenel, "Novel electrochemical xanthine biosensor based on chitosan-polypyrrole-gold nanoparticles hybrid bio-nanocomposite

platform," *Journal of Food and Drug Analysis*, vol. 25, no. 3, pp. 510–519, 2017.

[34] N. Prajda, "Imbalance of purine metabolism in hepatomas of different growth rates as expressed in behavior of xanthine oxidase (EC 1.2.3.2)," *Cancer Research*, vol. 36, pp. 4639–4646, 1976.

[35] N. Prajda and G. Weber, "Malignant transformation-linked imbalance: decreased xanthine oxidase activity in hepatomas," *FEBS Letters*, vol. 59, no. 2, pp. 245–249, 1975.

[36] F. Banki, W. N. Yacoub, J. A. Hagen et al., "Plasma DNA is more reliable than carcinoembryonic antigen for diagnosis of recurrent esophageal cancer," *Journal of the American College of Surgeons*, vol. 207, no. 1, pp. 30–35, 2008.

[37] J. Cao, B. Balluff, M. Arts et al., "Mass spectrometry imaging of L-[ring-13C6]-labeled phenylalanine and tyrosine kinetics in non-small cell lung carcinoma," *Cancer & Metabolism*, vol. 9, no. 1, p. 26, 2021.

[38] S.-C. Chuang, A. Fanidi, P. M. Ueland et al., "Circulating biomarkers of tryptophan and the kynurenine pathway and lung cancer risk," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 23, no. 3, pp. 461–468, 2014.

[39] C. Li and H. Zhao, "Tryptophan and its metabolites in lung cancer: basic functions and clinical significance," *Frontiers in Oncology*, vol. 11, p. no, 2021.

[40] M. K. Gould, J. Donington, W. R. Lynch et al., "Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines," *Chest*, vol. 143, no. 5, pp. e93S–e120S, 2013.

[41] B. Li, Y. Wu, Z. Wang et al., "Non-invasive diagnosis of Crohn's disease based on SERS combined with PCA-SVM," *Analytical Methods*, vol. 13, no. 44, pp. 5264–5273, 2021.

[42] S. Kang, I. Kim, and P. J. Vikesland, "Discriminatory detection of ssDNA by surface-enhanced Raman spectroscopy (SERS) and tree-based support vector machine (Tr-SVM)," *Analytical Chemistry*, vol. 93, no. 27, pp. 9319–9328, 2021.

[43] W. Dawuti, X. Zheng, H. Liu et al., "Urine surface-enhanced Raman spectroscopy combined with SVM algorithm for rapid diagnosis of liver cirrhosis and hepatocellular carcinoma," *Photodiagnosis and Photodynamic Therapy*, vol. 38, no. 102811, p. 102811, 2022.

Hindawi
BioMed Research International
Volume 2024, Article ID 9867136, 1 page
https://doi.org/10.1155/2024/9867136

Retraction

Retracted: A Study on Nonvisual Effects of Natural Light Environment in a Maternity Ward of a Hospital in Cold Area

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Song, W. Jiang, and P. Cui, "A Study on Nonvisual Effects of Natural Light Environment in a Maternity Ward of a Hospital in Cold Area," *BioMed Research International*, vol. 2022, Article ID 8608892, 16 pages, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 8608892, 16 pages
https://doi.org/10.1155/2022/8608892

RETRACTED

Research Article

A Study on Nonvisual Effects of Natural Light Environment in a Maternity Ward of a Hospital in Cold Area

Haihong Song, Wei Jiang, and Peng Cui

School of Landscape Architecture, Northeast Forestry University, Harbin 150040, China

Correspondence should be addressed to Wei Jiang; jiangdoudou@nefu.edu.cn

Received 8 April 2022; Revised 17 May 2022; Accepted 27 May 2022; Published 22 June 2022

Academic Editor: Chen Li

Copyright © 2022 Haihong Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Thanks to the discovery of human eye photoreceptor cells, ipGRC and human nonvisual channels, the nonvisual effects of light have gradually come to our vision and been rationally utilized. Along with this trend, people have expanded their demand for the light environment to both visual and nonvisual needs from only visual needs. With a good natural daylighting, parturients will find their psychological pressure and physiological pain relieved, their rehabilitation rate increased, and they get rehabilitated more quickly. This study was carried out in a representative maternity ward in Harbin based on the latest research results on nonvisual effects at home and abroad. Specifically, the nonvisual effects on the natural light environment in the ward throughout the year were simulated and analyzed from the aspects of equivalent melanopic lux (EML), stimulus frequency (Stim.freq), and circadian effective area (CEA). During the study, natural light in the ward was measured on site, and the evaluation tool and workflow of nonvisual effects were created with the aid of Grasshopper modeling platform, Ladybug +Honeybee, and VB script editor. The results show that the nonvisual effects of natural light on the body of parturients gradually weaken as they further go inside the ward. What is worse, in the most unfavorable all-overcast condition, daylighting on beds far away from the window does not meet the stimulus of human circadian rhythm from April to August. Therefore, additional light is required. The wards have the best nonvisual natural light environment when they are south facing and have a window to floor ratio of 0.3.

1. Introduction

Natural light is the main light source used by people. Besides its visual functions, people find that it can also have an impact on our circadian rhythm and psychological state, although no scientific explanation demonstrated. Until the retina's third kind of photoreceptor cells—intrinsically photosensitive ganglion cells (ipRGCs)—and nonvisual channels were found in 2002 [1], a rational explanation was delivered. Follow-up studies were carried out on the mechanism of such cells, allowing further verification of the nonvisual biological effects of light stimulation on the body [2].

More and more studies have shown that light's nonvisual effects can directly affect our mood, working efficiency, and physical health [3, 4]. This leads to a transition of the evaluation standard for a light environment from a single visual effect to a double evaluation standard of visual and nonvisual. As for parturients awaiting delivery in a hospital, ideal natural light not only alleviates their tension and fear but also has significant effects on their physiological changes, shortening their rehabilitation and reducing the incidence rate of postpartum depression [5, 6]. Therefore, an optimal natural light environment with both visual and nonvisual standards acts as an important factor concerning the overall environmental comfort of a maternity ward.

During pregnancy, due to their own physiological needs and the needs of fetal growth, the BMR of parturients is 20%~30% higher than that of ordinary people [7], with flood pressure and heart rate rising and hormone disrupting. The parturients become more sensitive to environmental changes, such as the color, light, and temperature, showing mood swing and a high risk of depression. According to statistics, about 5%-10% parturients suffer from depression [8]. Parturients who have given birth need a good rest to recuperate and stay in an enclosed space for a long time. Apart from the company of their family, the space environment

is of great importance. A comfortable environment has significantly positive effect on the regulation of their own emotions. Among a variety of environmental factors constituting the ward space, light environment is regarded as one of the most crucial factors [9]. Studies have shown that nonvisual effects of light affect the secretion of melatonin and the associated circadian rhythm, as well as the body temperature, heart rate, blood pressure, and sleep quality [10–12], all of which are related to loneliness, anxiety, depression, and other emotions [13]. In the case of parturients cannot take corresponding drugs, optimal natural light has certain phototherapy effect. Because of its high safety and quick effect, it can remarkably improve the physiological and psychological conditions of parturients.

Since the "Two-child Policy" and "Three-child Policy" have been released one after another, the birth rate of newborns is gradually rising. People's demand for maternity hospitals is increasing, and higher requirements and standards are put forward for the hospital environment. As an important part of the hospital physical environment, the impact of light environment on users' physiology and psychology cannot be ignored. Moreover, natural light is of great study significance based on its strengths such as no-cost, complete and continuous spectrum, energy saving, and environmental friendliness.

Previous studies on the function of natural light within the building space environment were limited to two aspects: visual effects and efficient utilization of energy. However, in 2012, The Society of Light and Lighting pointed out that natural light's nonvisual effects should also be taken into consideration [14]. According to the findings of Boubekri et al., optimal natural light and illumination may have positive effects on people's happiness within a building space. People who were less exposed to natural light might show less and poor sleep, less physical activity, and declining living quality [15]. Adequate sleep is of vital importance to parturients. It can not only improve their physical condition but also promote calcium absorption and the development of fetal brain. Hoof et al. studied on the illuminance range that met nonvisual effects. The results showed that in the daytime, a high-level light exposure (over 1000~2000 lx) was needed for human eyes to fully meet nonvisual needs, while excessive light exposure at night should be prevented from affecting sleep quality [16]. Kessel et al. found that photoreceptor cells of different age groups varied in their sensitivity to the light of the same wave band and circadian rhythm stimulation. Such sensitivity to light declined with the increase of age [17]. Therefore, elderly parturients needed more light stimulus compared to those of the right age. In their studies, Andersen et al. took people's activity patterns and viewing directions into consideration and proposed an evaluation model combining visual and nonvisual effects, further promoting the development of studies on nonvisual effects of natural light [18]. According to the current research trend, nonvisual functions of natural light will gradually surpass visual effects and energy utilization, and the therapeutic function of natural light environment will also be rediscovered [19]. Thus, a ward with sufficient natural lighting and nonvisual effects is essential.

Based on available research background, a double obstetrics ward in Harbin Maternity Hospital is taken as the experimental object. As Harbin is located in a harsh region of China, the winters are cold, dry, and long-lasting, receiving less sunlight, which has a greater impact on maternity and needs to be enhanced to benefit from natural light resources. Therefore, the winter time period was selected, and the ward was firstly analyzed with real measurements of natural light, and then, the dynamic simulation analysis of natural daylighting was carried out with the aid of Grasshopper modeling platform and VB editor. Moreover, the results are presented in Excel. On such a basis, the nonvisual effects of the natural light environment in the ward were evaluated in a scientific and quantitative way. By doing so, this paper will provide effective data support for researches on nonvisual effects of medical space in maternity hospitals and fill the gap in research on light environment in maternity wards in China.

2. Materials and Methods

2.1. Field Measurement of Natural Light

2.1.1. Selection of Sample Wards. After field visits and online data collection, considering the representativeness of the ward and the operability of field measurement, we chose general double-occupancy maternity wards within Harbin obstetrics and gynecology hospitals as experimental subjects, which meets the requirements of one-way side daylighting in *Standard for Daylighting Design of Buildings (GB5003-2013)* (daylight factor of side lighting shall not be less than 3% and indoor natural light illuminance shall not be less than 450 lx). Finally, two vacant wards on the 9^{th} floor of the inpatient department of Harbin Red Cross Hospital were selected as sample wards (Figure 1(a)). These two wards had basically the same parameters, such as indoor arrangement and spatial layout, with built-in toilets and no shelter outside. The difference was that one ward faces south, while the other faces north (Figure 1(b)).

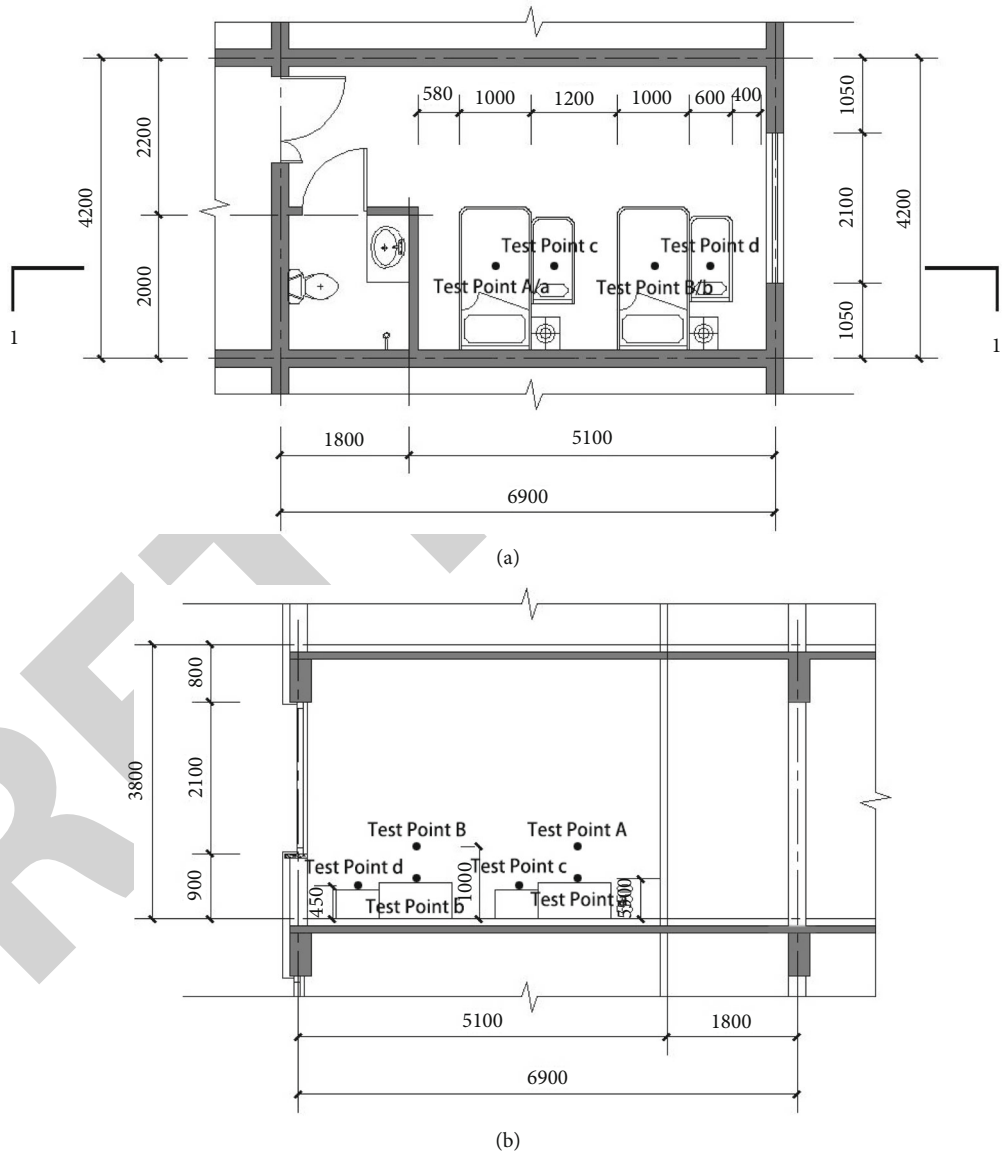
2.1.2. Layout of Measuring Points. One of the biggest differences between a maternity ward and a general ward is that the former is also equipped with a crib, apart from general beds for parturients to get rest. Due to the low ability of autonomic nerve and self-adjustment, newborns are quite vulnerable to the strong light. This may affect the physiological conditions of newborns such as mood fluctuation, sleep duration, and heart rate [20]. Compared to adults, photoreceptor cells in the eyes of newborns are more sensitive to the light stimulus [17]. Despite their sleep duration of 16~17 h a day, the nonvisual effects of natural light on their bodies cannot be ignored. Hence, apart from the ordinary hospital bed, it is also necessary to set up measuring points on the surface of the crib for illumination measurement.

For in-depth study of the nonvisual effects of natural light in a maternity ward, the behavior observation method was adopted to observe the daily activities of parturients. As such, their daily activity routines were obtained. According to the measured statistics, parturients showed three

RETRACTED

(a) (b)

Figure 1: Sample ward (a) site location of the hospital in which the ward was located and (b) on-site photo of the ward.



(a)

(b)

Figure 2: Layout of measuring points inside a ward (a) plane view and (b) sectional view.

Table 1: Material parameters.

Parameter	Wall	Ground	Ceiling	Bed surface	Glass
Material	White latex paint	Yellow PVC floor	Aluminum plate	White sheets	Ordinary double white glass
Reflection/transmission coefficient	0.7	0.4	0.7	0.6	0.8

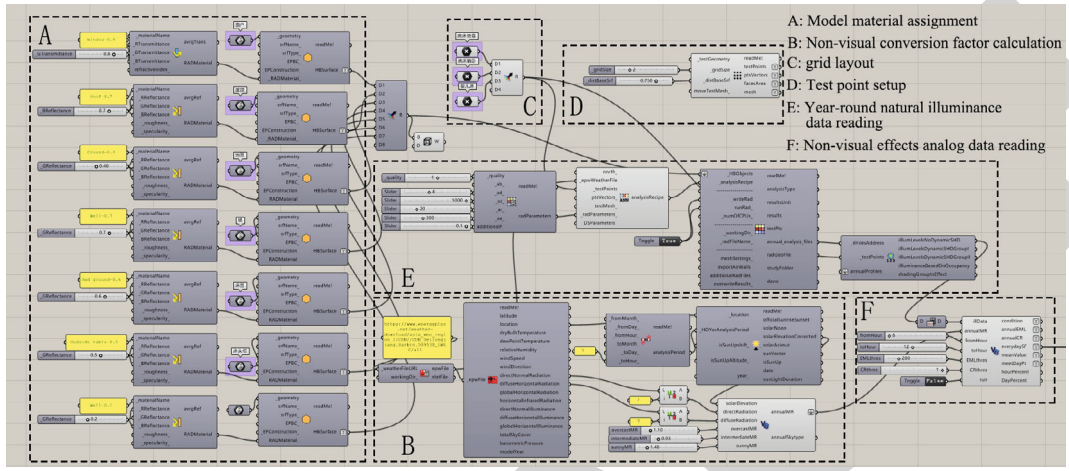


Figure 3: Simulation and analysis process of nonvisual effects of natural light.

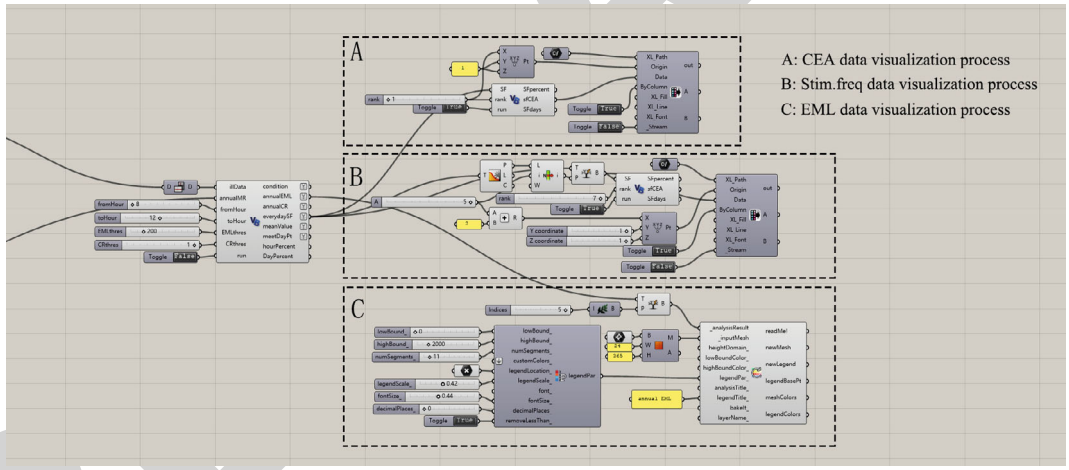
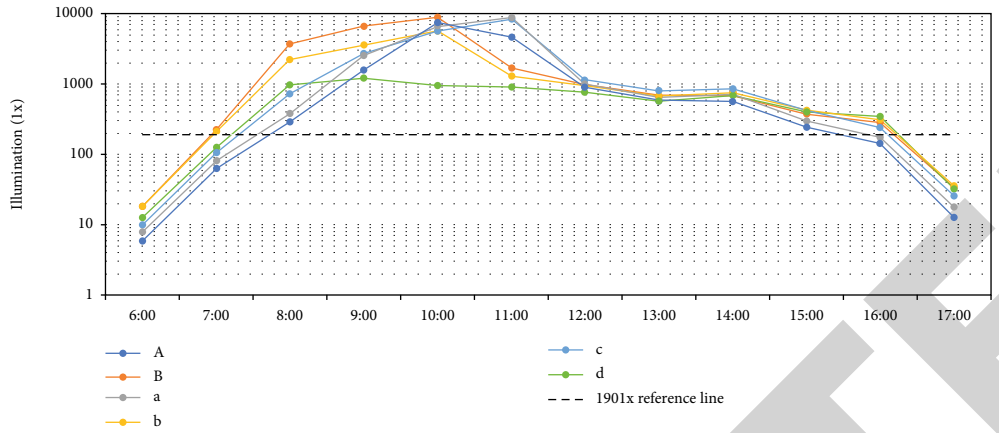


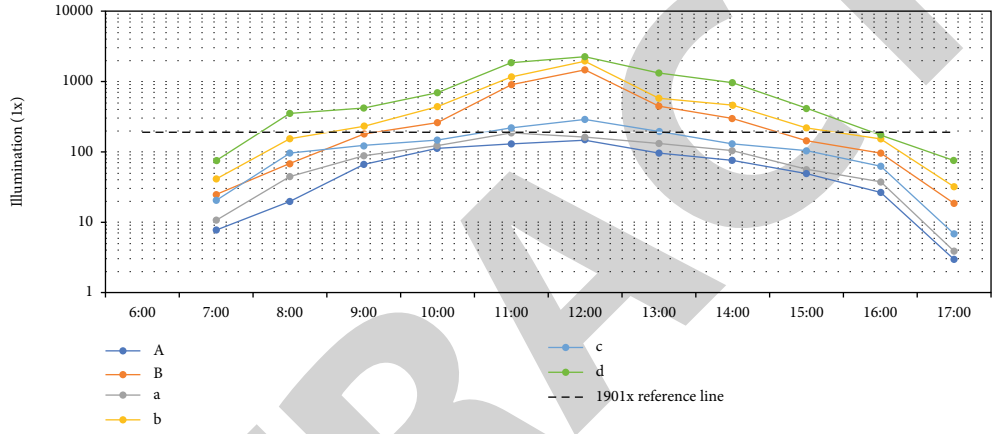
Figure 4: Visualization process of analytical data.

kinds of daily activities, namely, lying supine, leaning, and walking. Specifically, they lie supine for about 75% of their time, lean against the bed, and walk around for about 10% of their time, respectively, and do other things for about 5% of their time. According to parturients' activity routines and demand for light environment, the measuring points of nonvisual effects were arranged at different heights in the middle of the surface of ward beds and baby's cots. The measuring points A and B denote the leaning state, 1,000 mm away from the ground; a and b denote lying supine, 550 mm away from the ground; and c and d denote the selected measuring points on the crib, 450 mm away from the ground (Figure 2).

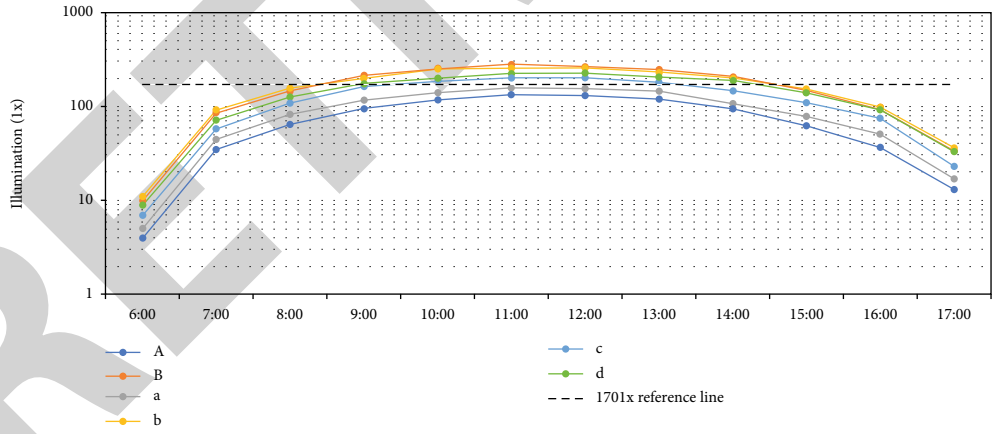
2.1.3. Field Measurement. In the study, we adopted TES1332A illuminometer (resolution: 0.1 lx; measuring range: $0.1 \sim 2 \times 10^5$ lx; margin of error: ±3%) to measure the illuminance at each measuring point. SNDWAYSW-M50 handheld infrared laser rangefinder (measuring range: 0-50 m; accuracy: ±1.5 mm) was used to measure window size, room depth, and bay net value. At the same time, a tape measure was also used for auxiliary measurement. The ward was measured on January 9-10 (two sunny days) and 12-13 (two completely cloudy days) in 2021. During measurement, the curtains were pulled back, and all indoor lights were turned off, reducing the influence of artificial light on measurement. According to the timetable of sunrise and sunset



(a)



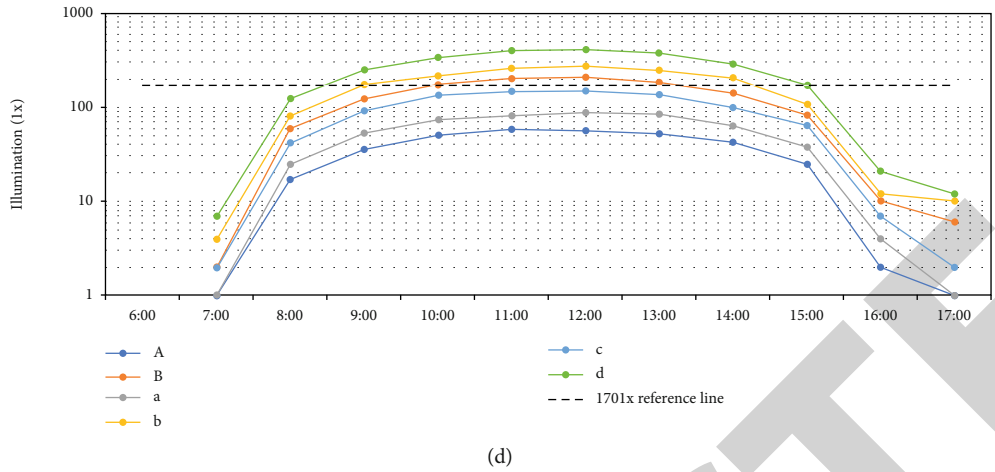
(b)



(c)

Figure 5: Continued.

RETRACTED



(d)

Figure 5: Illuminance values at each measuring point in the wards facing different directions under different weather conditions: (a) facing south, sunny; (b) facing south, cloudy all day; (c) facing north, sunny; and (d) facing north, cloudy all day.

in Harbin, the illuminance of the measuring points in the two wards is measured from 6 : 00 to 17 : 00 every day. The measurement was performed once every hour for three times at an interval of 5 s. Take the mean value of the three measurements as the final result.

Different from the previous evaluation of static daylighting and illumination, the evaluation of natural daylighting with nonvisual effects is badly sensitive to external factors. It seems that the illumination intensity and spectral distribution of natural light are invariably in a fluctuating state, but there are also certain inherent laws to follow [21]. In order to make the evaluation more accurate, natural light can be simplified to some extent. For example, D65, D75, D55, and other CIE standard light sources can substitute the sky in the south, north, northeast, and other directions, respectively. On account of the research conducted by Cajochen et al. on illumination intensity and human body vigilance [22], it is summarized that the starting threshold of natural light illuminance to stimulate the circadian rhythm is as follows. Among CIE standard light sources, D65 is 190 lux, D55 21 0lux, and D75 170 lux. Therefore, it can be concluded that the natural daylighting illuminance value of a room facing north, which can receive natural light from the north, reaches 170 lux, that of a room facing east or west reaches 210 lux, and that of a room facing south reaches 190 lux. Although there is no specific requirement for the daily lighting time that should meet the nonvisual effects of the human body, the WELL Building Standard defines that the minimum light stimuli standard for nonvisual effects is 250EML and lasts not less than 4 hours per day. As a result, in the actual measurement of this paper, the natural light of not less than 190 lux for at least 4 hours every day is regarded as the standard, on the basis of which the human body is believed to have an effective circadian rhythm on this day.

2.2. Computer Simulation

2.2.1. Simulation Technology Platform. Grasshopper is the most commonly used parametric design and visual programming platform for architects thanks to its great openness and efficiency and compatibility with many plug-ins. At present, DIVA and Ladybug+Honeybee, two mature daylighting computing plug-ins, are running on Grasshopper platform. Roudsari et al. found that the combination of Honeybee and Ladybug works better than DIVA in meteorological data analysis and visualization and also natural daylighting simulation. For this reason, Ladybug+Honeybee is selected for related daylighting calculation in the development of nonvisual evaluation tools.

Honeybee, mainly developed by Mostapha Sadeghipour Roudsari of the University of Pennsylvania, is a powerful port plug-in in Grasshopper toolkit for daylighting simulation and energy consumption analysis. With Honeybee, daylighting was mainly analyzed in line with the following procedures: setting weather conditions, importing models, assigning materials, setting simulation parameters, making simulated calculation, and visualizing results. In terms of the simulation type, users can analyze illuminance, brightness, shade, glare, etc. and then use Run daylight simulation, a module, to call Radiance or Daysim for daylighting calculation.

2.2.2. Evaluation Method. At present, there are three kinds of simulation evaluation methods for nonvisual effects of dynamic natural light in the world, namely, weighted DA value evaluation, cap model, and circadian effective area (CEA) evaluation. Among them, CEA evaluation takes into account the influencing factor of personal photosensitive history for the first time. It is convenient to calculate and has stronger graphical expression ability of the analysis results. Therefore, CEA evaluation is mainly adopted here for simulation analysis and evaluation of nonvisual effects of natural light.

2.2.3. Workflow Establishment of Simulation Evaluation. The simulation analysis of nonvisual effects mainly includes daylighting simulated calculation and visualization analysis. With the aid of plug-in Honeybee, the annual dynamic natural lighting is simulated, and data are processed with the

RETRACTED

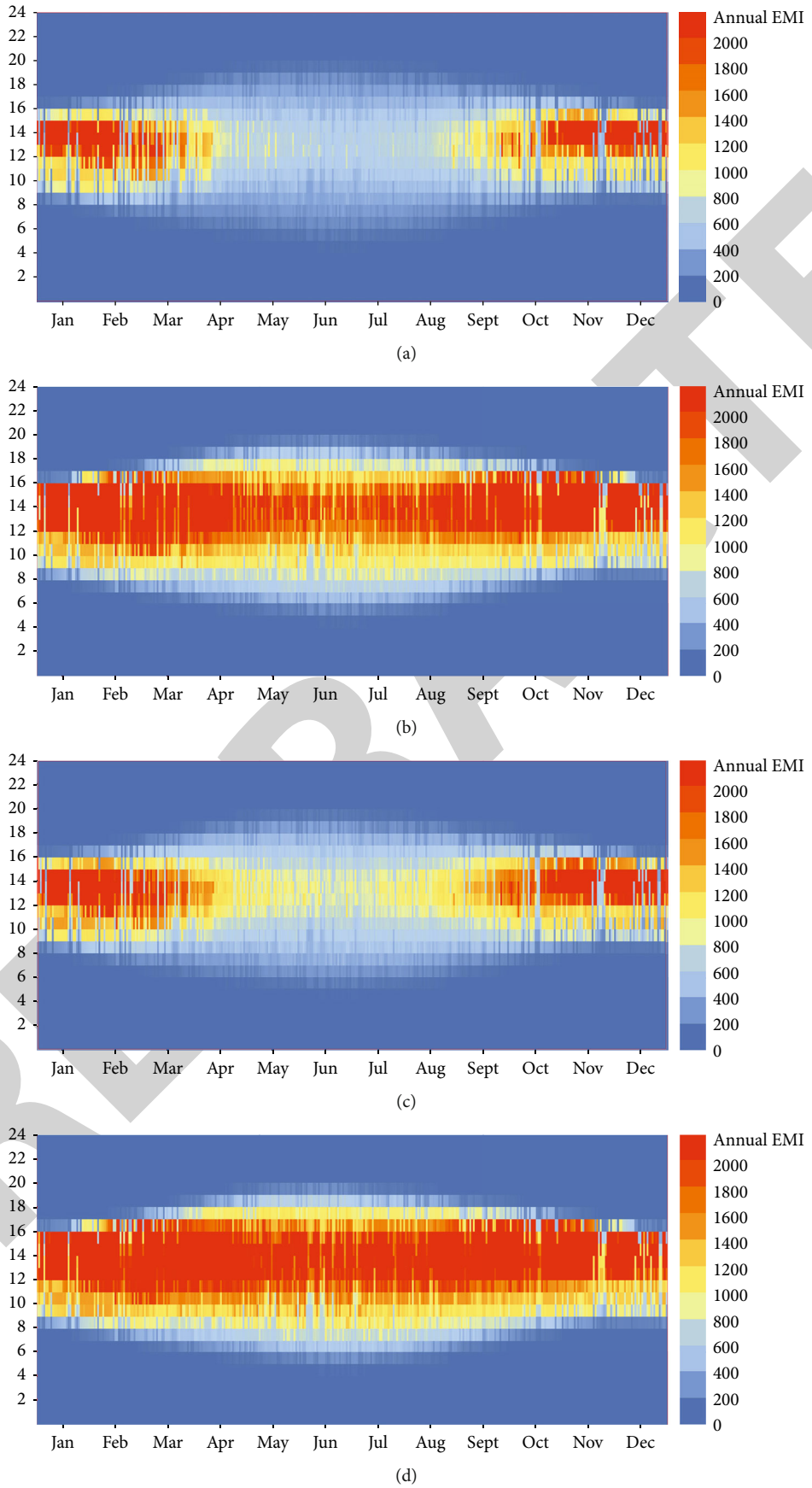


Figure 6: Continued.

RETRACTED

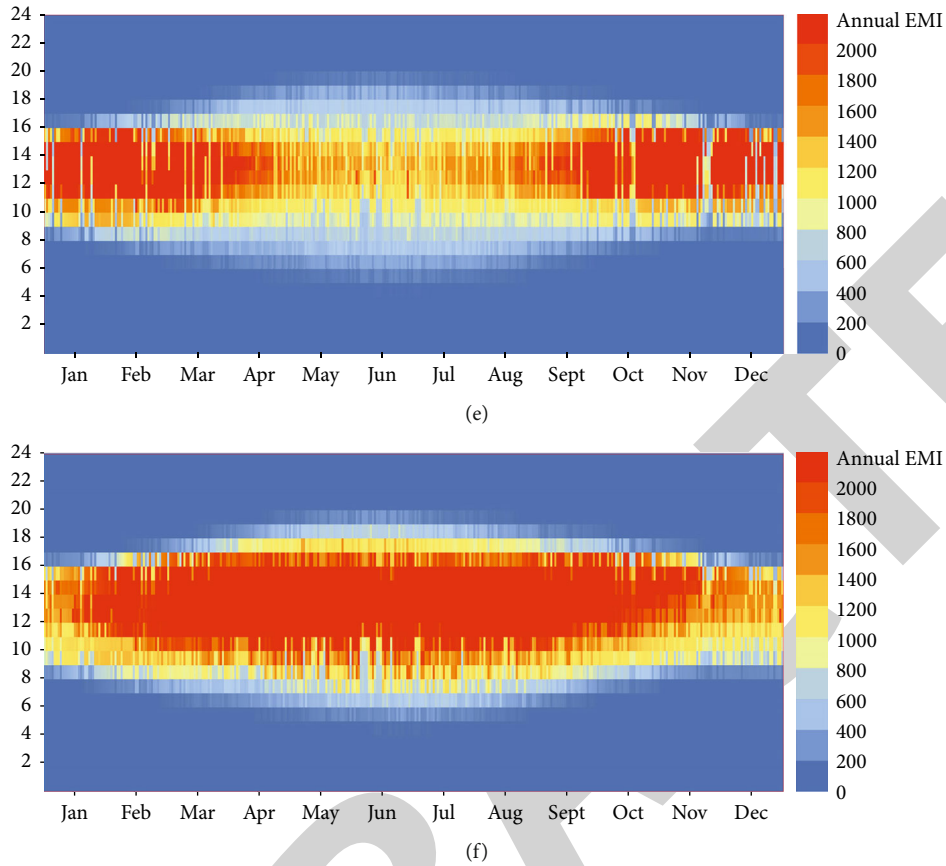


Figure 6: Year-round temporal mapping of EML at each measuring point: (a) measuring point A; (b) measuring point B; (c) measuring point a; (d) measuring point b; (e) measuring point c; and (f) measuring point d.

evaluation tool component produced by the VB editor. Then, the results can be visualized with the combination of Grasshopper's built-in components and Excel. According to the analysis of the measured data, it is found that the illuminance value of the ward facing the north is less sensitive to the weather change, and the illuminance value is low. Therefore, the ward facing the south is taken as the sample ward for computer simulation analysis.

(1) Daylighting Simulated Calculation. Based on the wards' measurement data, a 3D model was built in Rhino. The model was converted to the Honeybee Zone that can be identified by Honeybee, and then it was endowed with materials. The color, roughness, highlights, and other factors of materials will exert an impact on the indoor natural daylighting. Honeybee has a calling component of the Radiance material library, through which plastic and glass materials in Radiance are directly available. The reflectivity and transmittance of the indoor materials used in this paper are acquired with field measurements (Table 1).

Studies have shown that the human body can respond differently when stimulated by natural light at different times of a day, among which the light stimuli from 6 : 00 to 10 : 00 can affect the circadian rhythm at most and effectively advance the circadian rhythm cycle for most of the population [23]. Nearly 75% of the population, whose circadian rhythm cycle exceeds 24 hours, needs to advance their rhythm cycle every day to better adapt to the light and dark cycle in the natural environment. However, the natural light stimulation during the hours of 10 : 00 - 18 : 00 cannot change the circadian rhythm cycle but improve the arousal level of the human body [22]. The nonvisual illuminance requirement of EML_250 for 4 hours every day is proposed in the latest WELL Building Standard 2018, and it is stipulated that the requirement should be met from 9 : 00 to 13 : 00 every day to improve the circadian rhythm of users. However, considering the ward round beginning at 8 : 00 in the hospital and the rest of parturients, the daily analysis period was set at 8 : 00 - 12 : 00. Equivalent melanopic lux (EML) thresholds are at 250.

In this study, Annual Daylight Simulation was adopted, where Daysim was called to calculate the annual dynamic natural daylighting and to convert the visual illuminance into equivalent melanopic lux through setting conversion coefficient. By setting the time range and EML threshold for daily calculation, the whole-year hourly EML as well as the whole-year daily stimulus frequency (integer between 0 and 7) meeting EML threshold requirements can be obtained. It can be used as the basis for the follow-up analysis of year-round nonvisual effects in these wards. See Figure 3 for a detailed flow chart of the simulation process.

RETRACTED

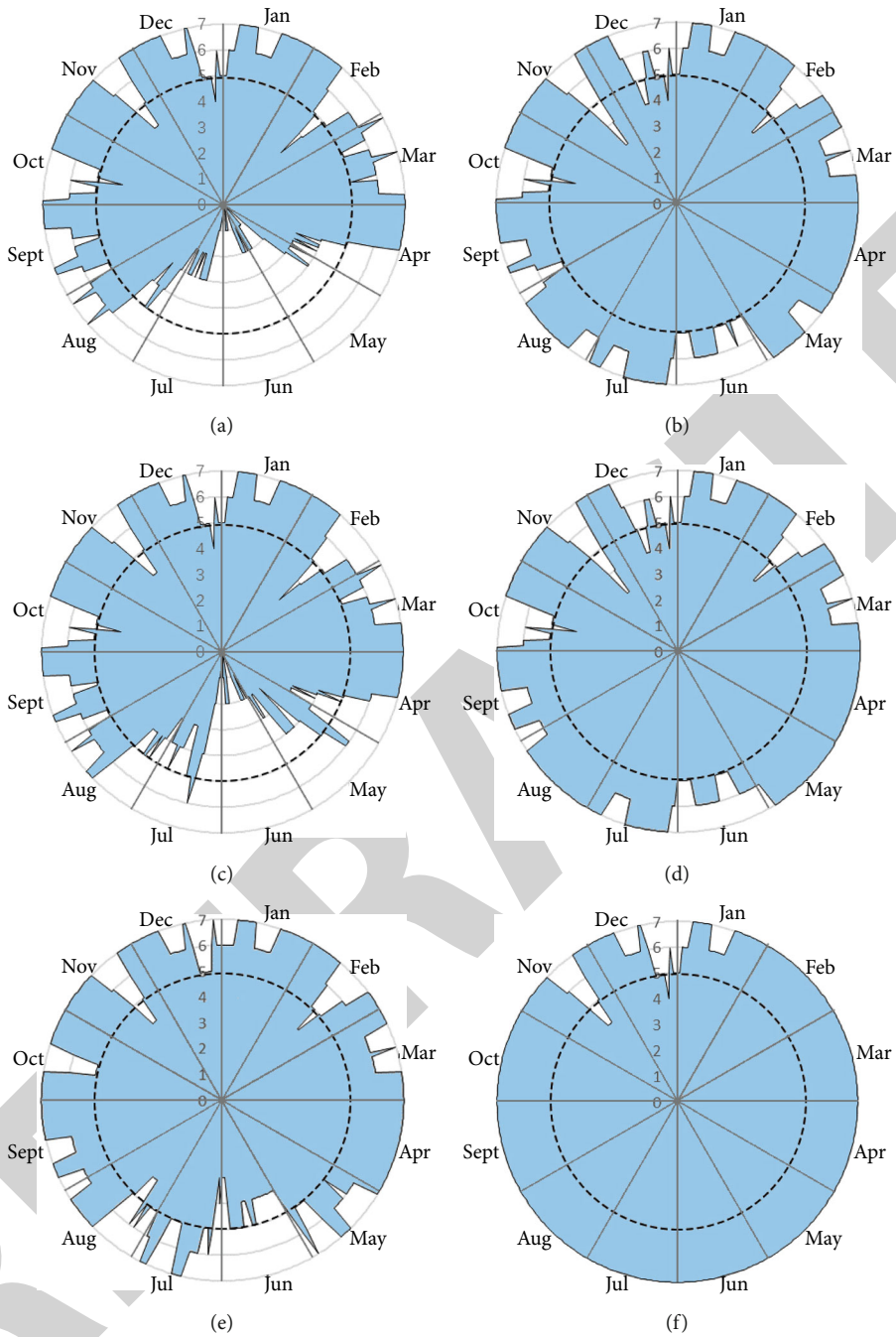


Figure 7: Year-round distribution of Stim.freq at each measuring point (8 : 00-12 : 00): (a) measuring point A; (b) measuring point B; (c) measuring point a; (d) measuring point b; (e) measuring point c; and (f) measuring point d.

(2) Visualization Analysis. In this paper, some evaluation components were compiled based on VB and combined with Grasshopper's built-in components for data processing. Then, Excel was used to present data, including the annual equivalent melanopic lux (EML), stimulus frequency (Stim.freq), and percentage of circadian effective area (CEA). The specific simulation process is shown in Figure 4.

The analysis chart of whole-year nonvisual illuminance can directly show the nonvisual illumination level of 8760 hours throughout the year, which can be taken as a basis for other analysis indicators. If the EML average of natural light received by a certain point in a certain direction reached 250EML, i.e., the minimal light stimulus standard for nonvisual effects, we deemed that circadian effective (CE) was generated at this measuring point this day. Moreover, in order to quantify the changes of availability of CE stimulus over a period of time (e.g., one week), the concept of stimulus frequency (Stim.freq for short) was proposed [24]. According to the number of days reaching effective CE on the current day and the following 6 days, the effect of CE maintenance on the current day was determined and

RETRACTED

Table 2: Year-round time percentage of Stim.freq at each level (8 : 00-12 : 00).

	Measuring point A	Measuring point B	Measuring point a	Measuring point b	Measuring point c	Measuring point d
0 d/wk	12%	0%	4%	0%	0%	0%
1 d/wk	9%	0%	3%	0%	0%	0%
2 d/wk	6%	0%	5%	0%	0%	0%
3 d/wk	5%	1%	6%	1%	2%	0%
4 d/wk	8%	4%	13%	8%	8%	1%
5 d/wk	11%	14%	14%	16%	16%	4%
6 d/wk	20%	25%	23%	24%	23%	6%
7 d/wk	29%	56%	32%	51%	51%	89%
≥5 d/wk	60%	95%	69%	91%	90%	99%

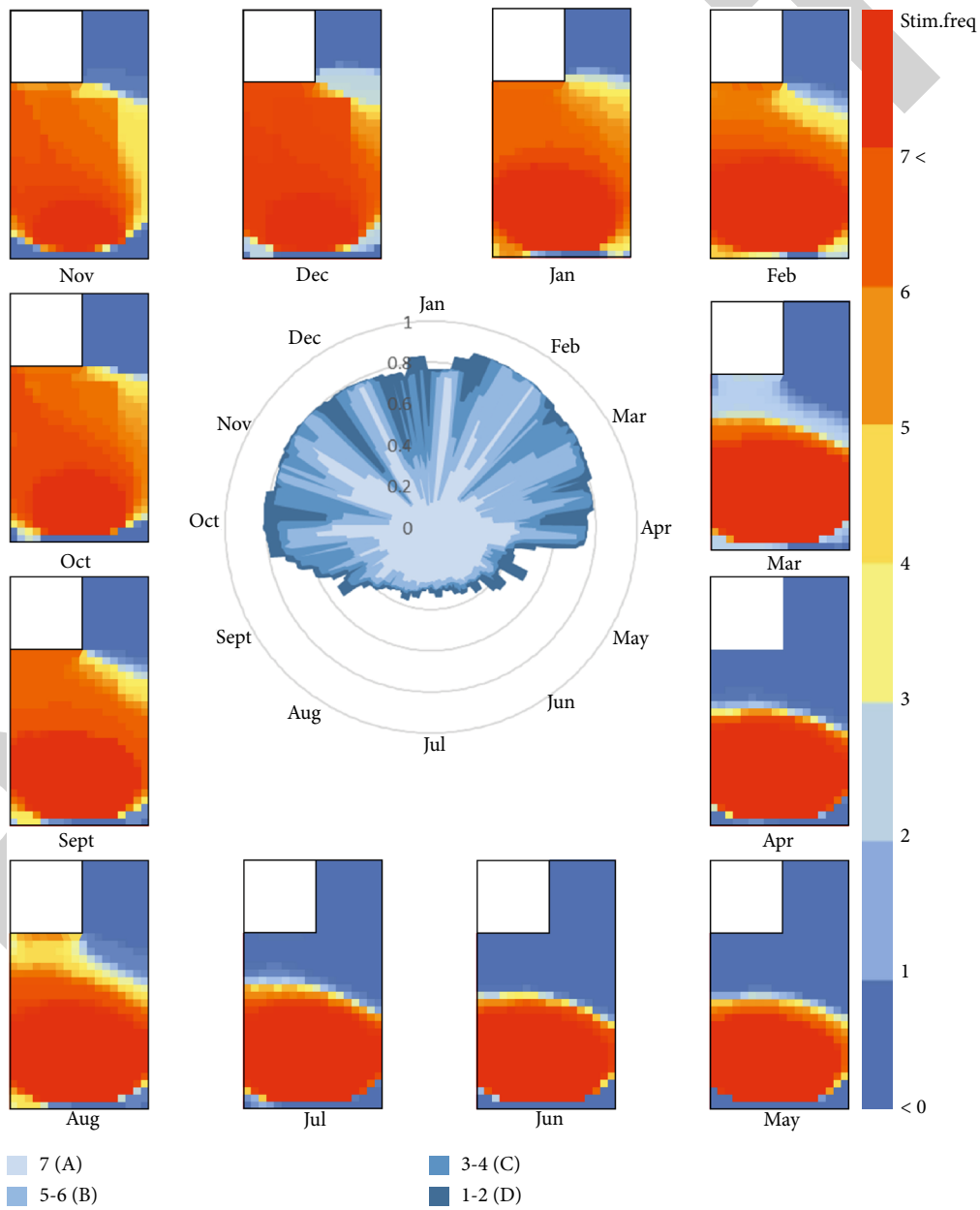


Figure 8: Spatially mapped distribution of Stim.freq of each measuring point in each month in the ward and superimposed expressions of annual CEA maps based on different Stim.freq.

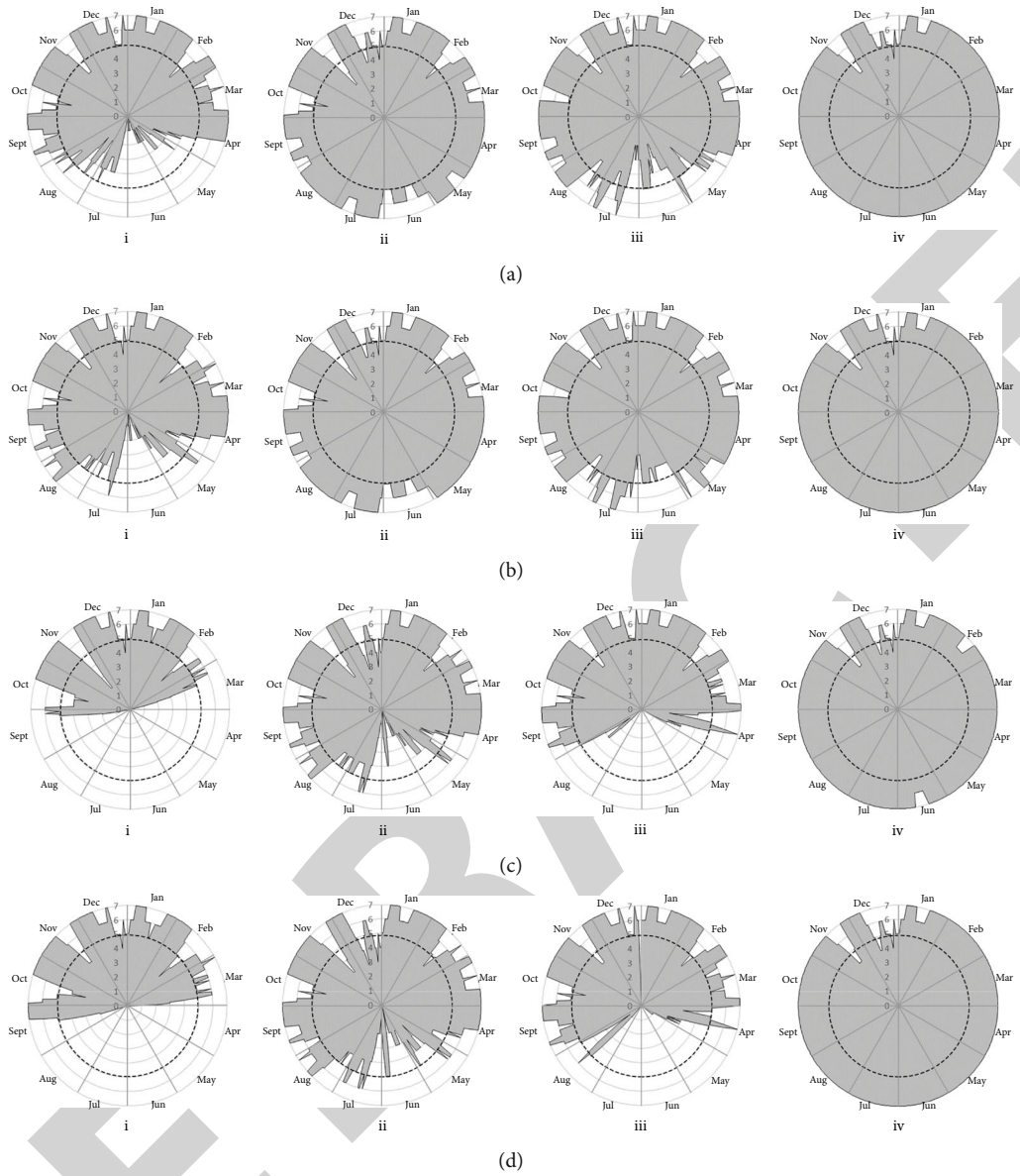


Figure 9: Whole-year Stim.freq distribution at each measuring point in the wards facing different directions: (a) facing east (i) measuring point a, (ii) measuring point b, (iii) measuring point c, and (iv) measuring point d; (b) facing south; (c) facing west; and (d) facing north.

divided into 5 levels (A: 7 days; B: 5-6 days; C: 3-4 days; D: 1-2 days; and F: 0 days). The higher the level, the better. The Stim.freq analysis chart mainly presented daily stimulus level at certain points throughout the year, only expressed in time dimension rather than spatial dimension. According to the WELL standard, the EML threshold of 200 and 150, respectively, required 75% and 100% working areas to meet the conditions. Hence, an indicator should be introduced to manifest the spatial condition during evaluation of nonvisual effects. In CEA evaluation, a CEA index has been put forward to represent the percentage of areas that met specific Stim.freq within the space. The CEA analysis chart clearly shows the percentage of measuring points to be calculated which met or exceeded the set Stim.freq within the space.

3. Results

3.1. Measurement and Analysis of Natural Light. Figure 5 shows the mean value of all day illuminance of each measuring point in both wards, respectively, facing south and north in sunny and cloudy days. In terms of the ward facing south, by comparing the 190 lx reference line, it can be seen that the natural light stimulus of each measuring point can reach the nonvisual threshold during 8 : 00-15 : 00 in sunny days (Figure 5(a)). Among them, an optimal illuminance value was determined at the bed near the window (test points B and b), which can meet the daily nonvisual needs of 10 hours. On the condition of cloudy day (Figure 5(b)), the average illuminance value at each measuring point in a ward

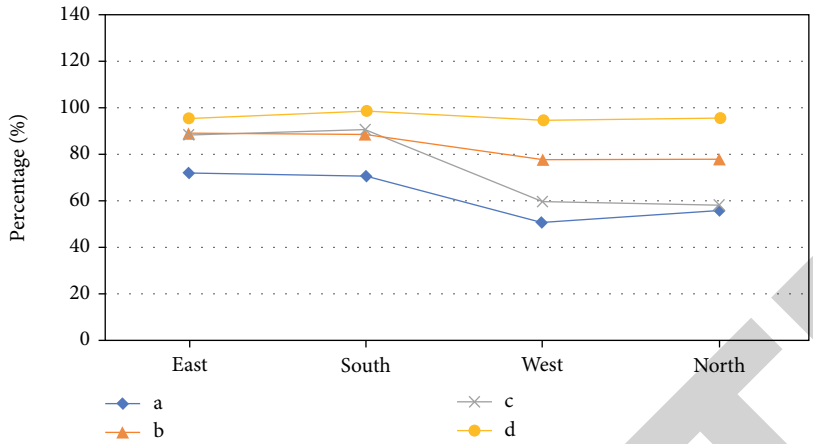


Figure 10: Percentage of Stim.freq ≥ 5 d/w at each measuring point in the wards facing different directions.

was lower than that of sunny day as a whole. The average illuminance value of measuring points close to the window was better than that far away from the window, basically reaching the illumination standard of more than 190 lux for 4 h a day, satisfying the demands of parturients' nonvisual effects. The illuminance values measured at those beds far away from the window (test point c) reached the threshold of nonvisual effects only 2 h daily, while those measured deeper from the room (test points A and a) failed to exceed 190 lx for an hour daily. Therefore, it can be considered that far away from the window (test points A, a, and C), light stimulus failed to reach the stimulus threshold of nonvisual effects on cloudy day, which is not conducive to maintaining the physical health of parturients. In terms of the ward facing north, there was no significant difference in the average illuminance value at each measuring point in the ward in sunny days (Figure 5(c)). Except for the two time periods of 6 : 00-7 : 00 a.m. and 16 : 00-17 : 00 p.m., the illuminance change range of each measuring point in time ranges is small. Compared with 170 lx reference line, it is clear that only the illumination values measured near the window (test points B, b, c, and d) between 10 : 00 and 13 : 00 meet the standard but did not exceed the standard by much. Hence, the ward was deemed with less ideal natural light environment and could not meet nonvisual needs of human body. On the condition of cloudy day (Figure 5(d)), it can be found that with the increase of room depth, the illuminance value of the measuring point decreased gradually, and the range was stable. Compared with the 170 lx reference line, in the whole ward, only two test points (test points b and d) near the window met the 4 h illuminance needs daily merely on the condition that parturients were lying. Basically, it cannot meet the nonvisual needs of parturients, which can be considered to be harmful to health.

3.2. Simulation Analysis

3.2.1. Equivalent Melanopic Lux throughout the Year. Figure 6 shows the annual time map of the measuring points representing the leaning and lying supine status of parturients as well as those on the surface of baby's cots. As illustrated, the all-year EML level of the bed (test points B, b, and d) near the window is quite high. Irrespective of a cot or a ward bed and the status of parturient, the natural daylighting of these three measuring points can meet the needs of nonvisual effects of the human body; that of the cot (test point c) in the middle can basically meet the nonvisual needs. In contrast, parturients, either lying supine or leaning against the bed, will expect variable and low equivalent EML levels at the measuring points in ward beds (test point A and a) away from the window and especially a lower level from April to August than that in other months. Therefore, it is hard to meet their physiological rhythm stimuli.

3.2.2. Stimulus Frequency. In order to sustain a healthy circadian rhythm cycle, Konis suggests that at least 5 days of effective physiological rhythmic stimulus within any 7 days are required to avoid upsetting the internal clock of human body [25]. Therefore, a threshold of Stim.freq ≥ 5 d/wk was selected here as the standard of Stim.freq analysis.

Figure 7 shows the year-round distribution of Stim.freq at each measuring point, whereas Table 2 shows the year-round time percentage of Stim.freq at each level in each measuring point. As illustrated below, from 8 : 00 to 12 : 00, the measuring points B, b, c, and d near the window basically sustain their respective critical value Stim.freq ≥ 5 d/wk during the whole year, and the percentage of time in which the actual value exceeds the critical value is over 90% throughout the year. In particular, the measuring point d sustains a stimulation frequency Stim.freq ≥ 7 d/wk per day throughout the year, except winter (December, the following January, and February). The measuring point c in the middle baby's cot hits the critical value at all times except for June. So, it follows that natural daylighting basically satisfies the nonvisual effects. Both measuring points A and a are far from the window. Their Stim.freq can reach the acceptable critical value about 70% of the time throughout the year. Unsatisfactorily, from April to August, the natural daylighting of nonvisual effects is quite poor and detrimental to the health of parturients. Comparatively, irrespective of being close to or away from the window, parturients who lie supine receive much more nonvisual natural light stimulus

RETRACTED

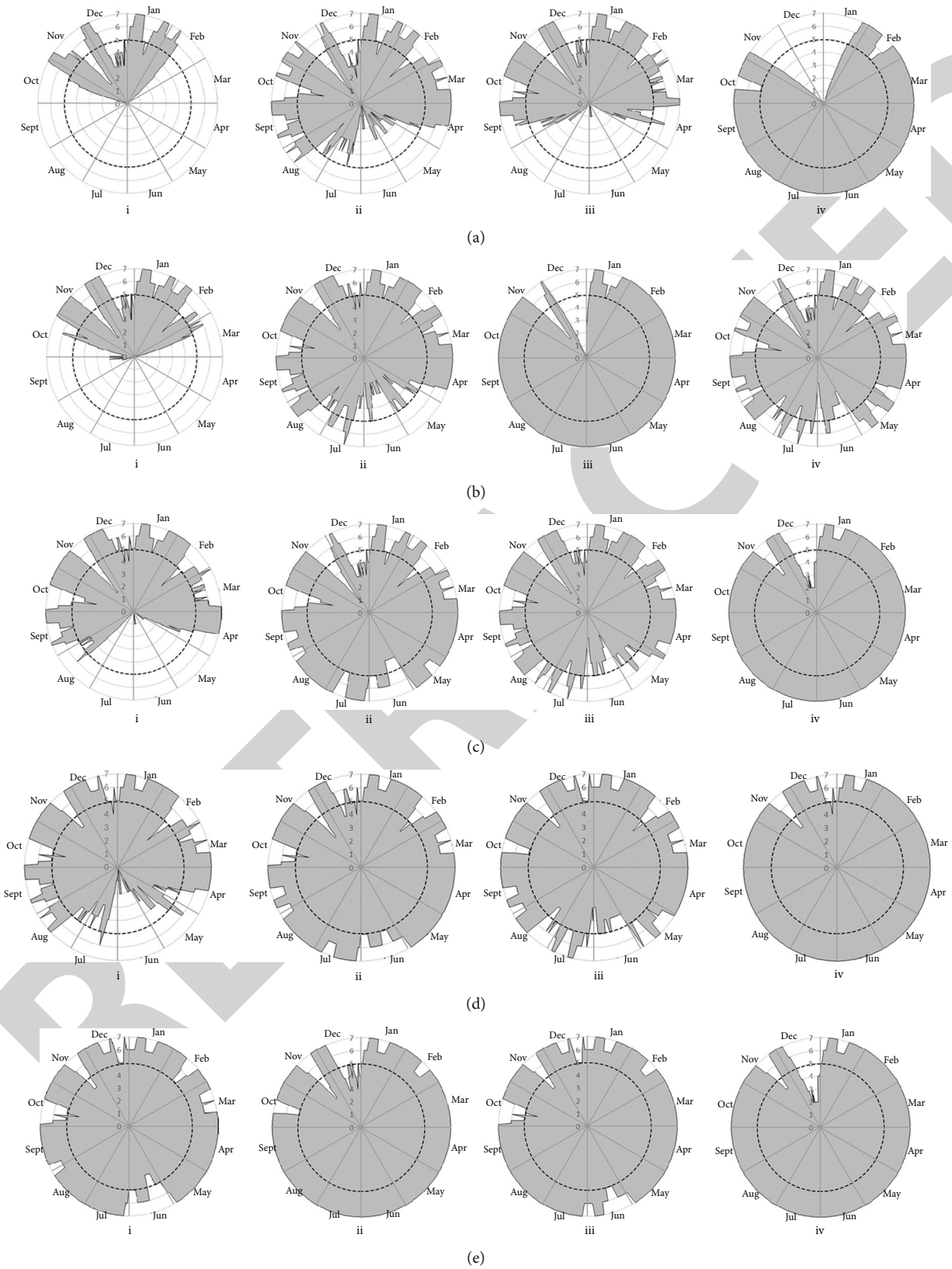


Figure 11: Whole-year Stim.freq distribution at each measuring point in the wards with different window-to-floor ratios. (a) 0.1 (i) measuring point a, (ii) measuring point b, (iii) measuring point c, and (iv) measuring point d; (b) 0.15; (c) 0.2; (d) 0.25; (e) 0.3, and (f) 0.35.

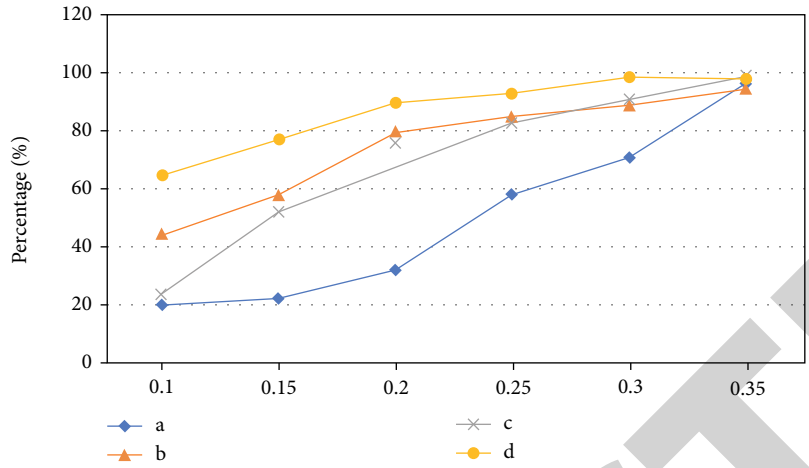


Figure 12: Percentage of Stim.freq ≥ 5 d/w at each measuring point in the wards with different window-to-floor ratios.

than those who lean against the bed. It is also evident that natural light in winter has little impact on nonvisual effects, and the distance from the window and parturients' behavior condition can only produce slight changes in the nonvisual natural light stimulus received by the body.

3.2.3. Effective Circadian Rhythm Percentage. In addition to lying supine and leaning against the bed, parturients tend to walk around. For the purpose of evaluating the nonvisual effects of natural light under such a state, it is necessary to evaluate the nonvisual natural light environment in a whole ward. With an available circadian effective area, the dynamic natural light in a building space can be evaluated. In this study, a factor of personal sensitization history is introduced for the first time. With this factor, the nonvisual effects of natural light in a space throughout the year can be analyzed. Hence, a grid at a certain distance from the ground is simulated and calculated in this daylighting simulation. In the nonvisual simulation, the WELL stipulates that the position of 1,500 mm from the ground denotes the height of human eyes when standing, typically the horizontal viewing direction. There are 522 grid measuring 200 mm × 200 mm.

As shown in Figure 8, the surrounding images show the distribution of the average CEA values of the month and the mapping of the Stim.freq of each measuring point of the grid in space. The image in the middle shows the CEA maps under each Stim.freq after categorization and merging based on the five levels (F-rank not shown). The effect of seasonal changes on the distribution of natural daylighting of nonvisual effects in the building is evident. Among others, the nonvisual stimulus is suboptimal from April to August, when the CEA above the critical value approaches 60% at the best, but only for a very few days. Overall, except for the four months from April to August, the Stim.fre in other months is 5 d/wk or above, a value quite satisfactory. The daylighting of nonvisual effects in the ward decreases as parturients go inside, especially in areas near the toilet and doorway. All these areas have quite low stimulus throughout the year, indicating that they are biologically dark and need to be supplemented with additional lighting.

4. Discussion

Based on the measurement data, in terms of the ward facing south, the standard was to meet the natural lighting of no less than 190 lx for at least 4 h every day. In sunny days, the light stimulus at each measuring point in the ward could meet parturients' needs for circadian rhythm stimulation. While in all cloudy days, the natural lighting environment near the window was not ideal. In terms of the ward facing north, the standard was natural lighting of no less than 170 lx for at least 4 h every day. Whether cloudy or sunny, natural daylighting in the ward failed to meet parturients' needs for circadian rhythm stimulation. It can thus be seen that although natural light can meet parturients' visual needs in the ward, it still fails to meet their nonvisual, requiring supplementary illumination. In the *Code for Design of General Hospital* (*GB51039-2014*), the orientation of the ward is not clearly specified. It only requires favorable daylighting, ventilation, and matching with terrain. But in fact, depending on the local climate and direction, the houses in some places are livable even they face east or west. Through software simulation, the nonvisual effects of different orientations of sample wards can be optimized and analyzed, to explore the orientation of hospital wards suitable for cold areas (Harbin).

It can be seen from Figures 9 and 10 and that the change of orientation has different effects on different measuring points. The farther away the hospital bed is from the window, the more sensitive the natural light nonvisual effect is to the change of orientation. Hence, when selecting a direction the ward faces, priority should be given to the natural daylighting having nonvisual effects near the entrance. In the ward facing east or south, percentage of test points exceeding the threshold (Stim.freq ≥ 5 d/w) exceeded 70%. In the ward facing west or north, only those measuring points near the entrance had poor performance of natural light's nonvisual effects, and percentage exceeding the threshold (Stim.freq ≥ 5 d/w) was about 50%. Therefore, in terms of natural light's nonvisual effects on parturients, the east and south direction is optimal and less different from

RETRACTED

each other, the west direction is the worst, and the north direction is on the middle level.

Through the aforesaid computer simulation, it can be concluded that the natural daylighting having nonvisual effects is problematic at beds away from the window. Especially from April to August, natural light basically cannot meet parturients' needs of circadian rhythm stimulation. This is because the ward is deep inside the war and the solar altitude angle is high from April to August, so the sunlight cannot directly reach the test points. The area near the entrance requires a supplement of high-intensity artificial illumination. Apart from the supplementary artificial illumination, the ward can also be optimized. However, it is quite complicated and demands a huge amount of work to implement optimization methods, such as changes of the ward's plane layout and proportion, or the combination between wards. In comparison, it is more simple and direct to improve the ward's daylighting by changing the ward's window-to-floor ratio.

According to Figures 11 and 12, the change of window-to-floor ratio had different impacts on the nonvisual light environment at different locations in the ward. It has the greatest impact on test point a near the entrance and the smallest impact on measuring point a near the window. That is, with the increase of room depth, the impact of the change of window-to-floor ratio on the stimulus frequency of test point increases. When the window-to-floor ratio reached 0.35, percentages of Stim.freq $\geq$ 5 d/w at each measuring point were basically the same, and the ratio rising above 0.35 had the same impacts on each test point. When the ratio was 0.3, Stim.freq values at each measuring point all maintained at a higher level throughout the year, and percentages of Stim.freq $\geq$ 5 d/w all exceeded 70% at each point. Compared to that of 0.35, the ratio of 0.3 is more energy-saving and environmentally friendly, so 0.3 was a more appropriate standard value out of these 6 ratios.

5. Conclusions

In this paper, we adopted WELL standard and CEA evaluation method that were used to quantify nonvisual effects of natural light in wards. According to the study findings, a ward's natural light environment that meets visual needs does not necessarily meet parturients' nonvisual needs. A ward's nonvisual natural light environment has something to do with parturients' behaviors and ward orientation. From the obtained results, the following conclusions can be drawn:

(1) Computer simulation has basically the same results with those of the field measurement—natural light's nonvisual effects on people gradually decrease with the increase of depth perpendicular to the ward's lighting surface width. Especially during April-September or on the whole-day cloudy condition, as for parturients away from the window, no matter lying supine or leaning, the natural light basically cannot meet their needs for circadian rhythm stimulation. In that way, supplementary illumination is required. However, during the field measurement, it was found that illuminance values near the window in the ward rose significantly in the morning. If the illuminance value is too high, it is prone to glare and visual fatigue, resulting in physical discomfort. Shading facilities need to be added appropriately to avoid such situations

(2) Natural daylighting of the overall nonvisual effects in the ward can vary significantly with seasonal changes. In summer, nonvisual stimuli received in the area near the window are remarkably better than those near the entrance. While in winter, there are no great changes. Although some areas received a higher value of nonvisual illuminance and a longer duration every day in summer, in terms of meeting the threshold standard and the uniformity, nonvisual natural light environment in the ward is better in winter than that in summer

(3) The change of orientation and window-to-floor ratio has a certain effect on the natural lighting of nonvisual effect in maternity ward. According to the field measurement, the ward facing south has better nonvisual effects than that facing south, which maintain at a higher level. Computer simulation analysis shows that the priority of different directions that wards face in the cold area (Harbin) is as follows: south>east>north>west. When the window-to-floor ratio is 0.3, the nonvisual natural light environment is optimal

Data Availability

The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author/s.

Ethical Approval

Ethical review and approval was not required for this study on human participants in accordance with local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Firstly, I would like to thank my supervisor, Associate Professor Song Haihong, for her patient guidance and assistance during the writing of my dissertation and for her valuable comments on revisions. Secondly, I would also like to thank Mr. Cui Peng for his help in the revision process of my thesis. This work is supported by the Opening Fund of Key Laboratory of

RETRACTED

Hindawi
BioMed Research International
Volume 2024, Article ID 9793629, 1 page
https://doi.org/10.1155/2024/9793629

Retraction

Retracted: Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Xu, D. Liu, M. Cui et al., "Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma," *BioMed Research International*, vol. 2022, Article ID 1893351, 21 pages, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 1893351, 21 pages
https://doi.org/10.1155/2022/1893351

RETRACTED

Research Article

Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma

Shuo Xu, Dingsheng Liu, Mingming Cui, Yao Zhang, Yu Zhang, Shiqi Guo, and Hong Zhang

Department of General Surgery, Shengjing Hospital of China Medical University, 36 Sanhao Street, Heping District, Shenyang 110004, China

Correspondence should be addressed to Hong Zhang; 1994180@cmu.edu.cn

Received 14 April 2022; Accepted 1 June 2022; Published 21 June 2022

Academic Editor: Chen Li

Copyright © 2022 Shuo Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Colon adenocarcinoma (COAD) is among the most common digestive system malignancies worldwide, and its pathogenesis and gene signatures remain unclear. This study explored the genetic characteristics and molecular mechanisms underlying colon cancer development. Three gene expression data sets were obtained from the Gene Expression Omnibus (GEO) database. GEO2R was used to determine differentially expressed genes (DEGs) between COAD and normal tissues. Then, the intersection of the data sets was obtained. Metascape was used to perform the functional enrichment analyses. Next, STRING was used to build protein-protein interaction (PPI) networks. Hub genes were identified and analysed using Cytoscape. Next, survival analysis and expression analysis of the hub genes were performed. ROC curve analysis was performed for further test of the diagnostic efficacy. Finally, alterations in the hub genes were predicted and analysed by cBioPortal. Altogether, 436 DEGs were detected. The DEGs were mainly enriched in cell cycle phase transition, nuclear division, meiotic nuclear division, and cytokinesis. Based on PPI networks, 20 hub genes were selected. Among them, 6 hub genes (*CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE*) showed significant prognostic value in colon cancer ($P < 0.05$), while 5 hub genes (*CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*) were associated with early colon cancer diagnosis and ROC curve analysis showed good diagnostic accuracy. In conclusion, integrated bioinformatics analysis was used to identify hub genes that reveal the potential mechanism of carcinogenesis and progression of colon cancer. The hub genes might be novel biomarkers for early diagnosis, treatment, and prognosis of colon cancer.

1. Introduction

Colon adenocarcinoma (COAD) is among the most common digestive system malignancies worldwide. There were 1,096,601 new colon cancer cases and 551,269 deaths worldwide in 2018 [1]. In the last decade, both the incidence and mortality of colon cancer increased in rapidly transitioning countries including the Baltic countries, Russia, China, and Brazil [2]. As previously reported, the 5-year survival rate was more than 90% for patients diagnosed with stage I, but only 12% for patients diagnosed with stage IV [3]. Thus, early diagnosis and surgical resection of colon cancer will greatly improve disease prognosis. The current early screening tests included noninvasive tests of stool and blood-based tests, radiologic tests, and invasive test like colonoscopy. However, the participation and adherence rates of screening were low, mainly due to the unreliable accuracy of noninvasive tests and low acceptance of the invasive tests as well as the expensive cost [4]. Computed tomographic colonography (CTC) with bowel preparation was reported to have a diagnostic sensitivity of 68.5% and specificity of 88.8% for adenoma ≥ 6 mm, while overall sensitivity (55.3%) and specificity (34.1%) were much lower for adenomas of all sizes [5]. Another study reported that the sensitivity of faecal immunochemical test (FIT) in detecting adenoma, advanced neoplasm, and cancer was 9.5%, 35.1%, and 25.0%, respectively,

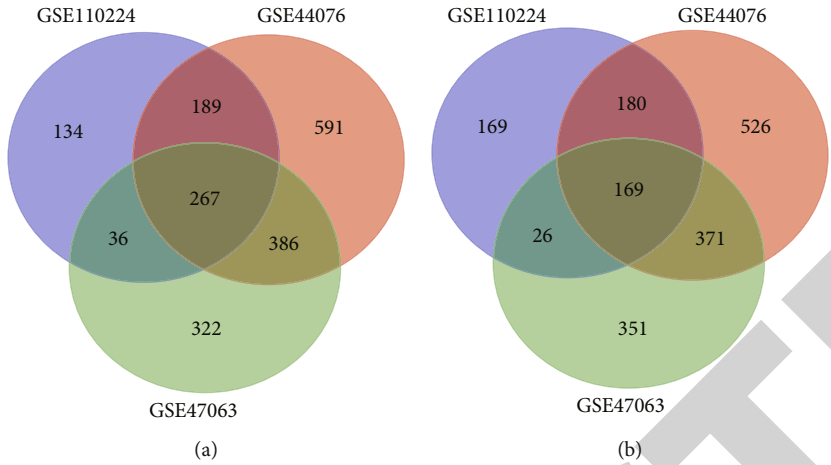


Figure 1: Venn diagram of DEGs from three datasets. (a) 267 downregulated DEGs. (b) 169 upregulated DEGs. Abbreviations: DEGs: differentially expressed genes.

Table 1: Gene ontology (GO) annotation of DEGs in COAD.

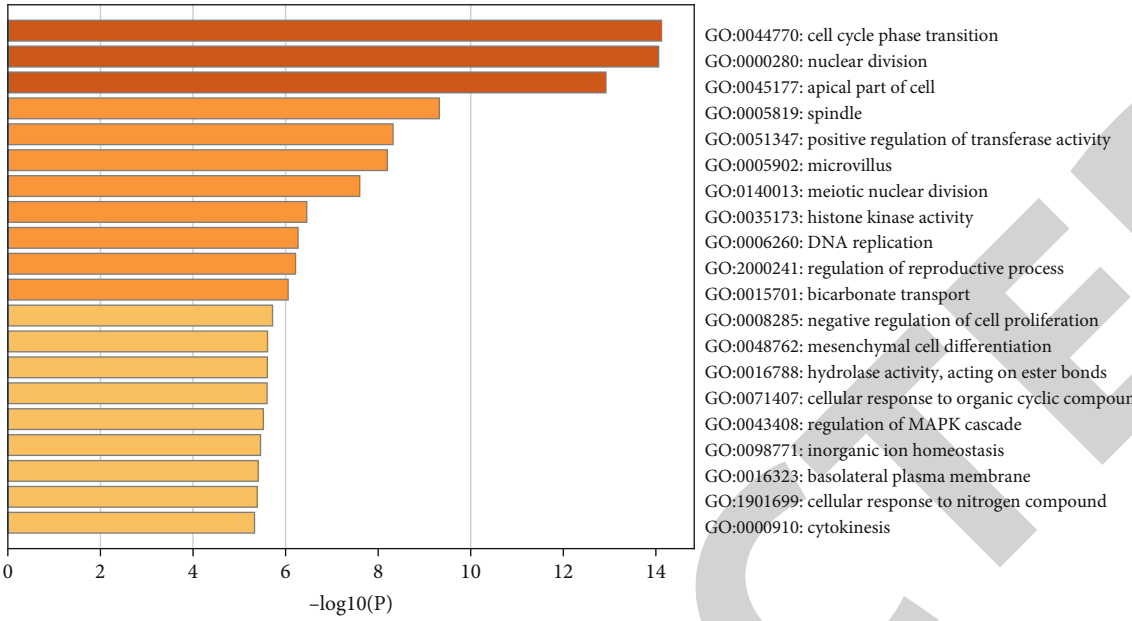
GO	Category	Description	Count	%	Log10(*P*)	Log10(*q*)
GO:0044770	GO biological processes	Cell cycle phase transition	45	10.32	-14.12	-10.06
GO:0000280	GO biological processes	Nuclear division	36	8.26	-14.06	-10.06
GO:0051347	GO biological processes	Positive regulation of transferase activity	37	8.49	-8.32	-5.38
GO:0140013	GO biological processes	Meiotic nuclear division	17	3.90	-7.60	-4.94
GO:0006260	GO biological processes	DNA replication	19	4.36	-6.27	-3.79
GO:2000241	GO biological processes	Regulation of reproductive process	14	3.21	-6.22	-3.75
GO:0015701	GO biological processes	Bicarbonate transport	8	1.83	-6.06	-3.60
GO:0008285	GO biological processes	Negative regulation of cell proliferation	34	7.80	-5.72	-3.29
GO:0048762	GO biological processes	Mesenchymal cell differentiation	16	3.67	-5.61	-3.19
GO:0071407	GO biological processes	Cellular response to organic cyclic compound	27	6.19	-5.60	-3.19
GO:0043408	GO biological processes	Regulation of MAPK cascade	33	7.57	-5.52	-3.13
GO:0098771	GO biological processes	Inorganic ion homeostasis	33	7.57	-5.46	-3.08
GO:1901699	GO biological processes	Cellular response to nitrogen compound	30	6.88	-5.39	-3.02
GO:0000910	GO biological processes	Cytokinesis	14	3.21	-5.33	-2.97
GO:0045177	GO cellular components	Apical part of cell	34	7.80	-12.92	-9.27
GO:0005819	GO cellular components	Spindle	27	6.19	-9.32	-6.23
GO:0005902	GO cellular components	Microvillus	13	2.98	-8.20	-5.35
GO:0016323	GO cellular components	Basolateral plasma membrane	16	3.67	-5.41	-3.04
GO:0035173	GO molecular functions	Histone kinase activity	6	1.38	-6.46	-3.97
GO:0016788	GO molecular functions	Hydrolase activity, acting on ester bonds	33	7.57	-5.61	-3.19

Abbreviations: DEGs: differentially expressed genes; COAD: colon adenocarcinoma.

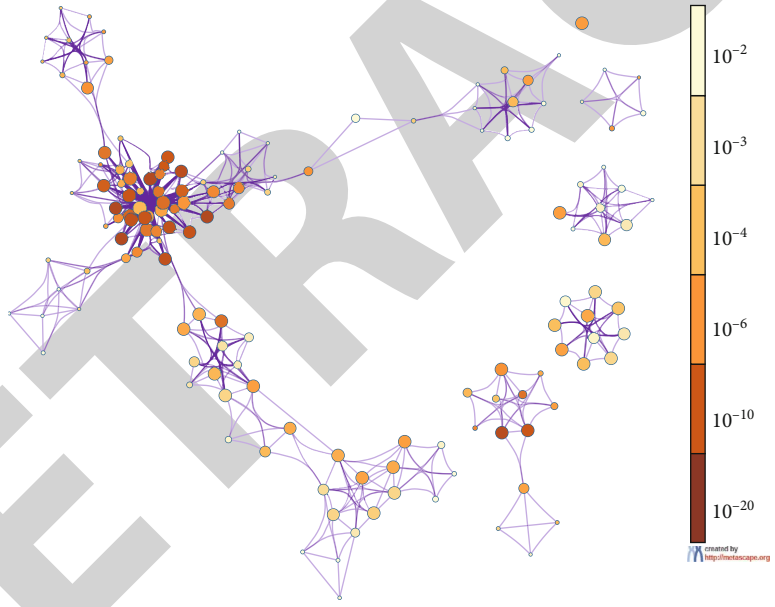
RETRACTED

which showed a low diagnostic accuracy [6]. As a result, only 39% of tumours were diagnosed at an early stage, and the colon cancer remained a serious health burden worldwide [7]. Thus, it is essential to uncover the molecular mechanism and to explore novel biomarkers for early colon cancer diagnosis.

At present, molecular biomarkers are mainly divided into three categories [8]: prognostic biomarkers such as tumour suppressor p53, vascular endothelial growth factor (*VEGF*), and epidermal growth factor receptor (*EGFR*); diagnostic biomarkers such as telomerase and pyruvate kinase M2 (*PKM2*); and predictive biomarkers such as *KRAS* and *B-Raf V600E*. Currently, some molecular markers have been applied in clinical practice. A study confirmed prostaglandin E receptor 4 (PTGER4)/short stature homeobox 2 (SHOX2) DNA methylation as a biomarker for early detection of lung cancer [9]. The panel of trefoil factor (TFF) 1, TFF2, and TFF3 may be potential biomarkers for early screening of breast cancer [10]. However, the accuracy and reliability of many markers were not satisfactory [8, 11]. Therefore, it is urgent to explore a single or a series of accurate and effective markers for early diagnosis and better individualized



(a)



(b)

Figure 2: Continued.

RETRACTED

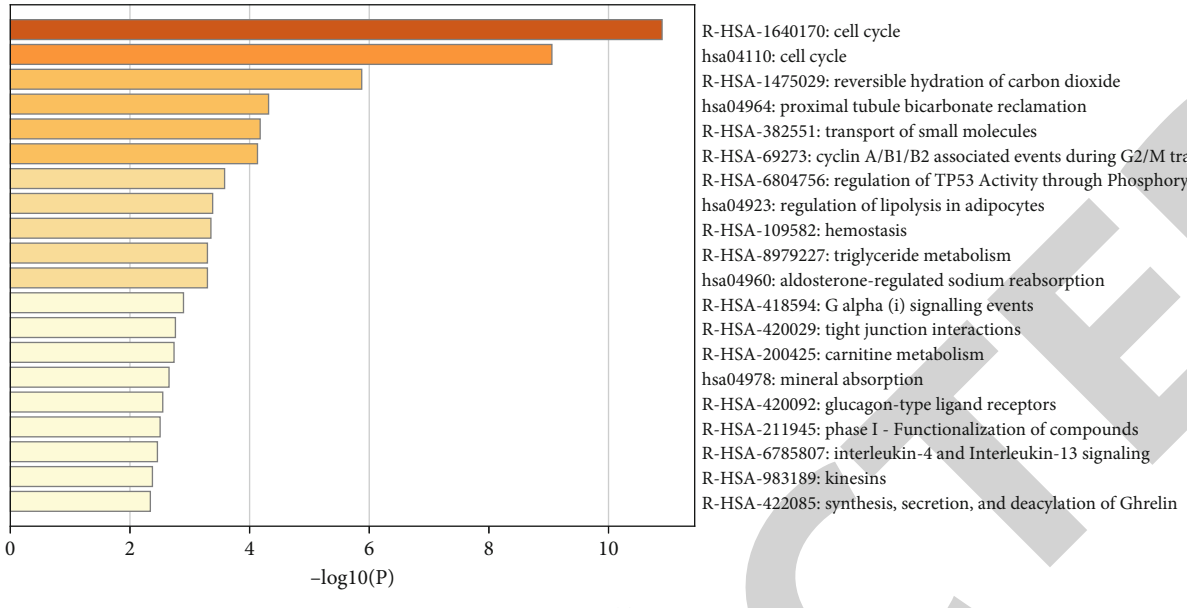


Figure 2: DEG and neighbouring gene enrichment analysis in COAD using Metascape. (a) Heatmap of GO enriched terms coloured by *P* value. (b) Network of GO enriched terms coloured by *P* value. Each node represents an enriched term. Dark colours indicate increased statistical significance. (c) Heatmap of KEGG and Reactome enriched terms coloured by *P* value. (d) Network of KEGG and Reactome enriched terms coloured by *P* value. Each node represents an enriched term. Darker colour indicates more statistical significance. Abbreviations: DEGs: differentially expressed genes; COAD: colon adenocarcinoma; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

RETRACTED

treatment of colon cancer [12]. RNA sequencing and gene expression microarrays were widely applied in cancer studies. Bioinformatics analysis of these data can be used to identify significant biomarkers which may improve cancer early diagnosis, predict prognosis, and inform therapeutic responses [13, 14]. Although there were some previous studies of gene expression in colon cancer, but few studies involved multiple gene expression files and focused on an early diagnosis of the disease. Hence, we performed this study in order to deepen the understanding of the underlying mechanism and provide novel biomarkers for early diagnosis and prognosis of the disease.

2. Materials and Methods

2.1. Microarray Data. We first searched the GEO database [15] and identified three microarray datasets (GSE110224, GSE44076, and GSE47063) [16–18] describing gene expression differences between COAD and normal colon tissue. GSE110224 is based on platform GPL570 ([HG-U133_

Table 2: KEGG and Reactome annotation of DEGs in COAD.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
hsa04110	KEGG pathway	Cell cycle	16	3.67	-9.05	-6.08
hsa04964	KEGG pathway	Proximal tubule bicarbonate reclamation	5	1.15	-4.32	-2.26
hsa04923	KEGG pathway	Regulation of lipolysis in adipocytes	6	1.38	-3.38	-1.5
hsa04960	KEGG pathway	Aldosterone-regulated sodium reabsorption	5	1.15	-3.3	-1.48
hsa04978	KEGG pathway	Mineral absorption	5	1.15	-2.65	-1.03
R-HSA-1640170	Reactome gene sets	Cell cycle	41	9.4	-10.9	-7.44
R-HSA-1475029	Reactome gene sets	Reversible hydration of carbon dioxide	5	1.15	-5.87	-3.47
R-HSA-382551	Reactome gene sets	Transport of small molecules	29	6.65	-4.18	-2.14
R-HSA-69273	Reactome gene sets	Cyclin A/B1/B2 associated events during G2/M transition	5	1.15	-4.13	-2.11
R-HSA-6804756	Reactome gene sets	Regulation of TP53 activity through phosphorylation	8	1.83	-3.58	-1.64
R-HSA-109582	Reactome gene sets	Haemostasis	24	5.5	-3.35	-1.5
R-HSA-8979227	Reactome gene sets	Triglyceride metabolism	5	1.15	-3.3	-1.48
R-HSA-418594	Reactome gene sets	G alpha (i) signalling events	17	3.9	-2.89	-1.16
R-HSA-420029	Reactome gene sets	Tight junction interactions	4	0.92	-2.76	-1.09
R-HSA-200425	Reactome gene sets	Carnitine metabolism	3	0.69	-2.74	-1.09
R-HSA-420092	Reactome gene sets	Glucagon-type ligand receptors	4	0.92	-2.55	-0.95
R-HSA-211945	Reactome gene sets	Phase I—functionalization of compounds	7	1.61	-2.5	-0.91
R-HSA-6785807	Reactome gene sets	Interleukin-4 and interleukin-13 signalling	7	1.61	-2.46	-0.88
R-HSA-983189	Reactome gene sets	Kinesins	5	1.15	-2.38	-0.82
R-HSA-422085	Reactome gene sets	Synthesis, secretion, and deacylation of ghrelin	3	0.69	-2.34	-0.8

Abbreviations: KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: differentially expressed genes; COAD: colon adenocarcinoma.

Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), GSE44076 is based on platform GPL13667 ([HG-U219] Affymetrix Human Genome U219 Array), and GSE47063 is based on platform GPL6102 (Illumina human-6 v2.0 expression beadchip). All data are freely available online.

2.2. DEG Identification. GEO2R is commonly used to process sample information from GEO series and to identify DEGs among user-defined groups. After screening the sample information in the three data sets, only the COAD samples and the corresponding normal tissues were included. After GEO2R analysis, DEGs were obtained by intersecting genes with an adjusted $P < 0.05$ and $|\log FC| \geq 1$ in each data set using a Venn diagram.

2.3. Gene Ontology and Pathway Enrichment Analysis of DEGs. Metascape [19] is an open access online tool for comprehensive gene list annotation and analysis. In this study, DEG pathway and process enrichment analyses were performed using Metascape. The parameters were set as follows: 3 for min overlap, 1.5 for min enrichment, and P value cutoff of 0.05. The enrichment results were presented as bar charts. Corresponding network graph nodes with similarity degree more than 0.3 were connected with curved edges. Edge thickness was positively correlated with the degree of similarity.

2.4. PPI Network Construction and Module Analysis. The Search Tool for the Retrieval of Interacting Genes (STRING) database [20] was used to construct the PPI network with an interaction score > 0.4. Then, Cytoscape (Version 3.7.2) [21] software was used to visualise and analyse PPI networks. Molecular Complex Detection (MCODE) (Version 1.6) [22], a Cytoscape plugin, was used to identify the most significant gene module in colon cancer. Then, we annotated the function of the module genes using Metascape.

2.5. Hub Gene Selection and Analysis. CytoHubba (Version 0.1) [23], a Cytoscape plugin, was used to identify the network hub genes. We used a degree-ranked method to identify hub genes with a criterion of degree no less than 67. ClueGO [24] is another Cytoscape plugin that can creates and visualises functionally grouped networks of biological terms and pathways. The CluePedia [25] Cytoscape plugin is a functional extension of ClueGO and a search tool for new markers potentially associated with pathways. In our study, ClueGO (Version 2.5.6) and CluePedia (Version 1.5.6) were used to analyse the biological processes and pathway enrichment of hub genes.

2.6. Analysis of Prognostic Value of Hub Genes. GEPIA [26] is an integrated bioinformatics analysis tool which was designed for transforming genomic big data into intuitive graphics. In this study, GEPIA was used to perform survival analysis based on gene expression. $P < 0.05$ was considered statistically significant.

2.7. Hub Gene Expression Analysis and ROC Curve Analysis. UALCAN [27] is a comprehensive interactive online resource which contains clinical data from 31 cancer types

RETRACTED

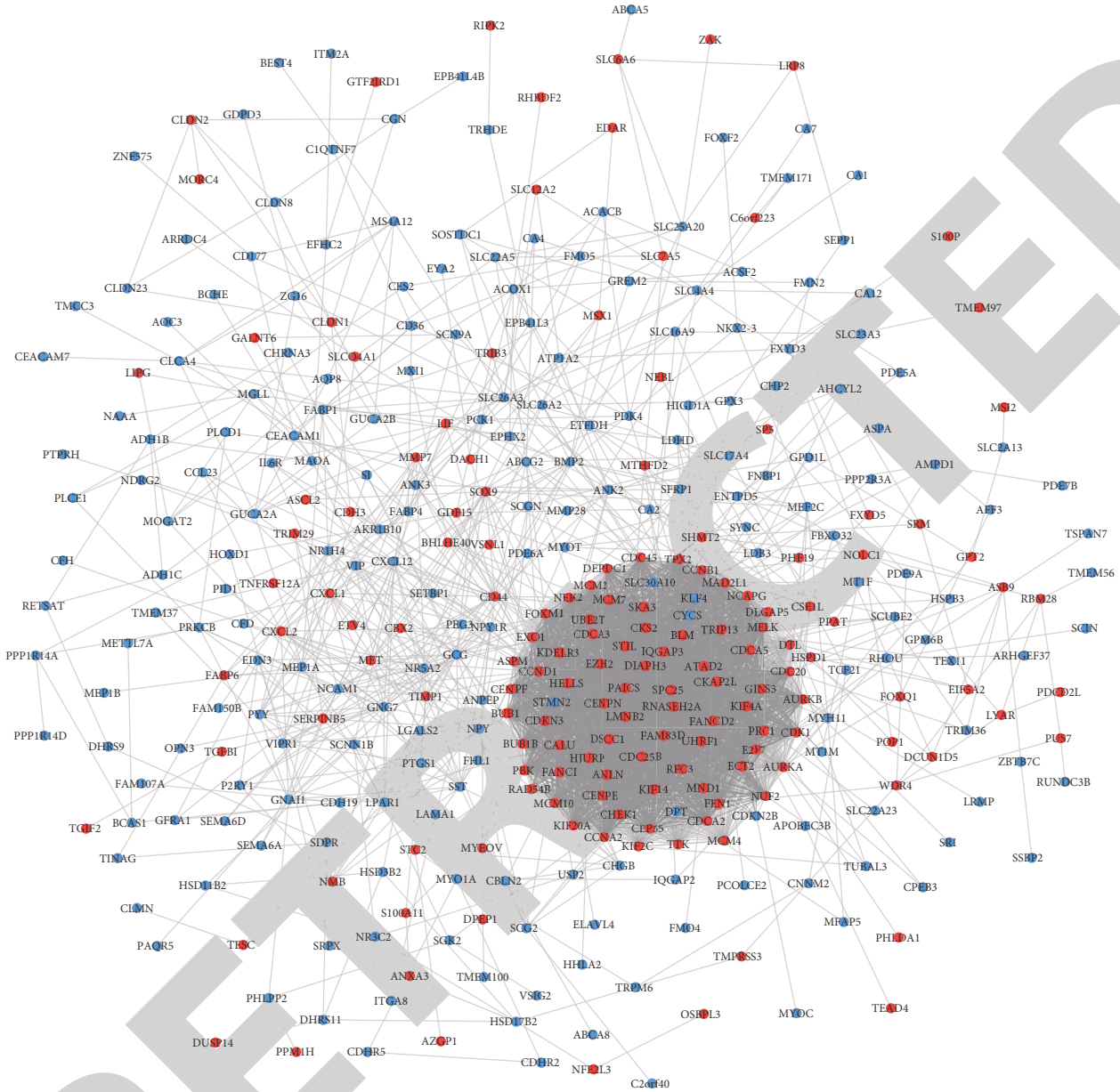


Figure 3: PPI network of DEGs, containing 369 nodes and 2708 edges. Red represents upregulated genes. Blue represents downregulated genes. Abbreviations: PPI: protein-protein interaction; DEGs: differentially expressed genes.

from the TCGA database. We used UALCAN to perform differential expression analysis of the hub genes and their association with clinicopathological parameters of COAD patients. Moreover, the Human Protein Atlas [28] is a website for users to freely access data for exploration of the human proteome, which contains transcriptome data from 17 main cancer types using data from nearly 8000 patients. In this study, histopathological data of the hub genes were downloaded and used for direct comparison the protein expression. We selected an additional dataset for ROC curve analysis of diagnostic accuracy for the hub genes. GSE87211 [29] is based on platform GPL13497 (Agilent-026652 Whole Human Genome Microarray 4x44K v2). All data are freely available online.

2.8. Analysis of Alterations of Hub Genes. cBioPortal [30] is a free web server for interactively exploring cancer genomics datasets. In this study, cBioPortal was utilised to predict the genetic alterations of eight hub genes in 378 COAD samples (TCGA, PanCancer Atlas) which contained mutations and putative copy-number alterations from GISTIC and mRNA expression z-scores (RNASeq V2 RSEM) with a z-score threshold ±2.0.

2.9. Statistical Analysis. Microarray data analysis was performed by using GEO2R. GEOquery R package was used to transform the original data into R data structure, and then, the statistical test of limma (linear models for microarray analysis) R package was used to identify DEGs. Survival

RETRACTED

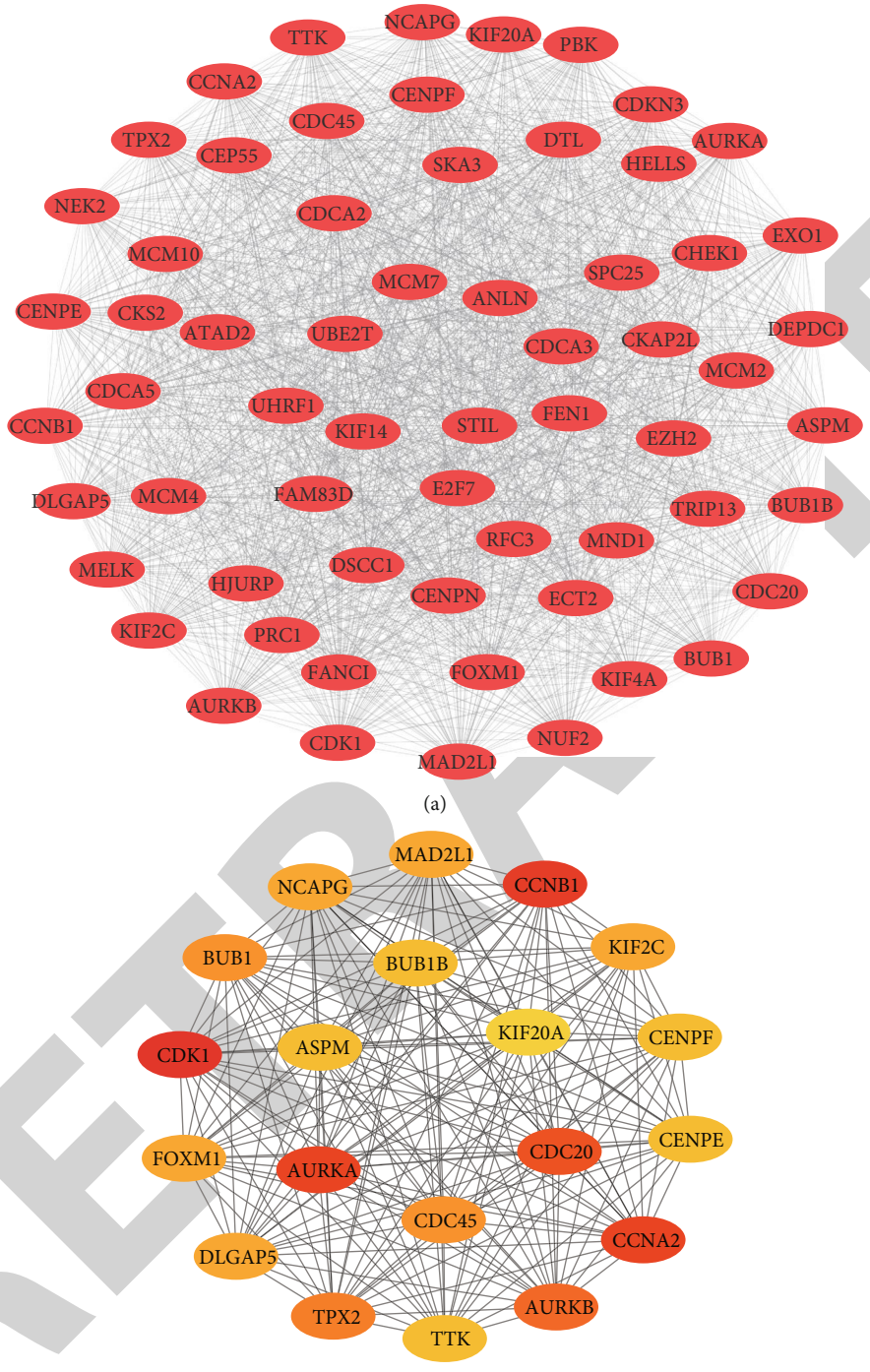
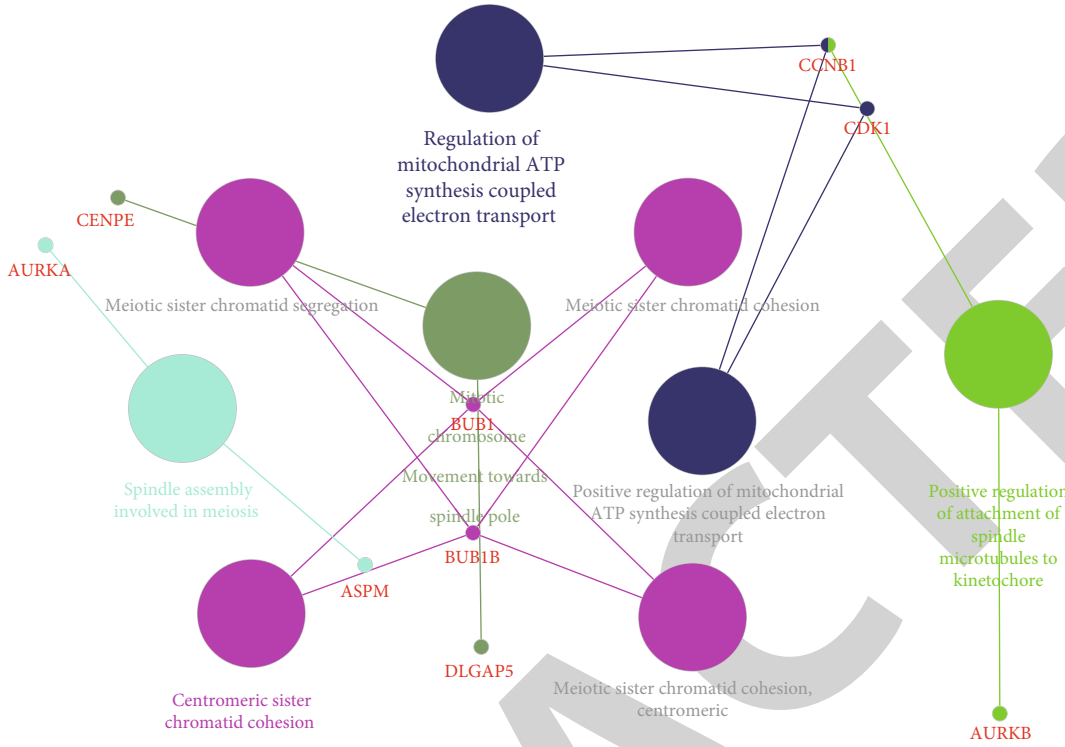
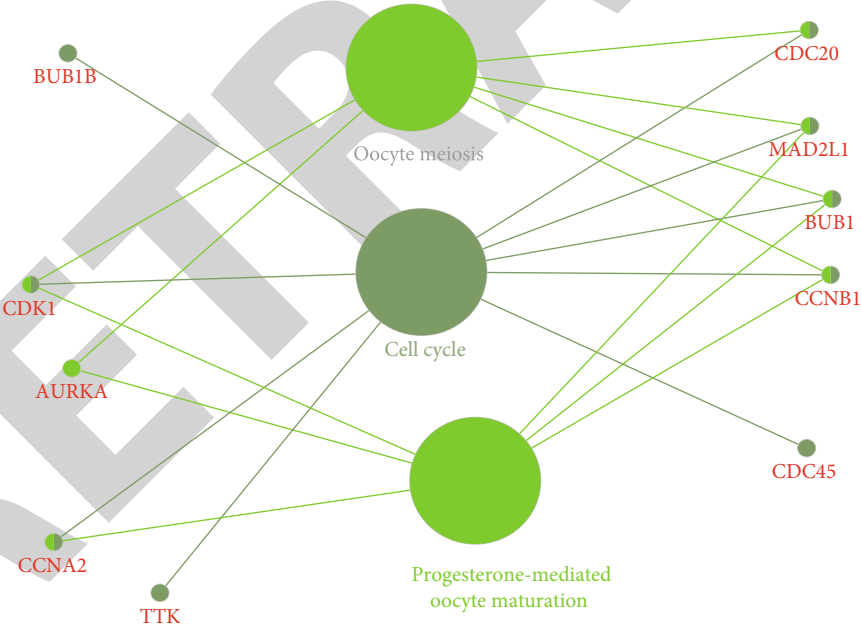


Figure 4: Continued.

RETRACTED



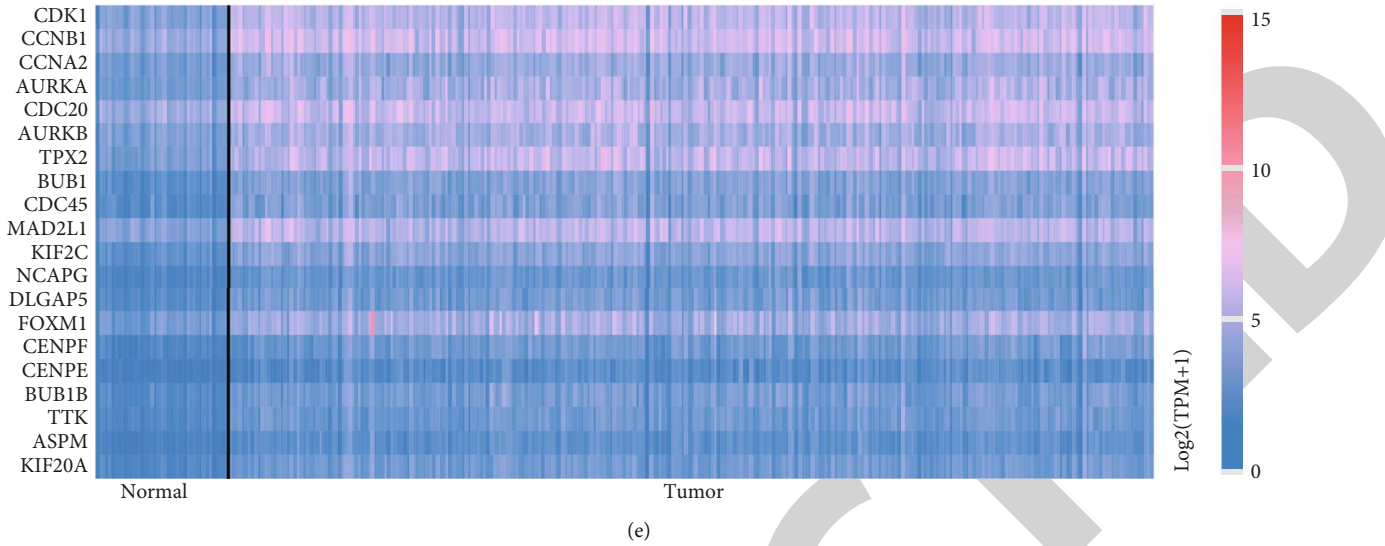
(c)



(d)

Figure 4: Continued.

RETRACTED



(e)

Figure 4: The most significant module gene network and hub genes analysis. (a) The most significant module in the PPI network contains 62 nodes and 1708 edges. (b) Network of 20 hub genes. Darker colours represent higher scores. (c) Biological process annotation of hub genes using ClueGO and CluePedia. $P < 0.01$ was considered statistically significant. (d) KEGG annotation of hub genes using ClueGO and CluePedia. $P < 0.01$ was considered statistically significant. (e) Heatmap of the top 20 hub genes was constructed using the UALCAN database. Abbreviations: PPI: protein-protein interaction; KEGG: Kyoto Encyclopedia of Genes and Genomes.

analysis was performed by using GEPIA and log-rank test. The transcripts per million (TPM) expression value and *t*-test were used for analysis of the relationship between hub genes expression and clinicopathological parameters. SPSS 26.0 was used for ROC curve analysis. $P < 0.05$ was considered statistically significant.

3. Results

3.1. DEGs in Colon Cancer. Among the three datasets (GSE110224, GSE44076, and GSE47063), there were 127 COAD tissues and 117 normal tissues. After GEO2R analysis, we screened 1617 DEGs (745 upregulated and 872 downregulated) from GSE110224, 4450 DEGs (2095 upregulated and 2355 downregulated) from GSE44076, and 2259 DEGs (1056 upregulated and 1203 downregulated) from GSE47063. Then, 436 DEGs were obtained by overlapping the three dataset results, including 267 downregulated genes (Figure 1(a)) and 169 upregulated genes (Figure 1(b)).

3.2. DEG Gene Ontology (GO) and Pathway Enrichment in Colon Cancer. The top 20 GO items were divided into 3 categories: biological processes (14 items), cellular components (4 items), and molecular functions (2 items; Table 1 and Figures 2(a) and 2(b)). The DEGs were mainly enriched in cell cycle, transcriptional regulation, and ion transport. Enriched biological processes included cell cycle phase transition, nuclear division, meiotic nuclear division, cytokinesis, DNA replication, negative regulation of cell proliferation, regulation of reproductive process, regulation of MAPK cascade, positive regulation of transferase activity, bicarbonate transport, inorganic ion homeostasis, cellular response to organic cyclic compound, cellular response to nitrogen compound, and mesenchymal cell differentiation. Cellular component analysis showed that the DEGs were significantly enriched in the apical part of the cell, spindle, microvillus, and basolateral plasma membrane. Molecular functions of these genes were histone kinase activity and activity of hydrolase acting on ester bond.

The top 20 Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathways were shown in Table 2 and Figures 2(c) and 2(d). DEGs were mainly enriched for terms associated with the cell cycle, reversible hydration of carbon dioxide, proximal tubule bicarbonate reclamation, transport of small molecules, cyclin A/B1/B2 associated events during G2/M transition, and regulation of TP53 activity through phosphorylation pathway.

3.3. DEG PPI Network and Modules. A PPI network composed of 369 nodes and 2708 edges was constructed (Figure 3). Then, MCODE was used to isolate the significant network modules. We selected the most significant module with the highest degree (Figure 4(a)) and functionally annotated the involved genes (Table 3). GO enrichment analysis showed that the genes were mainly enriched in biological processes, including chromosome segregation, cell cycle phase transition, positive regulation of cell cycle, DNA replication, meiotic cell cycle, attachment of spindle microtubules to kinetochore, DNA conformation change, signal transduction by p53 class mediator, positive regulation of transferase activity, sister chromatid cohesion, cytokinetic process, and protein localisation to cytoskeleton. Cellular component analysis showed that these genes were mainly enriched in the spindle, midbody, kinesin complex, and intercellular bridge. Molecular function analysis showed that these genes were mainly enriched in catalytic activity, acting on DNA, and chromatin binding. Pathway analysis revealed that these genes were mainly enriched in cyclin A/B1/B2-

RETRACTED

Table 3: Functional annotation of the genes involved in the most significant module.

GO	Category	Description	Count	%	Log10(*P*)	Log10(*q*)
GO:0007059	GO biological processes	Chromosome segregation	29	46.77	-37.06	-32.66
GO:0044770	GO biological processes	Cell cycle phase transition	33	53.23	-35.09	-31.16
GO:0045787	GO biological processes	Positive regulation of cell cycle	17	27.42	-15.9	-13.21
GO:0006260	GO biological processes	DNA replication	14	22.58	-14.21	-11.7
GO:0051321	GO biological processes	Meiotic cell cycle	12	19.35	-11.78	-9.35
GO:0008608	GO biological processes	Attachment of spindle microtubules to kinetochore	7	11.29	-11.68	-9.27
GO:0071103	GO biological processes	DNA conformation change	12	19.35	-10.45	-8.08
GO:0072331	GO biological processes	Signal transduction by p53 class mediator	10	16.13	-8.81	-6.51
GO:0051347	GO biological processes	Positive regulation of transferase activity	14	22.58	-8.75	-6.45
GO:0007062	GO biological processes	Sister chromatid cohesion	6	9.68	-7.87	-5.62
GO:0032506	GO biological processes	Cytokinetic process	5	8.06	-7.29	-5.1
GO:0044380	GO biological processes	Protein localisation to cytoskeleton	5	8.06	-6.41	-4.27
GO:0005819	GO cellular components	Spindle	23	37.1	-25.83	-22.66
GO:0030496	GO cellular components	Midbody	14	22.58	-16.65	-13.9
GO:0005871	GO cellular components	Kinesin complex	5	8.06	-6.53	-4.38
GO:0045171	GO cellular components	Intercellular bridge	5	8.06	-5.97	-3.86
GO:0140097	GO molecular functions	Catalytic activity, acting on DNA	7	11.29	-5.87	-3.78
GO:0003682	GO molecular functions	Chromatin binding	10	16.13	-5.86	-3.76
R-HSA-69273	Reactome gene sets	Cyclin A/B1/B2 associated events during G2/M transition	4	6.45	-6.32	-4.18
R-HSA-179409	Reactome gene sets	APC-Cdc20 mediated degradation of Nek2A	4	6.45	-6.32	-4.18

Abbreviations: GO: Gene Ontology.

associated events during G2/M transition and APC-Cdc20-mediated degradation of Nek2A.

3.4. Hub Genes. According to the node degree calculated by CytoHubba, 20 hub genes were screened out, and they were all upregulated (Figure 4(b)). The gene symbols and corresponding degree were shown in Table 4. Functional annotation of the 20 hub genes was shown in Figures 4(c) and 4(d). Heat map visualisation showed that the expression of these 20 hub genes in COAD tissues was higher than in normal tissues (Figure 4(e)).

3.5. Survival Based on Hub Gene Expression. Because several hub genes were closely related to the cell cycle, we further analysed their survival curves using the GEPIA database. Our results showed that overexpression of six hub genes influenced COAD prognosis, including *CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE*. Overexpression of the six genes was associated with favourable overall survival (OS) of colon cancer patients (Figures 5(a)–5(f)). Additionally, *AURKA* and *CENPE* overexpressions showed a favourable prognosis of disease-free survival (DFS) in COAD patients (Figures 5(g) and 5(h)).

3.6. Differential Expression of Hub Genes. UALCAN was used to analyse mRNA expression of the identified hub genes. We found 5 hub genes were related to clinicopathological parameters, including *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*. Additionally, we observed that these five genes were significantly overexpressed in tumour

Table 4: Top 20 hub genes and corresponding degree.

Gene symbol	Gene description	Score
CDK1	Cyclin dependent kinase 1	78
CCNB1	Cyclin B1	76
CCNA2	Cyclin A2	75
AURKA	Aurora kinase A	75
CDC20	Cell division cycle 20	74
AURKB	Aurora kinase B	72
TPX2	TPX2 microtubule nucleation factor	71
BUB1	BUB1 mitotic checkpoint serine/threonine kinase	70
CDC45	Cell division cycle 45	70
MAD2L1	Mitotic arrest deficient 2 like 1	69
KIF2C	Kinesin family member 2C	69
NCAPG	Non-SMC condensin I complex subunit G	69
DLGAP5	DLG associated protein 5	69
FOXM1	Forkhead box M1	69
CENPF	Centromere protein F	68
CENPE	Centromere protein E	68
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	68
TTK	TTK protein kinase	68
ASPM	Abnormal spindle microtubule assembly	68
KIF20A	Kinesin family member 20A	67

RETRACTED

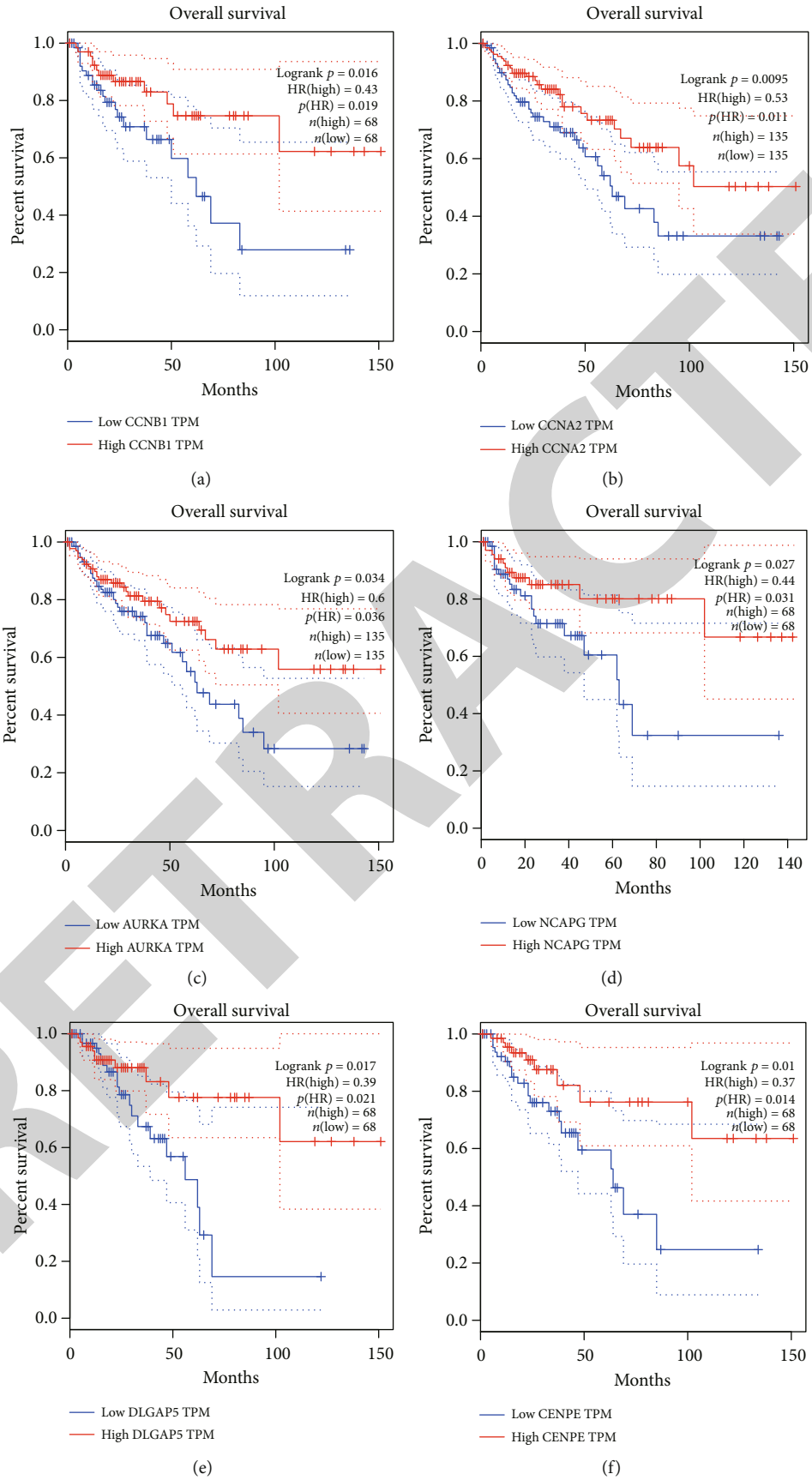


Figure 5: Continued.

RETRACTED

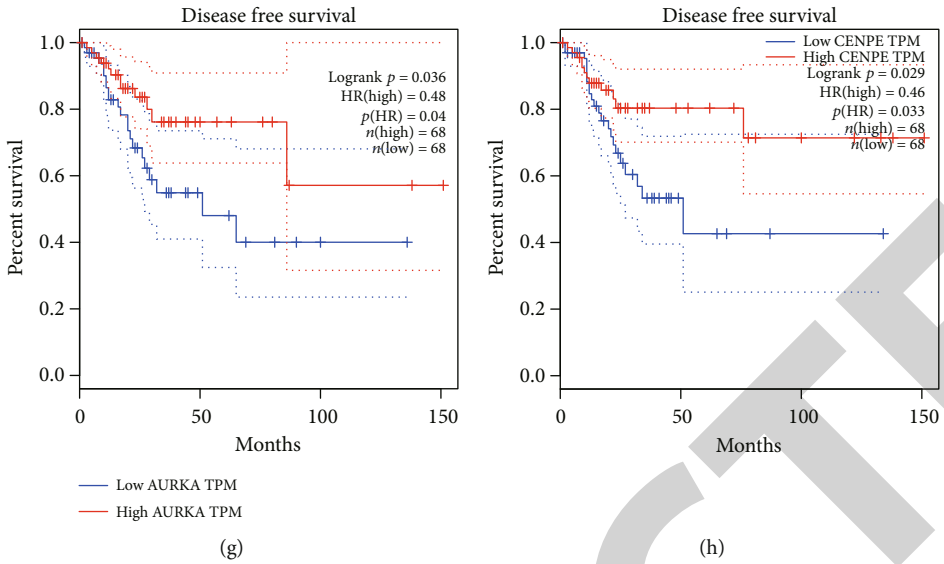


Figure 5: Overall survival of the hub genes in COAD patients. (a)–(f) *CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE* showed a significant difference in overall survival (OS). High expression of the 6 genes indicated favourable OS in COAD ($P < 0.05$). (g, h) *AURKA* and *CENPE* showed statistically significant association with disease-free survival (DFS) and indicated favourable disease-free survival in COAD ($P < 0.05$). Abbreviations: COAD: colon adenocarcinoma.

tissues (Figures 6(a), 6(d), 6(g), 6(j), and 6(m)). Then, we analysed their mRNA expression under different clinicopathological parameters. Our results revealed that the mRNA expression of the five genes was significantly correlated with the clinical stage, and that the highest mRNA expression appeared in the first tumour stage (Figures 6(b), 6(e), 6(h), 6(k), and 6(n)). Moreover, the mRNA expression of the five genes showed a significant correlation with lymph node metastasis, and the highest mRNA expression appeared at the N0 phase (Figures 6(c), 6(f), 6(i), 6(l), and 6(o)).

Moreover, we analysed the protein expressions of hub genes using histopathological images from HPA. Our results showed that CDK1 staining was low in normal tissues and moderate in COAD tissues (Figure 7(a)). CCNB1 and CCNA2 staining were moderate in normal colon tissues, whereas high staining was observed in COAD tissues (Figures 7(b) and 7(c)). DLGAP5 staining was not detected in normal tissues, while moderate staining was observed in COAD tissues (Figure 7(d)). MAD2L1 was moderately stained in both tumour and normal tissues (Figure 7(e)).

In order to further test the diagnostic efficacy of these hub genes for colon cancer, ROC curve analysis was performed on these five genes (Figure 8). We used gene expression data from GSE87211 for analysis. The dataset contained 363 cases (203 colon tumours and 160 healthy mucosa). AUCs were used to assess the diagnostic accuracy. ROC analysis showed that AUCs for *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5* were 0.928 (95% CI: 0.901-0.956), 0.931 (95% CI: 0.905-0.956), 0.904 (95% CI: 0.847-0.934), 0.917 (95% CI: 0.887-0.947), and 0.911 (95% CI: 0.881-0.940), respectively.

3.7. Alteration of Hub Genes. We also analysed alterations of the six prognostic hub genes *CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE* together with the five hub genes which were associated with clinicopathological parameters: *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*. Eight hub genes including *CDK1*, *CCNB1*, *CCNA2*, *AURKA*, *MAD2L1*, *NCAPG*, *DLGAP5*, and *CENPE* were detected by cBioPortal.

Altogether, 378 samples of COAD were included, and our analysis revealed that the hub genes were altered in 42.86% of the 378 samples. *AURKA* (28%) was the most frequently altered gene of the eight hub genes (Figure 9).

4. Discussion

Colon cancer was the fourth most commonly diagnosed malignant tumour worldwide in 2018, with increasing incidence in countries undergoing major developmental transition [31]. Due to a lack of specific symptoms for early detection, patients are usually diagnosed at an advanced stage which leads to a poor prognosis [32]. Therefore, it is crucial to uncover the underlying molecular mechanism and to explore key biomarkers for early colon cancer diagnosis.

In this study, we analysed three microarray datasets that included 127 tumours and 117 normal samples. A total of 436 DEGs were screened. Functional annotation showed that the DEGs were mainly enriched in biological processes associated with cell cycle phase transition, nuclear division, positive regulation of transferase activity, meiotic nuclear division, and DNA replication. These results suggested that these genes were closely related to the cell cycle. Many studies indicated that dysregulation of cell cycle progression was closely related to cancer progression [33, 34]. Finetti et al. [35] found that several genes participated in regulating the cell cycle, like *CDK1* and *AURKA*. Moreover, their expressions were correlated with breast cancer prognosis. In our colon cancer study, we obtained many DEGs involved in cell

RETRACTED

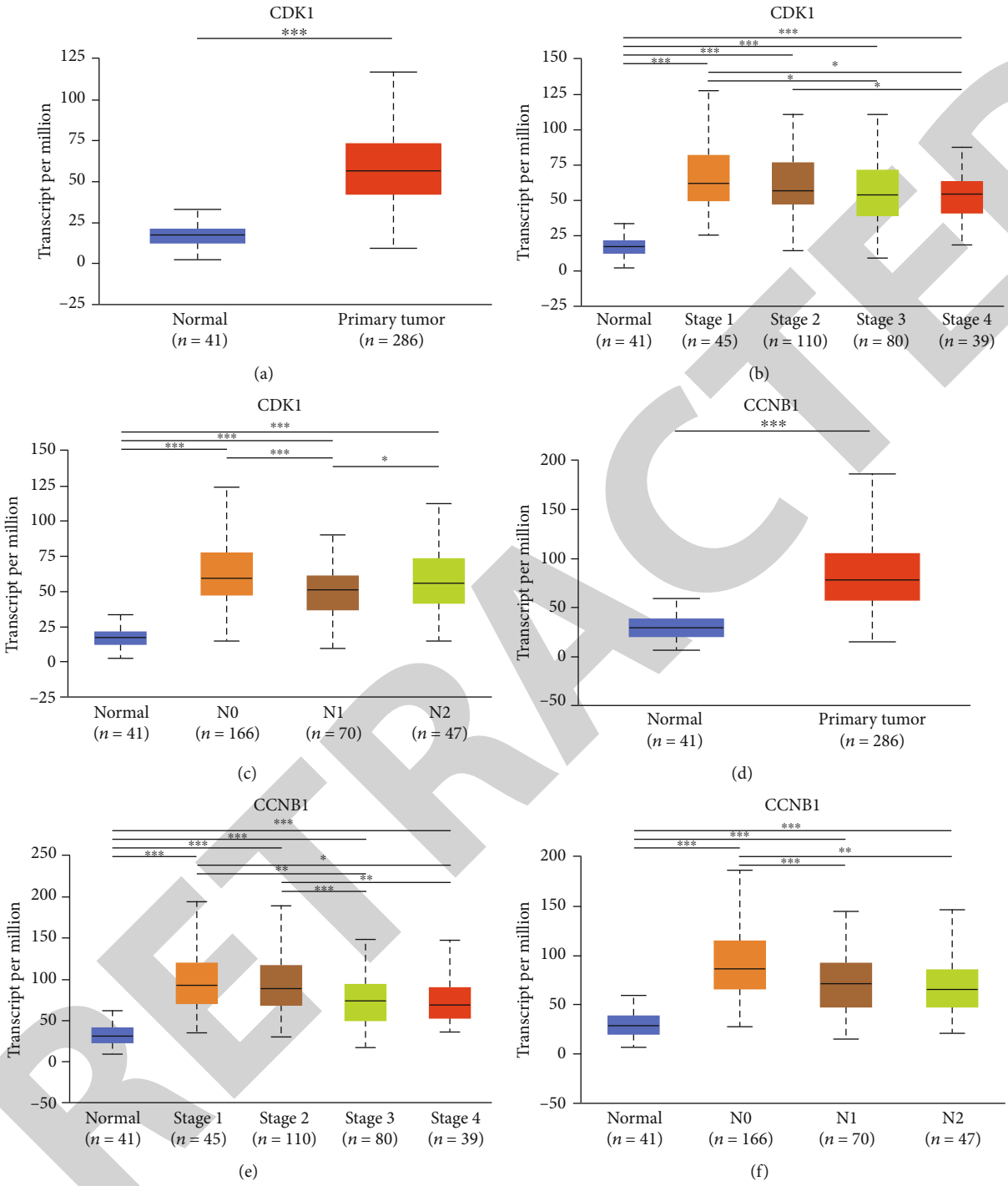


Figure 6: Continued.

RETRACTED

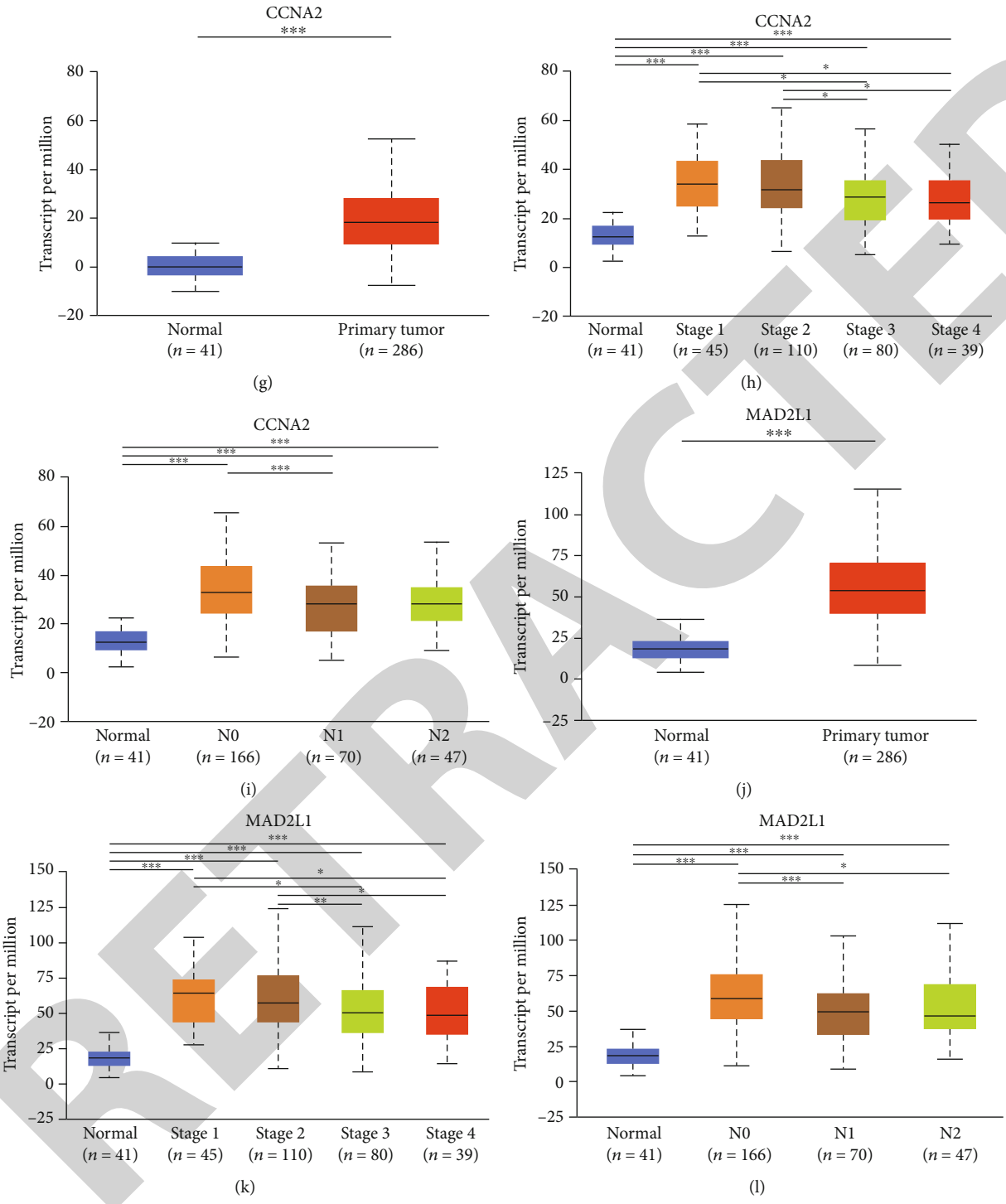


Figure 6: Continued.

RETRACTED

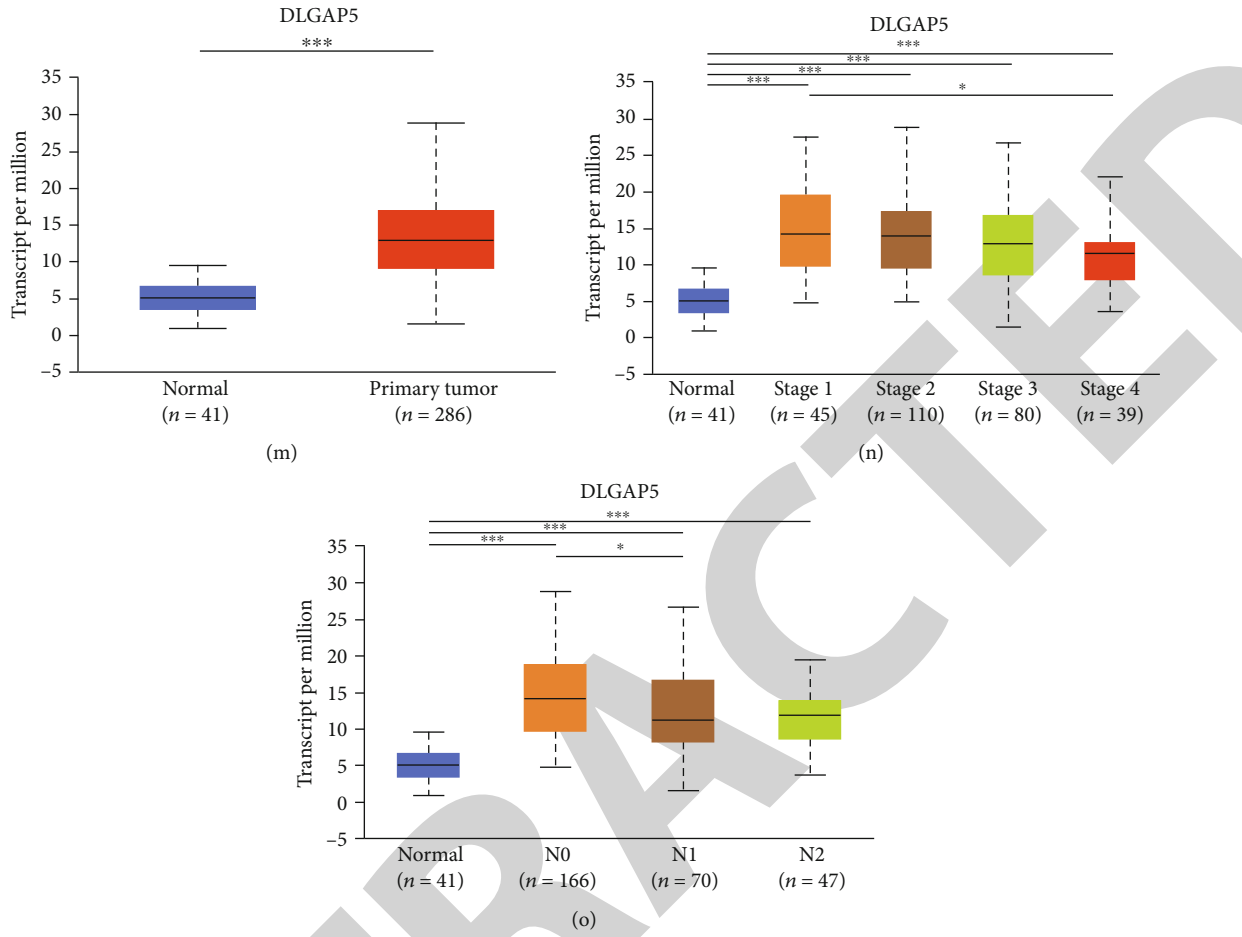


Figure 6: Differential expression analysis of the 5 hub genes was performed by UALCAN. (a, d, g, j, and m) mRNA expression of the five genes was overexpressed in colon cancer compared to normal colon tissues. (b, e, h, k, and n) mRNA expression of the five genes was significantly related to individual cancer stage, with the highest expressions tending to appear at stage 1. (c, f, i, l, and o) mRNA expression of the five genes was significantly related to nodal metastasis status, and the highest mRNA expression tended to appear at the N0 phase. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

RETRACTED

cycle progression, including *CCND1*, *BLM*, *BUB1*, *BUB1B*, *CCNA2*, *CCNB1*, *CDK1*, and *CDC20*. Some genes were closely related to the transformation of cancer. For example, *CCND1* belonged to the cyclin family whose members were characterised by dramatic periodicity in protein abundance throughout the cell cycle. Deregulation of *CCND1* was observed frequently in numerous human cancers, including pancreatic cancer, head and neck squamous cell carcinoma, breast cancer, and colorectal carcinoma [36, 37]. Accumulation of *CCND1* in the nucleus caused uncontrolled cell cycle progression and acted as a tumour-initiating event [38]. Overexpression of cyclin D1 (T286A), an oncogenic mutant allele of *CCND1*, promoted stabilization and overexpression of the DNA replication licensing factor, Cdt1, by inhibiting its proteolysis. This caused DNA rereplication and damage and resulted in cellular aneuploidy, genomic instability, and further neoplastic growth [39]. Cyclin dependent kinases (CDKs) were necessary functional partner kinases with cyclin D1. Thus, CDK inhibitors would be an effective drug for targeting malignant tumours [40]. However, given the development of resistance and side effects of CDK inhibitors, further research is warranted [36].

Pathway analysis also revealed that DEGs were mainly enriched for terms associated with the cell cycle pathway. Cyclin A/B1/B2-associated events in the "G2/M transition" and "Regulation of TP53 Activity through Phosphorylation" pathways were closely related to tumourigenesis. Like the cyclin D1 mentioned above, cyclins A/B1/B2 were also cyclin members that binded to CDKs and regulated the cell cycle. Abundant evidence showed that G2/M phase arrest was closely related to the inhibition of tumour cell proliferation [41, 42]. Additional studies focusing on cyclins are aimed at identifying novel therapeutic strategies for cancer treatment. Ma [43] revealed that the microRNA miR-219-5p downregulated *CCNA2* expression and induced G2/M phase arrest to inhibit tumour formation in oesophageal cancer. Tu et al. [44] found *CCNA2* was downregulated by the small molecule FH535 in colorectal cancer, which caused G2/M phase arrest and inhibited tumour proliferation. Thus, inhibiting *CCNA2* and *CCNB1* may contribute to the

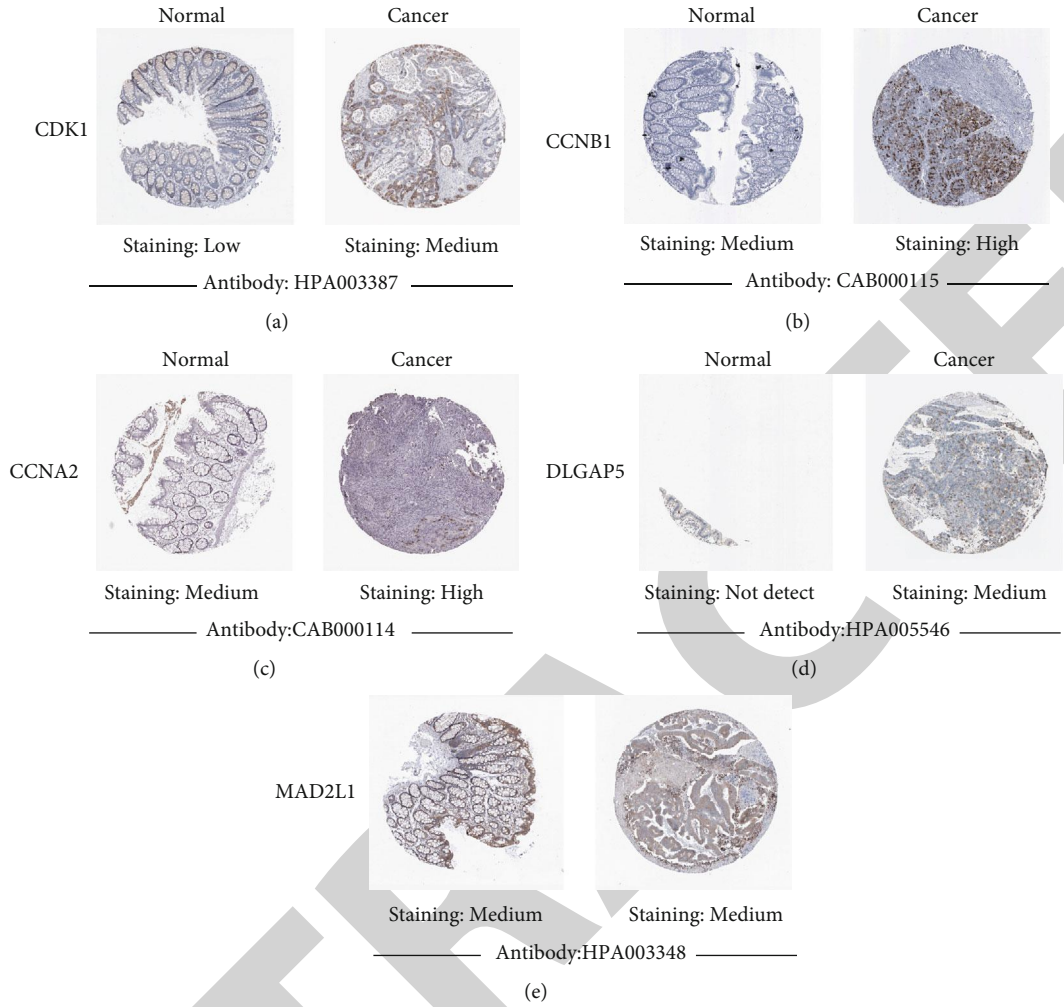


Figure 7: Protein expression analysis of the 5 hub genes was performed using the HPA database. Except for MAD2L1, the other 4 proteins showed a higher degree of staining in tumour tissue compared to normal tissues.

development of novel anticancer drugs. The p53 signalling pathway significantly contributed to cell cycle regulation, suppression of tumour expression, metabolism, aging, development, and reproduction [45]. Phosphorylation of p53 protein stabilized the protein and extended its half-life, thus, causing cell cycle arrest, apoptosis, and inhibited tumour cell proliferation [46]. A study of natural polyphenols as anticancer agents revealed that polyphenols could induce apoptosis, which was achieved by stabilizing p53 protein through phosphorylation and showed remarkable effects in human gastric carcinoma cells [47]. We also identified some pathways associated with metabolism, including triglyceride metabolism, carnitine metabolism, regulation of lipolysis in adipocytes, and phase I—functionalization of compounds. Among these pathways, we found that *FABP4*, which encoded fatty acid binding protein, was involved in fatty acid uptake, transport, and metabolism and was related to tumour metastasis. Gharpure et al. [48] observed that overexpression of *FABP4* played a key role in aggressive metastasis of ovarian cancer via various metabolites and protein pathways. Likewise, *FABP4* had crucial effects on adipocyte-induced cholangiocarcinoma metastasis [49]. Collectively, metabolic disorder was among the leading causes of tumour development. Thus, the study of tumour metabolism may provide new targets for tumour treatment.

The PPI network was built using STRING. Twenty hub genes were screened, and their functional annotations were most closely related to the cell cycle. Survival analysis showed that higher mRNA expression of six hub genes was significantly related to longer OS in colon cancer patients, including *CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE*. Moreover, *AURKA* and *CENPE* exhibited favourable effects on both OS and DFS. Studies showed that *CCNB1* was highly expressed in colorectal cancer tissues and was negatively correlated with tumour invasion and distant metastasis, which may be caused by regulating the expression of E-cadherin [50]. This was consistent with our findings. A murine colorectal cancer model showed that *CCNA2* deletion in colonic epithelial cells promoted the development of dysplasia and adenocarcinomas [51]. Analysis of *CCNA2* expression in clinical samples revealed that higher expression of *CCNA2* in tumours of stage 1 or 2 colon cancer patients is compared with stage 3 or 4 patients [51], which was also consistent with our results. However,

RETRACTED

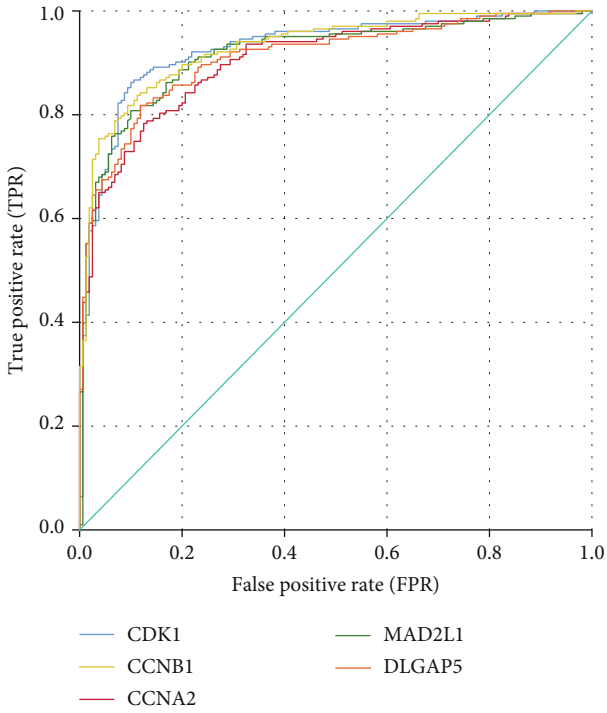


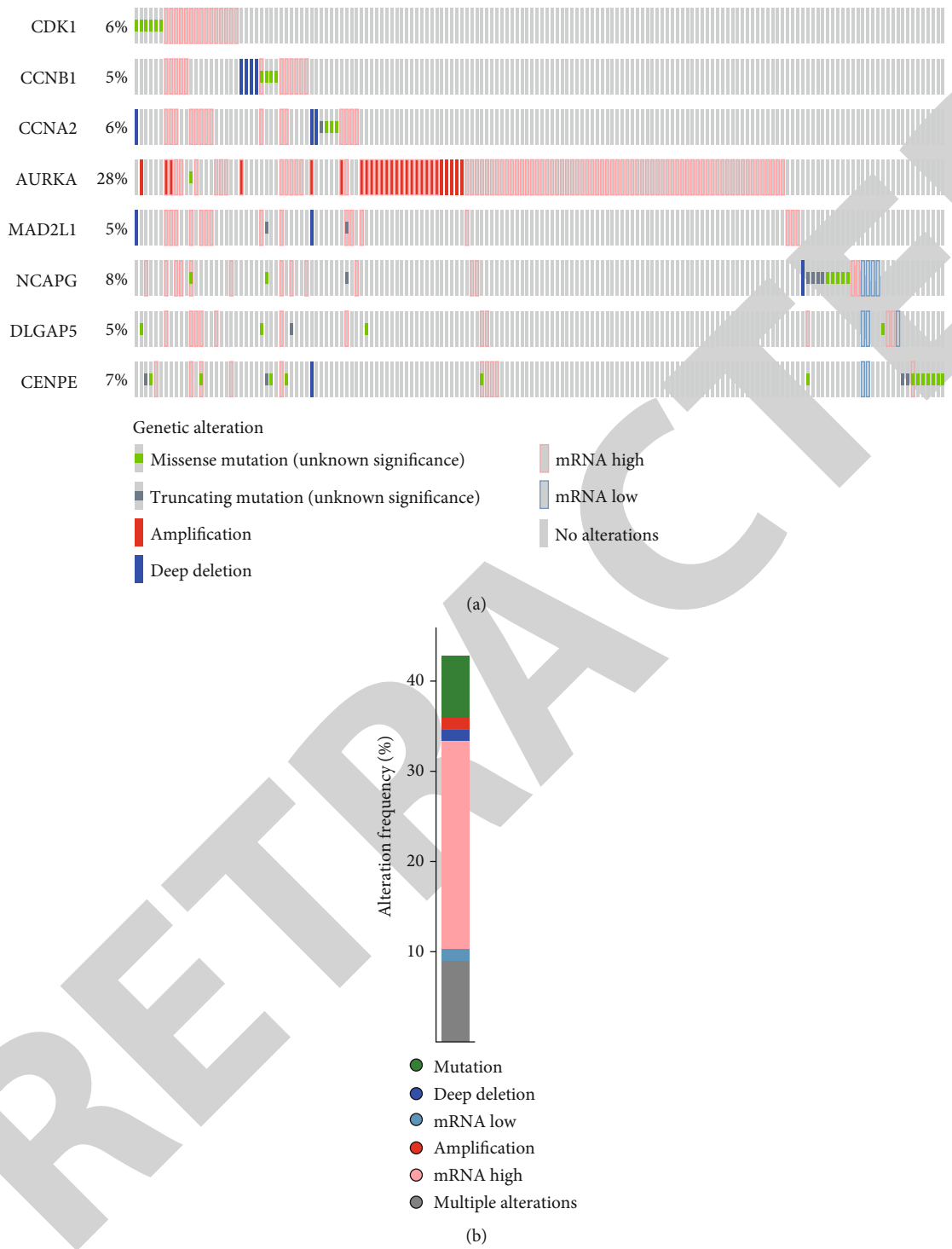
Figure 8: ROC curves analysis of the five hub genes, *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*. AUCs were used to assess the five hub genes, and the results showed high diagnostic accuracy.

previous studies had shown that *CCNA2* was tumour-promoting and associated with advanced tumour stage and tumour development [52, 53]. This was inconsistent with our results, which may be due to the heterogeneity of the sample. Besides, high expression of *DLGAP5* was associated with poor prognosis in well differentiated colon cancer, whereas the prognosis was better in some molecular subtypes of colon cancer, such as patients with a stem cell gene signature [54] and Budinska subtypes A (surface crypt-like) [55]. In our study, *AURKA* exhibited favourable prognostic effects. Interestingly, *AURKA* was upregulated across cancer types, but was only positively associated with prognosis in colon cancer patients [56]. Current studies supported that *AURKA* was associated with the development of colorectal cancer by causing genomic instability [57], but high expression of *AURKA* in colon cancer enhanced the chemotherapy sensitivity of platinum drugs by inhibiting the expression of TP53-regulated DNA damage response genes, which may explain the corresponding better prognosis [56]. However, it has also been reported that high expression of *AURKA* is associated with poor prognosis in colon cancer patients with liver metastasis [58]. Therefore, there was still controversy, and further exploration was needed. *NCAPG and CENPE* have also been reported to play a role in various types of cancer [59, 60], but the underlying mechanisms behind the observed changes in prognosis remain unknown. In summary, these 6 hub genes were significantly associated with the prognosis of colon cancer and may serve as potential prognostic markers as well as therapeutic targets, but further studies were needed to explain and verify their underlying mechanisms.

For early COAD diagnosis, we identified *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*, which were closely related to clinicopathological parameters. *CDK1* plays a key role in the regulation of eukaryotic cell cycle and is essential for G1/S and G2/M transition of eukaryotic cell cycle [61]. Many biological experiments have demonstrated that *CDK1* is highly expressed in colon cancer cells [62, 63] and participates in apoptosis. *CDK1* may act as a potential diagnostic and therapeutic target in view of its extensive involvement in the regulation of colorectal cancer development and progression [62]. *CCNB1* and *CCNA2* are closely related to mitosis. In addition to colon cancer, they have also been found to be highly expressed in pancreatic cancer [64], breast cancer [65], lung cancer [66], and many other cancers, suggesting their potential diagnostic value. *MAD2L1* was highly expressed in active proliferating colon cancer cells, and its expression level gradually increased with the stage of colon cancer [67]. *DLGAP5* was involved in cell proliferation (ClueGO analysis: mitotic chromosome movement towards spindle pole) which was highly expressed in colon cancer cells [54, 68]. One study showed that *DLGAP5* was overexpressed in 293 T cells, resulting in excessive cell proliferation, which may play a potential role in carcinogenesis [69]. In summary, our results showed that both the mRNA and protein expressions of these five hub genes were higher in tumour tissue than in normal tissue, which indicated that the hub genes may be closely related to COAD progression and the possibility of five gene biomarkers in the diagnosis of CRC. Previous studies observed that the expression of these genes was correlated with tumour size and stage [52, 54, 70]. In our study, we found that mRNA expression of the five hub genes was significantly related to mild clinical pathological parameters, so these genes may play an important role in the early diagnosis of colon cancer. In addition, AUCs of these five genes were all greater than 0.9 in ROC curve analysis, which further verified the favourable diagnostic accuracy of these five genes. The relationship between these genes and COAD has not yet been fully determined, but our data indicate that the increased expression in early COAD stages may provide an indicator for early diagnosis. At present, machine learning and deep learning are widely used in disease diagnosis [71, 72]. Deep learning, with its ability to process large-scale data, is a powerful solution for tissue classification and segmentation of histopathological images of colon cancer and other diseases [73, 74].

We finally performed alteration analysis of eight hub genes which showed significant effects on survival analysis, including *CDK1*, *CCNB1*, *CCNA2*, *AURKA*, *MAD2L1*, *NCAPG*, *DLGAP5*, and *CENPE*. The result showed that more than 40% of the patient tumours analysed had at least one hub gene alteration. *AURKA* was the most frequently altered (28%) of the 8 hub genes. The protein encoded by this gene is a cell cycle-regulated kinase that appears to be involved in spindle assembly, cytokinesis, centrosome maturation, and separation [75]. In our study, *AURKA* exhibited favourable effects on both OS and DFS. Previous studies showed that

RETRACTED



RETRACTED

Figure 9: Alterations of the eight hub genes analysed by cBioPortal. (a) OncoPrint of genetic alterations in 378 COAD cases. (b) Alteration frequency of eight hub genes. Gene expression was altered in 42.86% of 378 cases. Abbreviations: COAD: colon adenocarcinoma.

AURKA was frequently upregulated and correlated with prognosis in several types of cancers, which may reveal an important role in human cancer [76, 77].

There were some limitations in this study. First, all the data analysed in our study was retrieved from online databases. Thus, further studies with larger sample sizes and biological experiments were required to validate our findings. Our future research will focus on experimental verification of these results. Second, we did not explore the underlying mechanisms of hub genes in COAD. Future studies should investigate the detailed mechanism between hub genes and COAD.

In conclusion, our study identified and analysed DEGs and 20 core genes associated with COAD, which might deepen the understanding of carcinogenesis and provide indicators for prognosis and early diagnosis of the disease.

Data Availability

All The data used to support the findings of this study are available online.

Disclosure

A preprint has previously been published [Xu, S et al. 2021] [78].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by grant from the Livelihood Science and Technology project of Liaoning Province (Grant no. 2021JH2/10300053).

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global patterns and trends in colorectal cancer incidence and mortality," *Gut*, vol. 66, no. 4, pp. 683–691, 2017.

[3] K. D. Miller, L. Nogueira, A. B. Mariotto et al., "Cancer treatment and survivorship statistics, 2019," *CA: a Cancer Journal for Clinicians*, vol. 69, no. 5, pp. 363–385, 2019.

[4] I. A. Issa and M. Noureddine, "Colorectal cancer screening: an updated review of the available options," *World Journal of Gastroenterology*, vol. 23, no. 28, pp. 5086–5096, 2017.

[5] Y. S. Kim, N. Kim, S. H. Kim et al., "The efficacy of intravenous contrast-enhanced 16-raw multidetector CT colonography for detecting patients with colorectal polyps in an asymptomatic population in Korea," *Journal of Clinical Gastroenterology*, vol. 42, no. 7, pp. 791–798, 2008.

[6] S. C. Ng, J. Y. L. Ching, V. Chan et al., "Diagnostic accuracy of faecal immunochemical test for screening individuals with a family history of colorectal cancer," *Alimentary Pharmacology & Therapeutics*, vol. 38, no. 7, pp. 835–841, 2013.

[7] P. Favoriti, G. Carbone, M. Greco, F. Pirozzi, R. E. M. Pirozzi, and F. Corcione, "Worldwide burden of colorectal cancer: a review," *Updates in Surgery*, vol. 68, no. 1, pp. 7–11, 2016.

[8] V. Das, J. Kalita, and M. Pal, "Predictive and prognostic biomarkers in colorectal cancer: a systematic review of recent advances and challenges," *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, vol. 87, 2017.

[9] G. Weiss, A. Schlegel, D. Kottwitz, T. König, and R. Tetzner, "Validation of the *SHOX*2 / *PTGER*4 DNA methylation marker panel for plasma- based discrimination between patients with malignant and nonmalignant lung disease," *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, vol. 12, no. 1, pp. 77–84, 2017.

[10] Y. Ishibashi, H. Ohtsu, M. Ikemura et al., "Serum TFF1 and TFF3 but not TFF2 are higher in women with breast cancer than in women without breast cancer," *Scientific Reports*, vol. 7, no. 1, p. 4846, 2017.

[11] S. Ning, W. Wei, J. Li et al., "Clinical significance and diagnostic capacity of serum TK1, CEA, CA 19-9 and CA 72-4 levels in gastric and colorectal cancer patients," *Journal of Cancer*, vol. 9, no. 3, pp. 494–501, 2018.

[12] T. Gangadhar and R. L. Schilsky, "Molecular markers to individualize adjuvant therapy for colon cancer," *Nature Reviews. Clinical Oncology*, vol. 7, no. 6, pp. 318–325, 2010.

[13] V. Kulasingam and E. P. Diamandis, "Strategies for discovering novel cancer biomarkers through utilization of emerging technologies," *Nature Clinical Practice. Oncology*, vol. 5, no. 10, pp. 588–599, 2008.

[14] C. C. Huang, M. Du, and L. Wang, "Bioinformatics analysis for circulating cell-free DNA in cancer," *Cancers*, vol. 11, no. 6, p. 805, 2019.

[15] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets–update," *Nucleic Acids Research*, vol. 41, no. Database issue, pp. D991–D995, 2013.

[16] S. Uribe-Lewis, R. Stark, T. Carroll et al., "5-Hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer," *Genome biology*, vol. 16, p. 69, 2015.

[17] E. I. Vlachavas, E. Pilalis, O. Papadodima et al., "Radiogenomic analysis of F-18-fluorodeoxyglucose positron emission tomography and gene expression data elucidates the epidemiological complexity of colorectal cancer landscape," *Computational and Structural Biotechnology Journal*, vol. 17, pp. 177–185, 2019.

[18] X. Sole, M. Crous-Bou, D. Cordero et al., "Discovery and validation of new potential biomarkers for early detection of colon cancer," *PLoS One*, vol. 9, no. 9, p. e106748, 2014.

[19] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, p. 1523, 2019.

[20] D. Szklarczyk, A. L. Gable, D. Lyon et al., "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Research*, vol. 47, no. D1, pp. D607–D613, 2019.

[21] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.

[22] G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC Bioinformatics*, vol. 4, no. 1, p. 2, 2003.

[23] C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, and C. Y. Lin, "cytoHubba: identifying hub objects and sub-networks from complex interactome," *BMC Systems Biology*, vol. 8, Suppl 4, p. S11, 2014.

[24] G. Bindea, B. Mlecnik, H. Hackl et al., "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks," *Bioinformatics*, vol. 25, no. 8, pp. 1091–1093, 2009.

RETRACTED

[25] G. Bindea, J. Galon, and B. Mlecnik, "CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data," *Bioinformatics*, vol. 29, no. 5, pp. 661–663, 2013.

[26] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–W102, 2017.

[27] D. S. Chandrashekar, B. Bashel, S. A. H. Balasubramanya et al., "UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses," *Neoplasia*, vol. 19, no. 8, pp. 649–658, 2017.

[28] M. Uhlen, L. Fagerberg, B. M. Hallström et al., "Proteomics. Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, p. 1260419, 2015.

[29] Y. Hu, J. Gaedcke, G. Emons et al., "Colorectal cancer susceptibility loci as predictive markers of rectal cancer prognosis after surgery," *Genes, Chromosomes & Cancer*, vol. 57, no. 3, pp. 140–149, 2018.

[30] E. Cerami, J. Gao, U. Dogrusoz et al., "The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data," *Cancer Discovery*, vol. 2, no. 5, pp. 401–404, 2012.

[31] M. M. Fidler, I. Soerjomataram, and F. Bray, "A global view on cancer incidence and national levels of the human development index," *International Journal of Cancer*, vol. 139, no. 11, pp. 2436–2446, 2016.

[32] H. Brenner, L. Altenhofen, C. Stock, and M. Hoffmeister, "Prevention, early detection, and overdiagnosis of colorectal cancer within 10 years of screening colonoscopy in Germany," *Clinical Gastroenterology and Hepatology*, vol. 13, no. 4, pp. 717–723, 2015.

[33] Q. Zhou, F. Zhang, Z. He, and M. Z. Zuo, "E2F2/5/8 serve as potential prognostic biomarkers and targets for human ovarian cancer," *Frontiers in Oncology*, vol. 9, p. 161, 2019.

[34] J. Sun, R. Shi, S. Zhao et al., "Cell division cycle 45 promotes papillary thyroid cancer progression via regulating cell cycle," *Tumour Biology*, vol. 39, no. 5, p. 1010428317705342, 2017.

[35] P. Finetti, N. Cervera, E. Charafe-Jauffret et al., "Sixteen-kinase gene expression identifies luminal breast cancers with poor prognosis," *Cancer Research*, vol. 68, no. 3, pp. 767–776, 2008.

[36] S. Qie and J. A. Diehl, "Cyclin D1, cancer progression, and opportunities in cancer treatment," *Journal of Molecular Medicine*, vol. 94, no. 12, pp. 1313–1326, 2016.

[37] B. M. Moradi, A. Bahrami, M. Khazaei et al., "The prognostic value of cyclin D1 expression in the survival of cancer patients: a *meta* -analysis," *Gene*, vol. 728, p. 144283, 2020.

[38] J. R. Alt, J. L. Cleveland, M. Hannink, and J. A. Diehl, "Phosphorylation-dependent regulation of cyclin D1 nuclear export and cyclin D1-dependent cellular transformation," *Genes & Development*, vol. 14, no. 24, pp. 3102–3114, 2000.

[39] P. Aggarwal, M. D. Lessie, D. I. Lin et al., "Nuclear accumulation of cyclin D1 during S phase inhibits Cul4-dependent Cdt1 proteolysis and triggers p53-dependent DNA rereplication," *Genes & Development*, vol. 21, no. 22, pp. 2908–2922, 2007.

[40] M. C. Casimiro, M. Velasco-Velázquez, C. Aguirre-Alvarado, and R. G. Pestell, "Overview of cyclins D1 function in cancer and the CDK inhibitor landscape: past and present," *Expert Opinion on Investigational Drugs*, vol. 23, no. 3, pp. 295–304, 2014.

[41] J. Sun, Y. Guo, X. Fu et al., "Dendrobium candidum inhibits MCF-7 cells proliferation by inducing cell cycle arrest at G2/M phase and regulating key biomarkers," *Oncotargets and Therapy*, vol. 9, pp. 21–30, 2016.

[42] Z. Lin, L. Xiong, and Q. Lin, "Knockdown of eIF3d inhibits cell proliferation through G2/M phase arrest in non-small cell lung cancer," *Medical Oncology*, vol. 32, no. 7, p. 183, 2015.

[43] Q. Ma, "miR-219-5p suppresses cell proliferation and cell cycle progression in esophageal squamous cell carcinoma by targeting CCNA2," *Cellular & Molecular Biology Letters*, vol. 24, no. 1, p. 4, 2019.

[44] X. Tu, D. Hong, Y. Jiang et al., "FH535 inhibits proliferation and migration of colorectal cancer cells by regulating _CyclinA2_ and _Claudin1_ gene expression," *Gene*, vol. 690, pp. 48–56, 2019.

[45] F. Kruiswijk, C. F. Labuschagne, and K. H. Vousden, "p53 in survival, death and metabolic health: a lifeguard with a licence to kill," *Nature Reviews. Molecular Cell Biology*, vol. 16, no. 7, pp. 393–405, 2015.

[46] S. Shankar and R. K. Srivastava, "Involvement of Bcl-2 family members, phosphatidylinositol 3'-kinase/AKT and mitochondrial p53 in curcumin (diferulolylmethane)-induced apoptosis in prostate cancer," *International Journal of Oncology*, vol. 30, no. 4, pp. 905–918, 2007.

[47] H. H. Lin, H. P. Huang, C. C. Huang, J. H. Chen, and C. J. Wang, "Hibiscus polyphenol-rich extract induces apoptosis in human gastric carcinoma cells via p53 phosphorylation and p38 MAPK/FasL cascade pathway," *Molecular Carcinogenesis*, vol. 43, no. 2, pp. 86–99, 2005.

[48] K. M. Gharpure, S. Pradeep, M. Sans et al., "FABP4 as a key determinant of metastatic potential of ovarian cancer," *Nature Communications*, vol. 9, no. 1, p. 2923, 2018.

[49] J. Nie, J. Zhang, L. Wang et al., "Adipocytes promote cholangiocarcinoma metastasis through fatty acid binding protein 4," *Journal of Experimental & Clinical Cancer Research*, vol. 36, no. 1, p. 183, 2017.

[50] Y. Fang, X. Liang, W. Jiang, J. Li, J. Xu, and X. Cai, "Cyclin b1 suppresses colorectal cancer invasion and metastasis by regulating e-cadherin," *PLoS One*, vol. 10, no. 5, article e0126875, 2015.

[51] Y. Guo, M. Gabola, R. Lattanzio et al., "Cyclin A2 maintains colon homeostasis and is a prognostic factor in colorectal cancer," *The Journal of Clinical Investigation*, vol. 131, no. 4, 2021.

[52] S. Dong, F. Huang, H. Zhang, and Q. Chen, "Overexpression of BUB1B, CCNA2, CDC20, and CDK1 in tumor tissues predicts poor survival in pancreatic ductal adenocarcinoma," *Bioscience Reports*, vol. 39, no. 2, 2019.

[53] L. Brcic, M. Heidinger, A. Z. Sever et al., "Prognostic value of cyclin A2 and B1 expression in lung carcinoids," *Pathology*, vol. 51, no. 5, pp. 481–486, 2019.

[54] V. Branchi, S. A. García, P. Radhakrishnan et al., "Prognostic value of DLGAP5 in colorectal cancer," *International Journal of Colorectal Disease*, vol. 34, no. 8, pp. 1455–1465, 2019.

[55] E. Budinska, V. Popovici, S. Tejpar et al., "Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer," *The Journal of Pathology*, vol. 231, no. 1, pp. 63–76, 2013.

[56] B. Shan, R. Zhao, J. Zhou et al., "AURKA increase the chemosensitivity of colon cancer cells to oxaliplatin by inhibiting the TP53-mediated DNA damage response genes," *BioMed*

RETRACTED

Hindawi
BioMed Research International
Volume 2022, Article ID 4067581, 17 pages
https://doi.org/10.1155/2022/4067581

Review Article

Pooling *Operations in Deep* Learning: From "Invariable" to "Variable"

Zhou Tao,[1,2] **Chang XiaoYu**,[1] **Lu HuiLing**,[3] **Ye XinYu**,[1] **Liu YunCan**,[1] **and Zheng XiaoMin**[4]

[1]*School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China*
[2]*Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China*
[3]*School of Science, Ningxia Medical University, Yinchuan 750004, China*
[4]*Research Institute for Reproductive Medicine and Genetic Diseases, Wuxi Maternity and Child Health Hospital, Wuxi 214002, China*

Correspondence should be addressed to Chang XiaoYu; yu0410yu@163.com and Lu HuiLing; lu_huiling@163.com

Received 6 May 2022; Accepted 27 May 2022; Published 20 June 2022

Academic Editor: Chen Li

Copyright © 2022 Zhou Tao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep learning has become a research hotspot in multimedia, especially in the field of image processing. Pooling operation is an important operation in deep learning. Pooling operation can reduce the feature dimension, the number of parameters, the complexity of computation, and the complexity of time. With the development of deep learning models, pooling operation has made great progress. The main contributions of this paper on pooling operation are as follows: firstly, the steps of the pooling operation are summarized as the pooling domain, pooling kernel, step size, activation value, and response value. Secondly, the expression form of pooling operation is standardized. From the perspective of "invariable" to "variable," this paper analyzes the pooling domain and pooling kernel in the pooling operation. Pooling operation can be classified into four categories: invariable of pooling domain, variable of pooling domain, variable of pooling kernel, and the pooling of invariable "+" variable. Finally, the four types of pooling operation are summarized and discussed with their advantages and disadvantages. There is great significance to the research of pooling operations and the iterative updating of deep learning models.

1. Introduction

Deep learning was proposed by Hinton [1] in 2006. A Convolutional Neural Network (CNN) was proposed by Le et al. [2], which is the first real multilayer structure learning algorithm. Deep learning has been widely used in various fields of multimedia such as text, image, and video, and it has become a research hotspot in the field of images. Convolution operation is the main step of the CNN, but it makes the whole network have a large number of calculations [3]. And it takes a long time for network training. Therefore, the pooling operation is an essential step in CNN.

Pooling operation is a common step in CNN and is also called subsampling operation or downsampling operation. When CNN was constructed, the pooling operation is usually placed behind the convolution layer. From the perspective of lightweight, pooling operation can reduce feature dimensionality, computation complexes, parameter number, and memory consumption and suppress noise. From the perspective of improving image quality, pooling operation can improve the horizontal movement invariance, scale invariance, and rotation invariance of the model. The pooling operation improves the robustness to image bias and rotation. From the perspective of the deep learning model, pooling operation can prevent overfitting. Pooling operation can improve the generalization ability of the model and expand the receptive field. Hence, it is very interesting to discuss pooling operation.

2. The "Variable" and "Invariable" of Pooling Operation

There are five parts in pooling operation, such as pooling domain, pooling kernel, step length, activation value, and response value [4]. The process of the pooling operation is shown in Figure 1. The "invariable" of pooling operation means that there is no change in the shape and size of the pooling domain and the shape, size, and parameter number of the pooling kernel. The "variable" of pooling operation means that there are two types: first is pooling domain changing and second is pooling kernel changing. About pooling domain changing, the shape and size are changed in the pooling domain. About pooling kernel changing, the shape, size, and parameters are changed in the pooling kernel [5].

Let the target image be $F(m, n)$, m is the width of the image, and n is the height of the image. The pooling domain is $f(i, j)$, where $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$, which is the size of the receptive field in the pooling process. The pooling kernel is $w(i, j)$, where $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$, which is the filter of the pooling domain. The pooling step length is the step length of the pooling nucleus every time it moves. The activation value $x_{i,j}$ is the feature value of the i-th row and j-th column in the feature map. The pooling feature value of the i-th row and j-th column is obtained by pooling operation.

In the pooling domain of size $i \times j$, the value in the pooling domain is determined by the feature map. The pooling domain $f(i, j)$ is determined by the pooling kernel $w(i, j)$, and the pooling kernel is a matrix of size $i \times j$. Generally, the pooling kernel has no parameters. The pooling domain $f(i, j)$ can be represented as

$$f(i,j) = \begin{bmatrix} x_{p,q} & x_{p,q+1} & \cdots & x_{p,q+j} \\ x_{p+1,q} & x_{p+1,q+1} & \cdots & x_{p+1,q+j} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p+i,q} & x_{p+i,q+1} & \cdots & x_{p+i,q+j} \end{bmatrix}_{i \times j}. \quad (1)$$

It can be concluded that the response feature map $g(i, j)$ is determined by the pooling kernel $w(i, j)$ in the feature map F of size $m \times n$. It is expressed as

$$g(i,j) = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,j} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i,1} & y_{i,2} & \cdots & y_{i,j} \end{bmatrix}_{i \times j}, \quad (2)$$

$$y_{i,j} = \beta \text{down}(x(i,j)) + b,$$

where down$(.)$ represents the downsampling operation, which is the pooling kernel $w(i, j)$ for the pooling operation of values in the pooling domain $f(i, j)$. β is the pooling coefficient. $y_{i,j}$ is the response value of the pooling domain. The response value is the value of the response feature map in the i-th row and j-th column. And b is the bias value of pooling operation.

The development of pooling operation is a process from "invariable" to "variable," as shown in Figure 2. The most commonly used pooling operation is max pooling operation and average pooling operation. Max pooling operation and average pooling operation are calculated simply. They have few parameters and work well on most tasks. So far, these pooling operations are still the most commonly used pooling methods. In addition, max pooling operation and average pooling operation cannot achieve good results on specific tasks. Many scholars have further studied the pooling operation for different problems. Firstly, change the value of pooling selection. Secondly, change the pooling domain, which is including the change of shape and size. Thirdly, change the pooling kernel, which is including the change in shape, size, and parameters. Finally, there are various pooling methods that are combined, which have parallel methods and serial methods. The parallel method is the most commonly used.

In this paper, the "variable" and "invariable" of the pooling operation are divided into four categories: invariable of pooling domain, variable of pooling domain, variable of pooling kernel, and the invariable "+" variable of pooling. The overall framework is shown in Figure 3. The invariable of pooling domain means that there is no change in the shape and size of the pooling domain. The main methods are max pooling operation, maximum two mean pooling operation, K-max pooling operation, average pooling operation, median pooling operation, sum pooling operation, overlapping pooling operation, matrix 2-norm pooling operation, moment pooling operation, covariance pooling operation, dynamic adaptive pooling operation, and nested invariance pooling operation. The variable of pooling domain means that there is change in the shape and size of the pooling domain. The main methods are region of interest pooling method, fractional max pooling method, strip pooling method, chunk-max pooling method, and multiscale orderless pooling method. The variable of pooling kernel means that there is change in the parameters, selection method, and size of the pooling kernel. The main methods are generalized max pooling operation, parameter pooling operation, LP pooling operation, spatial attention pooling operation, stochastic pooling operation, spatial pyramid pooling operation, and rank-based pooling (rank-based average pooling operation, rank-based weighted pooling operation, and rank-based stochastic pooling operation) operation. The invariable "+" variable of pooling means that there are two or more pooling methods in pooling methods, including variable pooling operation and invariable pooling operation. The main methods are intermediate value pooling operation, mixed pooling operation, spatial pyramid and global average hybrid pooling operation, multiscale pooling operation, and scalable overlapping slide pooling operation.

3. Invariable of Pooling Domain

Invariable of pooling domain means that there is no change in the shape and size of the pooling domain. And there is

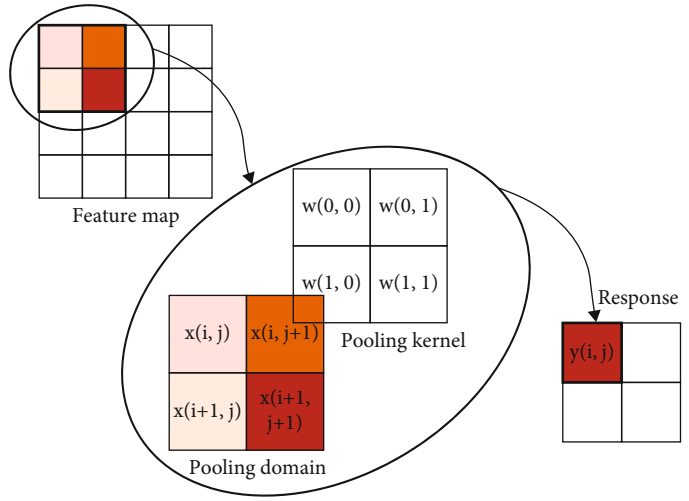


Figure 1: The process of pooling operation.

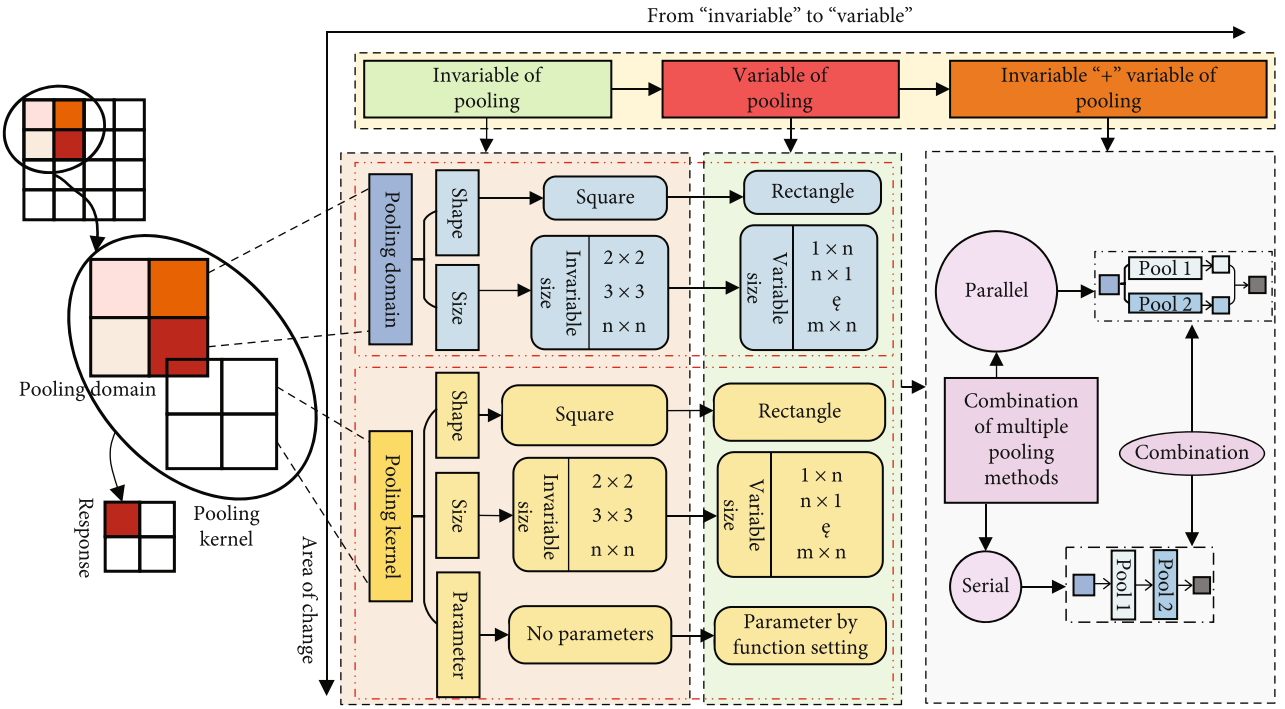


Figure 2: "Invariable" to "variable" of pooling operation.

change in the shape, size, and parameters of the pooling kernel. In this paper, the variable of pooling domain is divided into two methods: basic pooling operation and improved pooling operation. The basic pooling operation has the advantages of simple calculation process and wide application range. But the utilization rate is low in the feature map. The improved pooling operation uses different mapping rules to select the eigenvalues in the pooling domain. So the improved pooling operation can improve the utilization rate of the feature maps. But the calculation process is more complicated. The development process of invariable of pooling domain is shown in Figure 4.

3.1. Basic Pooling Operation. The basic pooling operation during the pooling operation is that the shape and size of the pooling domain are unchanged, and the shape, size, and parameters of the pooling kernel are unchanged. Select the eigenvalues with different numbers in the pooling domain, including max pooling operation, maximum two-mean pooling operation, K-max pooling operation, average pooling operation, median pooling operation, sum pooling operation, and overlapping pooling operation.

Firstly, max pooling operation selects the largest activation value in the pooling domain as the output response value. Other activation values are discarded at the same time.

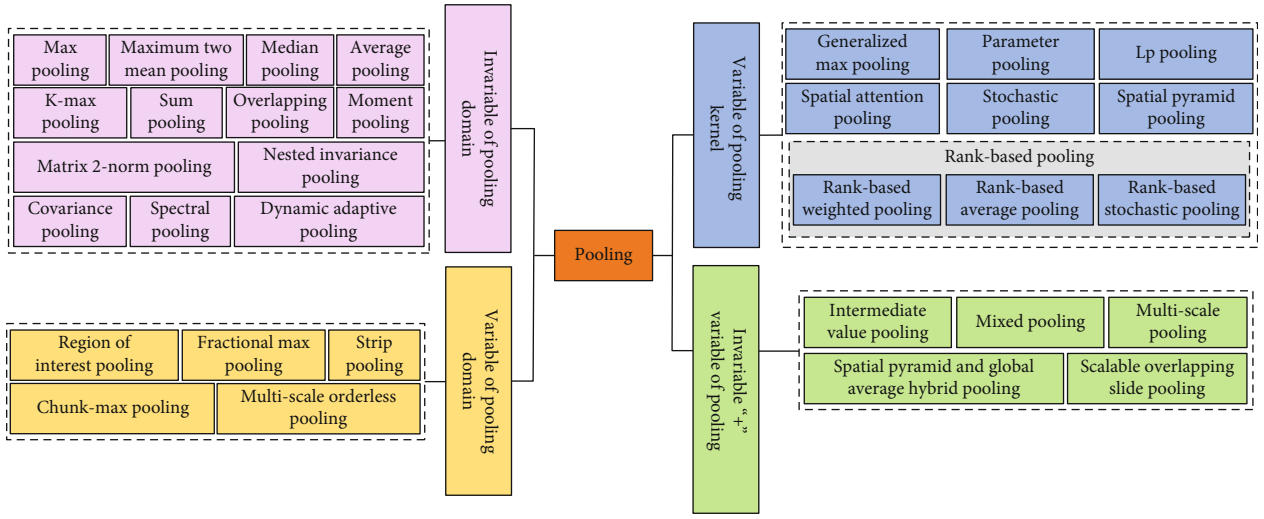


Figure 3: Summary of pooling methods.

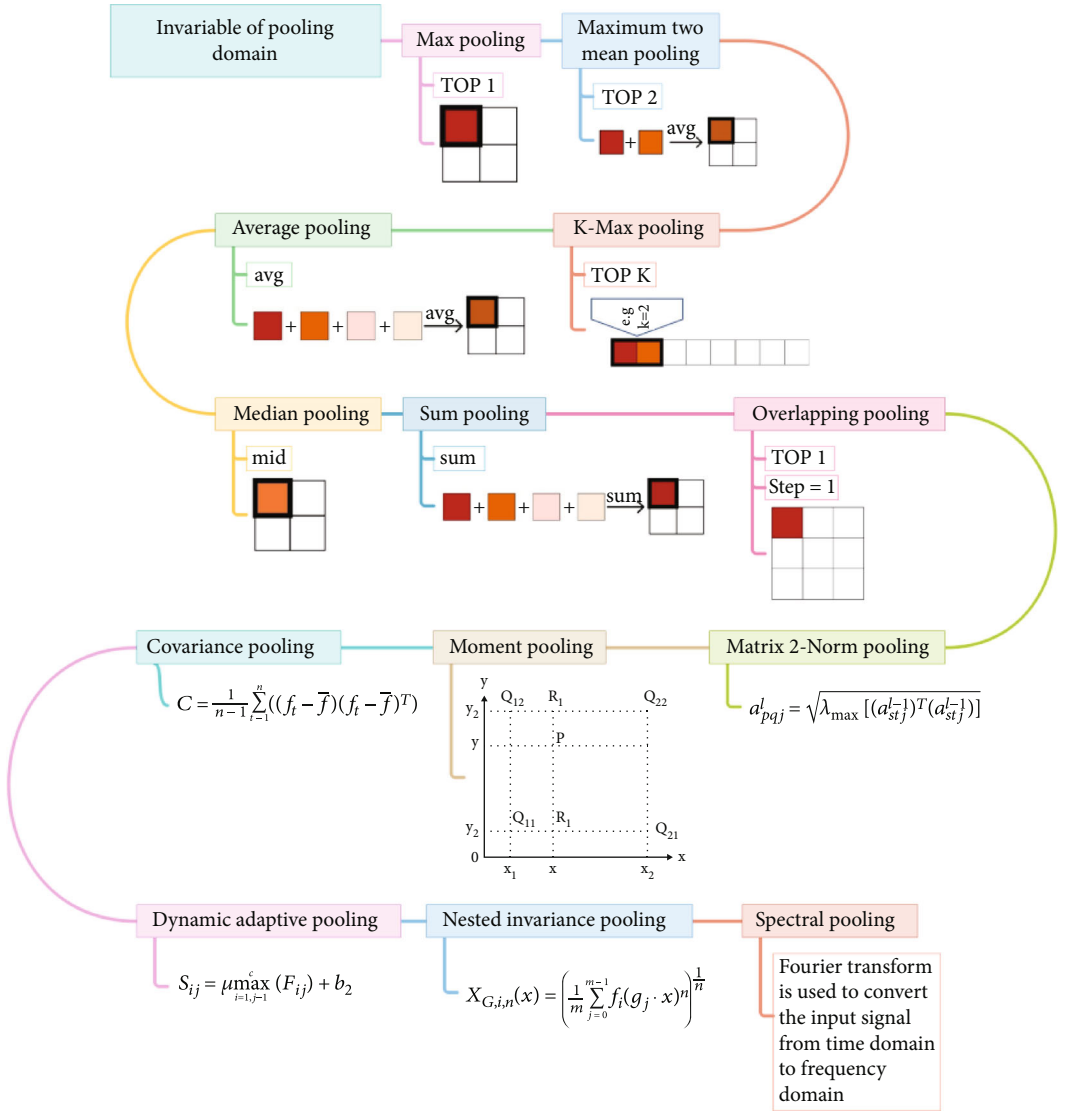


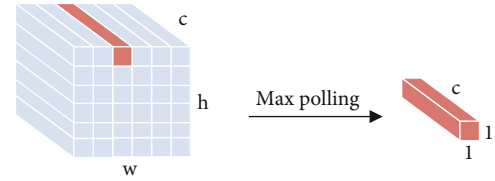
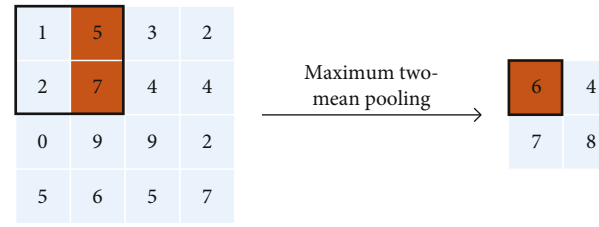
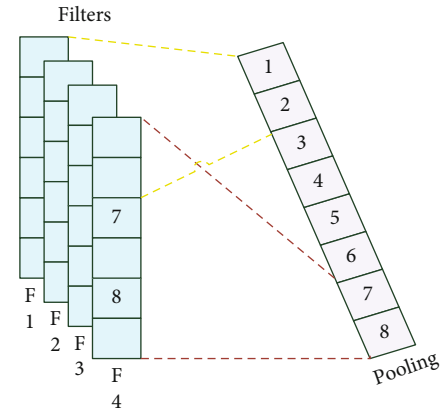
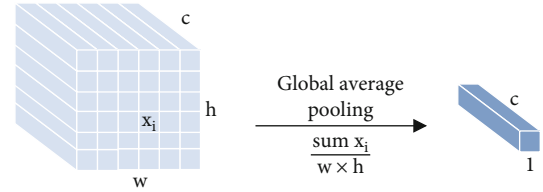
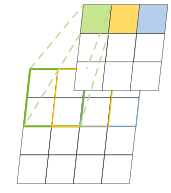
Figure 4: Invariable of pooling domain.

The advantage of max pooling operation is that salient features can be extracted, and the contour information of the image can be learned. However, max pooling operation loses too much features and information. The global max pooling operation is a case that max pooling operation is generalized to the global. The difference is that the size of the pooling domain is the same as the size of the entire feature map. Secondly, maximum two-mean pooling operation selects the two largest activation values in the pooling domain and averages them. Use the averaged value as the response value. The maximum two-mean pooling operation improves the disadvantage of ignoring features with larger influencing factors of max pooling. But there is some information missing. Thirdly, K-max pooling operation extracts the top K values in the pooling domain as the response value and preserves the order of these values. K-max pooling operation retains more information than max pooling, which can express the strength of a certain type of feature. K-max pooling operation retains position information, but it is the relative order between features, not absolute position information. Fourthly, average pooling operation means that the average of activation values in the pooling domain and the average value are used as the response value. Average pooling considers the global information, and there are no information loss and reducing overfitting. But it makes the features smooth and cannot extract the information of prominent features. The global average pooling operation is a case that average pooling operation is generalized to the global. The difference is that the size of the global average pooling domain is the same as the size of the entire feature map. Fifthly, median pooling operation takes the median of all activation values in the pooling domain as the response value. Median pooling operation has the characteristics of learning edge and texture structures and has a strong ability to resist noise, which can improve performance. Sixthly, sum pooling operation takes the sum value of all activation values in the pooling domain as the response value. Sum pooling operation can be applied to any encoding, and the impact of frequent descriptors is greater than that of rare descriptors. Seventhly, overlapping pooling [6] operation selects the largest activation value in the pooling domain as the output response value. The pooling domain of overlapping pooling operation has overlapping parts because its stride is 1. The overlapping pooling operation augments the output into multiple levels of smaller features. Overlapping pooling operation uses sparse coding to fuse multilevel features, which reduces the feature dimension of the output and improves the network performance. Table 1 shows the summary of basic pooling operation.

3.2. Improved Pooling Operation. The improved pooling operation means that there are no change in the shape and size of the pooling domain and change in the shape, size, and parameters of the pooling kernel. But the difference is the way to get the response value of eigenvalues in the pooling domain using the mapping rule. Improved pooling operation includes matrix 2-norm pooling operation, moment pooling operation, covariance pooling operation, dynamic adaptive pooling operation, nested invariance pooling operation, and spectral pooling operation.

Firstly, matrix2-norm pooling [7] operation makes the eigenvalues of feature map as the coefficient matrix of the singular equation system. Take the square root of the largest singular value obtained from this matrix as the response value of the pooling domain. When the image is geometrically distorted such as rotation, translation, and scaling, there are little changes in the singular value of the image matrix. This means that the energy information of the image has high stability, so the singular value of the matrix can make the geometric distortion of the image highly invariant. Secondly, moment pooling [8] operation calculates the central moment $c(i, j)$ of the pooling domain using the gray moment. Calculates the four neighborhoods of the central moment c. Calculate the probability of each neighborhood using the distances between the four neighborhoods $x(i, j)$, $x(i+1, j)$, $x(i, j+1)$, $x(i+1, j+1)$, and center $c(i, j)$. The closer the distance is, the greater the probability of being selected, which is conducive to the probability of selecting the response value. Because the randomness of moment pooling operation makes each selection different, it can effectively prevent oversuppression. Moment pooling operation randomly generates 2 probabilities. The computational complexity of the moment pooling operation is much lower than bilinear interpolation and close to the nearest neighbor interpolation. Thirdly, covariance pooling [9] operation calculates the covariance matrix using the eigenvalue matrix of the feature map and then adjusts the matrix to what you need. For example, the matrix is normalized to get the response values. Covariance pooling operation [10] belongs to the second-order pooling method, which can capture more information about the feature map than most first-order pooling methods. Fourthly, dynamic adaptive pooling operation [11] optimizes max pooling operation with a pooling factor μ. Dynamic adaptive pooling operation builds a mathematical model for function simulation, and other parameters are the same as max pooling operation. The pooling factor will not lose accuracy when it deals with pooling domains with obvious max characteristics. It can weaken the influence of max pooling operation when it deals with other pooling domains. Therefore, the CNN can extract more accurate features under different iterations. Fifthly, nested invariance pooling [12] operation changes the output of feature maps in the pooling domain using invariant transformation, which is the response feature map. It can be extended to any transform set. Lou et al. [13] obtain global descriptors by concatenating all feature maps. They construct nested invariance pooling operation from existing local translation invariant feature maps. Zhang et al. [14] showed that various types of transformation invariants can be linked. Sixthly, spectral pooling [15, 16] operation is performed for dimensionality reduction by projecting onto the frequency basis set and then truncating the representation. It is a pooling method based on fast Fourier transform, which keeps the same number of parameters and retains more information. Spectral pooling operation can be regarded as a denoising method that allows specifying any arbitrary output feature dimension. The truncation of frequency representation

Table 1: Summary of basic pooling operation.

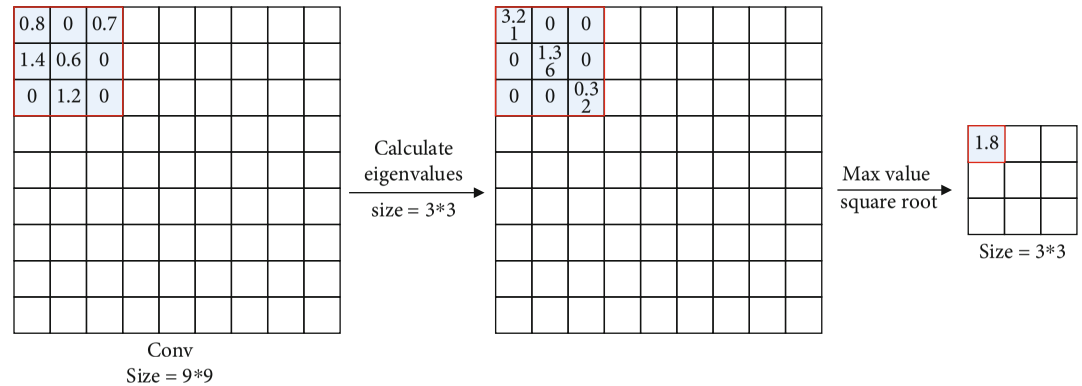
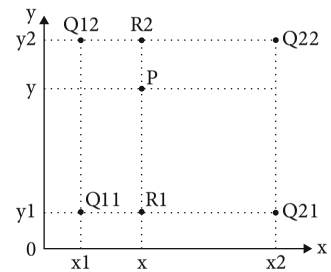
Pooling method	Characteristic	Sketch map
Max pooling	The max pooling operation is simple, but the features with strong influencing factors are ignored and lost information	
Maximum two-mean pooling	Maximum two-mean pooling operation improves the disadvantage of max pooling to a certain extent, which is ignoring features with larger influencing factors. But there is some problem of loss information	
K-max pooling	K-max pooling operation preserves more information than max pooling and preserves the relative order of eigenvalues, but there is no absolute position information	
Average pooling	Average pooling operation takes global information into account, no information is lost, and overfitting is reduced. However, the features tend to be smooth, and the information of prominent features cannot be extracted	
Median pooling	Median pooling operation can learn the characteristics of edge and texture structure and has a strong antinoise ability	$Y(F) = \text{Median}(S_1, S_2, \cdots, S_N)$ $S_i = \text{Median}\left(f_{i1}, f_{i1}, \cdots, f_{im_i}\right)$ $i = 1, 2, \cdots N$
Sum pooling	Sum pooling operation considers global information. There is no information loss, but it is vulnerable to extreme information	—
Overlapping pooling [6]	Overlapping pooling operation reduces the characteristic dimension of the output and reduces overfitting, but there are information that will be lost	

corresponds to the reduction in resolution. So spectral pooling operation can be supplemented with random regularization in the form of random resolution. Table 2 shows the summary of improved pooling operation.

4. Variable of Pooling Domain

Variable of pooling domain means that there is change in the shape and size of the pooling kernel. And there is no

Table 2: Summary of improved pooling operation.

Pooling method	Characteristic	Sketch map
Matrix2-norm pooling [7]	Matrix 2-norm pooling operation takes the energy of the image as the information transmitted to the next layer network to make geometric distortion of the image is highly invariable	
Moment pooling [8]	The randomness of moment pooling operation makes each choice different, to prevent over inhibition	
Covariance pooling [9]	Covariance pooling operation can capture more information on the feature map	$C = (1/n-1)\sum_{i=1}^{n}\left(\left(f_i - f_{_}\right)\left(f_i\bar{f}_{_}\right)^T\right)$
Dynamic adaptive pooling [11]	Dynamic adaptive pooling operation can adaptively adjust the weight of pooling and extract more accurate features	$y_{ij} = \mu \max_{i=1,j=1}^{c}\left(x_{ij}\right) + b_2$ $\mu = \rho\left(\lambda\left(x_{\max} - \lambda\right)/{x_{\max}}^2\right) + \theta\rho = c/1 + \left(n_{epo} - 1\right)c^{n_{epo}^2+1}$
Nested invariance pooling [12]	Nested invariance pooling operation can be extended to any arbitrary transformation set, allowing any arbitrary output feature dimension to be specified	$Y_{G,i,n}(x) = \left((1/m)\sum_{j=0}^{m-1} f_i\left(g_j.x\right)^n\right)^{1/n}$
Spectral pooling [15]	Spectral pooling operation has higher efficiency and lower cost	—

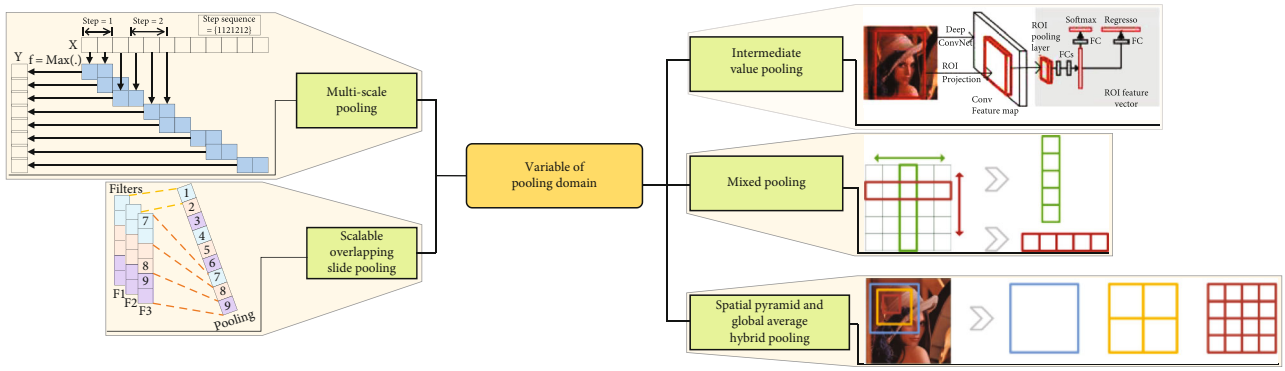


Figure 5: Variable of pooling domain.

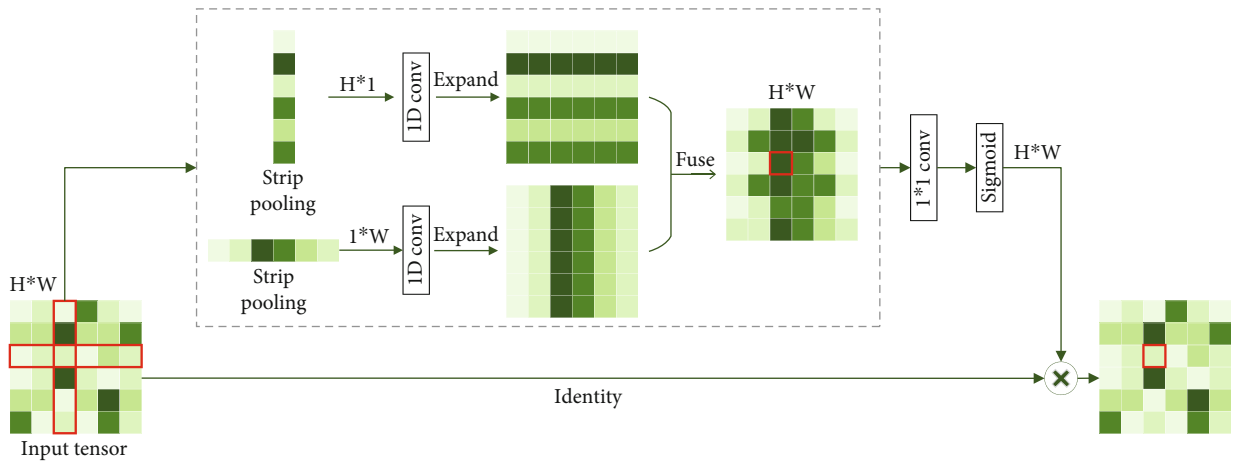


Figure 6: Schematic diagram of strip pooling [21].

change in the parameters of the pooling kernel. The shape and size of the pooling kernel are determined by the pooling domain. Variable of pooling domain can modify the size and shape of the pooling domain as needed, which is more flexible. But it leads to susceptibility to more irrelevant pixels when the pooling domain changes. Variable of pooling domain includes region of interest pooling operation, fractional max pooling operation, strip pooling operation, chunk-max pooling operation, and multiscale orderless pooling operation, as shown in Figure 5.

Firstly, region of interest pooling (ROI pooling) [17] is pooling just for the region of interest. There is no fix in the shape and size of the pooling domain which is depending on the region of interest. The pooling domain is the region of interest in which a max pooling operation in the pooling domain is done. ROI pooling operation was used in fast RCNN [17, 18] first. Its advantage is that the location area and shape size of the pooling domain can be determined due to the problem, and the size of the output feature map can be determined. And just the region of interest in the feature map needs to be pooled, which greatly reduces the amount of computation. In the calculation process, the position of the region of interest is easily selected due to the quantization operation. Secondly, fractional max pooling [19] operation divides the feature map of input size $b \times b$ into $a \times a$ blocks. These feature blocks of the $a \times a$ are $a \times a$ pooling domains, and then, select the max value of each pooling domain to obtain the response value. Fractional max pooling operation allows the pooling domain to be non-integer values. It avoids the problem of limited performance due to disjoint pooling kernels when the pooling domain is integer. Yue et al. [20] make random number sequences or pseudorandom sequences as the size of each pooling domain, and the integer sequence generates 1 or 2. It reduces the overfitting of the dataset due to the randomness of the sequence. Thirdly, strip pooling [21] operation performs pooling operations along the horizontal or vertical dimension using a strip pooling domain. The pooling domain of strip pooling operation is an $h \times 1$ or $1 \times w$ strip domain, and average values in the pooling domain rowwise or columnwise are shown in Figure 6. Strip pooling alleviates the problem of merging irrelevant regions in the process of processing objects with irregular shapes about average pooling. It can be easily embedded into any building block without training from scratch because the strip pooling module is lightweight. Strip pooling has long and narrow pooling field shapes. It is easy to establish long-range dependencies between discretely distributed regions and encode the strip-shaped regions. Fourthly, chunk-max pooling [22] operation segments the feature vector on the feature map

as pooling domain. Chunk-max pooling operation performs max pooling operation for each pooling domain, obtaining multiple response values. This method of segmenting eigenvectors can be set in advance or dynamically divided according to eigenvalues. Static segmentation means that the number of paragraphs can be set in advance. Dynamic segmentation means that the boundary positions between chunks can be dynamically divided according to different inputs. Chunk-max pooling operation retains relatively coarse-grained fuzzy location information. Feature intensity can be captured if a strong feature occurs multiple times. Fifthly, multiscale orderless pooling [23] operation extracts deep activation features through three pooling domains of different sizes. Then, the local responses are aggregated on a finer scale by local aggregation vector (VLAD) coding. Finally, the original global deep activations are connected with the VLAD feature to form response value. The features obtained after activation of global CNNs lack geometric invariance, which includes rotation invariance and translation invariance. Multiscale orderless pooling operation is proposed to solve this problem. It limits their robustness to the classification and matching of variable scenes. Table 3 shows the summary of variable of pooling domain.

5. Variable of Pooling Kernel

Variable of pooling kernel means that there is change in the shape, size, and parameters of the pooling kernel. Variable of pooling kernel can get a better solution for the problem to be solved, and the pooling changes the pooling kernel. In this paper, the variable of pooling kernel is divided into four categories: pooling kernel parameter invariable, random selection, rank-based pooling, and pooling kernel parameter variable, as shown in Figure 7. The pooling kernel parameter invariable is to set the parameters of the pooling kernel. The response value is obtained by multiplying the parameters in the pooling kernel with the activation value in the pooling domain. Random selection randomly selects values in the pooling domain as the output response value. Rank-based pooling sorts all the values in the pooling domain and then performs averages, weights, or random pooling on the ordered activation values. The pooling kernel parameter variable is the improvement of the pooling operation in front of the fully connected layer. The pooling kernel parameter variable uses three pooling kernels of different sizes to generate a fixed-length feature vector for the max pooling operation. The obtained result is passed to the fully connected layer.

Firstly, generalized max pooling [24] operation is a process of obtaining weights by solving the optimization problem. The weights are obtained according to the similarity between each element and other elements. Rare pixels with lower similarity will have larger weights. Finally, each element will be weighted and summed as the output. Generalized max pooling operation achieves the same effect as max pooling operation. It is also suitable for places other than the visual word package (BOV), especially for the Fisher vector. Secondly, parameter pooling [25] operation uses the operation rules of the pooling kernel parameters to obtain the parameters of the pooling kernel. The parameters of the pooling kernel are matrix multiplied by the activation value to obtain the response value. Generally, the pooling kernel has no parameters. Parameter pooling operation allows the model to choose the pooling method autonomously because it has a few parameters. Parameter pooling operation preserves feature information as much as possible. It reduces network errors, and the network performance is improved. Thirdly, LP pooling [26] operation uses a Gaussian kernel to weight the pixel values of the input feature map. At this time, the Gaussian kernel is the pooling kernel of the pooling operation, and the weighted value is the response value. The optimal pooling method for most problems is neither average pooling operation nor max pooling operation but a certain type between them. When $p = 2$, it is L2 pooling operation. There are good results that can be achieved in most image classification problems. Fourthly, spatial attention pooling [27] operation generates an attention parameter matrix for the entire feature map. The parameter matrix acts as a pooling kernel to perform a global pooling operation on the feature map. The value on the attention parameter matrix is multiplied by the values of the feature map to obtain a new matrix as the response result of spatial attention pooling operation. Spatial attention pooling operation is able to highlight regions that require attention. Fifthly, rank-based pooling [28] operation sorts the eigenvalues in the pooling domain from large to small. It performs average pooling operation, weighted pooling operation, and stochastic pooling operation on the sorted values, as shown in Figure 8. The traditional pooling methods operate directly on the values in the feature map. There are many possible values in a pooling domain, but the ranking of eigenvalues is relatively stable. Therefore, rank-based pooling operation can avoid the influence caused by the change of the value. Rank-based average pooling [29] (RAP) operation sorts the values of the feature maps from large to small and selects the t with the largest value to average the response value. RAP can alleviate the problem of information loss from max pooling operation and discriminative information loss from average pooling operation. Rank-based weighted pooling [30] (RWP) operation assigns weight p_r according to the size of the value. The value as the response value is obtained by multiplying the feature map matrix and the weight matrix. It can be assigned the larger weights for the higher activation value. The value ordering method of RWP can be a linear function or a nonlinear function. RWP assigns a reasonable weight value to each activation value in the pooling domain, which can improve the performance significantly. Rank-based stochastic pooling [31] (RSP) computes the probability p_r using exponential ranking on the resulting ranking values. It selects activation values by sampling from a multinomial distribution based on p_r. It is proposed to solve the problem that random pooling is just suitable for nonnegative values, and feature probability calculated by standardized eigenvalues is out of control, which is easy to lead to overfitting.

Sixthly, stochastic pooling [29] operation obtains the probability according to the value. The probability is used to select randomly, and the selected activation value is output. The higher the probability value, the greater the

Table 3: Summary of variable of pooling domain.

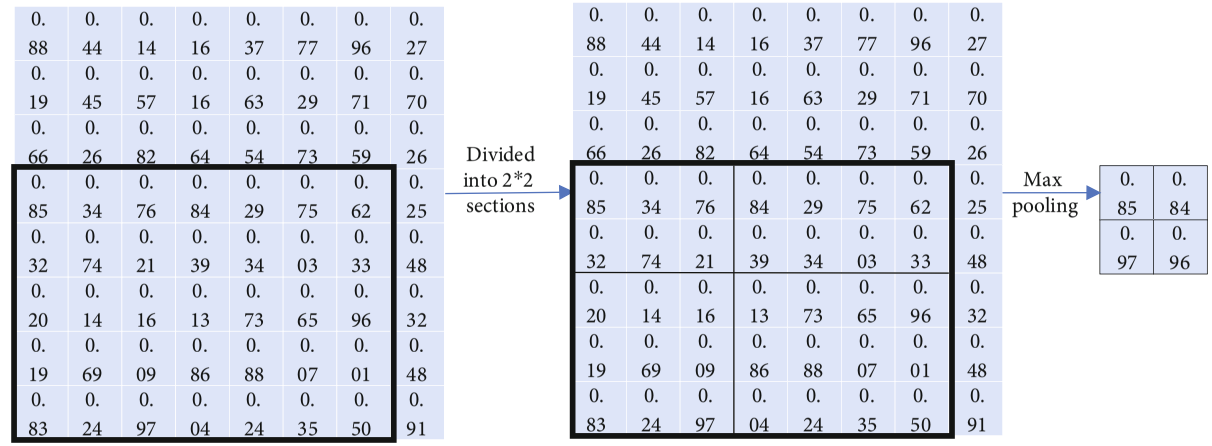
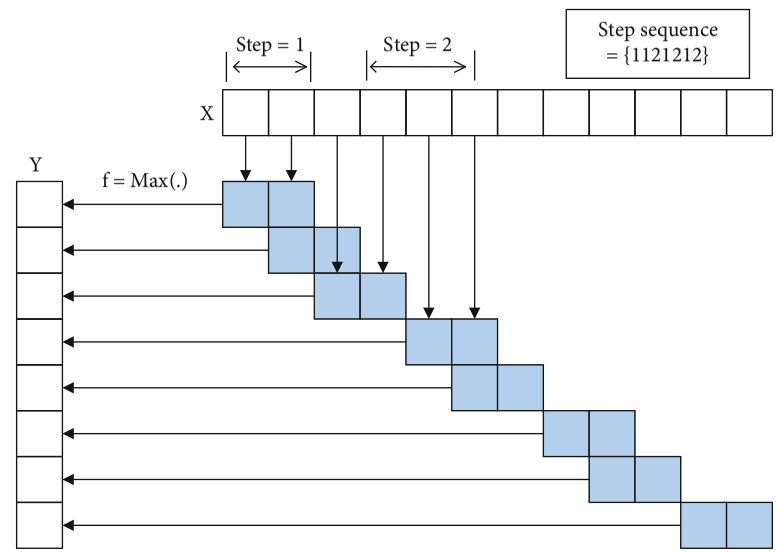
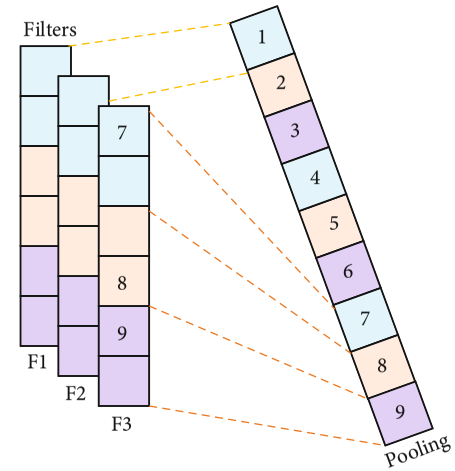
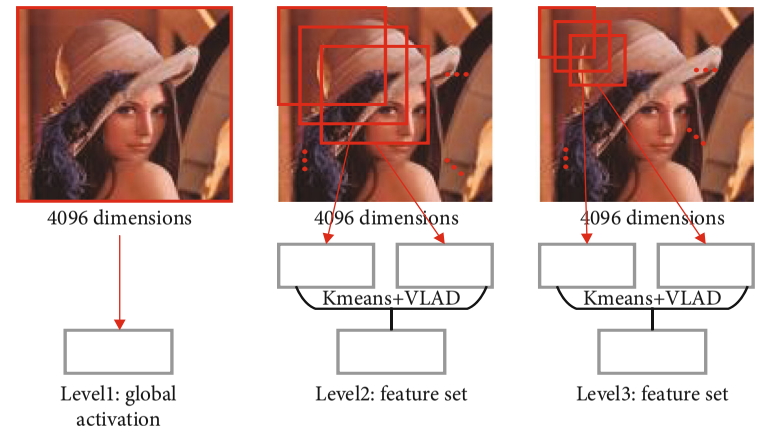
Pooling method	Characteristic	Sketch map
Region of interest pooling [17]	ROI pooling operation converts feature maps within a region of interest of any size into feature maps of fixed size	
Fractional max pooling [19]	Fractional max pooling operation allows the pooling domain to be noninteger values and reduces overfitting	
Strip pooling [21]	Strip pooling operation is easy to establish remote dependencies between discrete distributed regions. It can capture local details and can be easily embedded into any building block	$y_i^h = 1/w \sum_{0 \le j \le W} x_{i,j}, y_j^v = 1/H \sum_{0 \le i \le H} x_{i,j}$

Table 3: Continued.

Pooling method	Characteristic	Sketch map
Chunk-max pooling [22]	Chunk-max pooling operation retains the relative order information of multiple local max eigenvalues but does not retain the absolute position information	
Multiscale orderless pooling [23]	Multiscale orderless pooling operation improves the invariance of neural network activations without reducing its resolution	

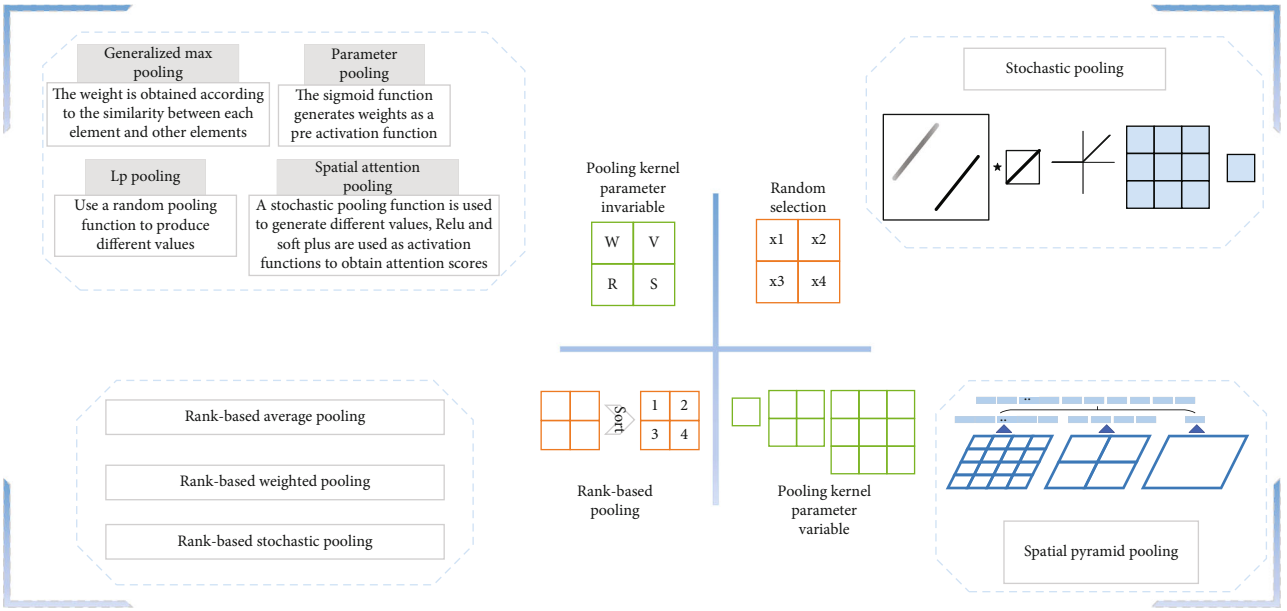


Figure 7: Variable of pooling kernel.

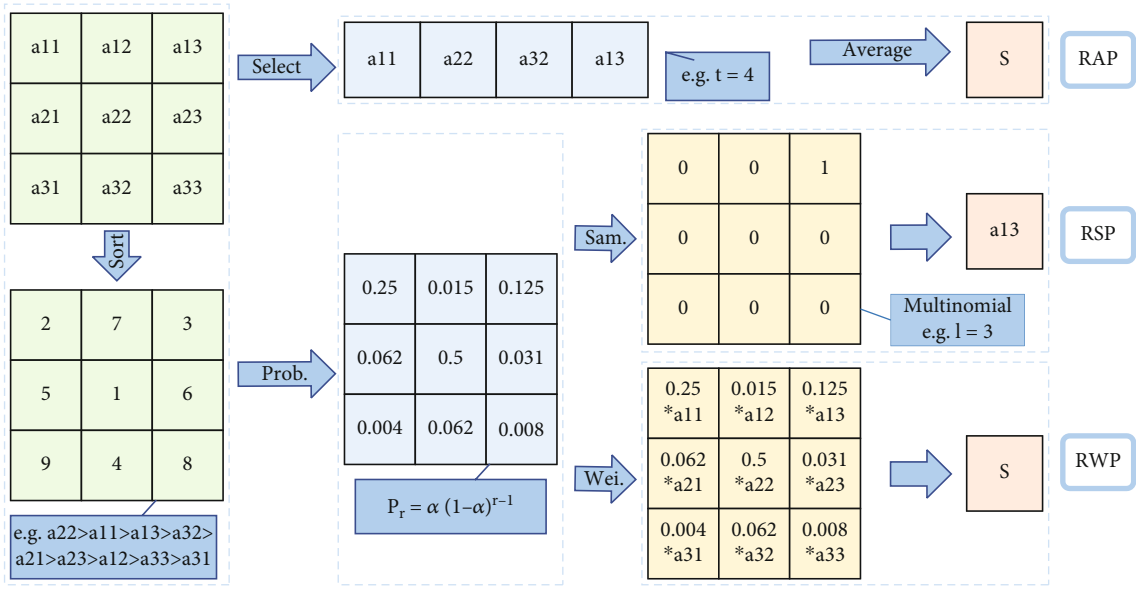


Figure 8: Rank-based pooling [28].

probability of being selected. Stochastic pooling operation is the process of random selection. When the selected pooling value is the max value in the pooling domain, the calculation process is equivalent to max pooling operation. Seventh, spatial pyramid pooling [33] operation performs max pooling operations on feature maps at three different scales. It replaces the last pooling layer with a spatial pyramid pooling layer, which can process information of different scales in the image to generate the fixed length feature information. Spatial pyramid pooling avoids the fact that most CNNs need to input fixed size images artificially, thus solving the limitation of image recognition accuracy due to fixed size. No matter how the feature map size changes, as long as the output length is fixed, the final number of features will be fixed. Images of different scales can be processed by Spatial Pyramid Pooling operation, so it is very flexible and can effectively prevent overfitting. However, some problems cannot be calculated in spatial pyramid pooling operation. When setting the output length, it is necessary to take into account the limitations of the actual situation. Table 4 shows the summary of variable of the pooling kernel.

6. The Invariable "+" Variable of Pooling

The invariable "+" variable of pooling means that it is a pooling method that uses different pooling methods for pooling separately and then fuses multiple pooling results.

Table 4: Summary of variable of the pooling kernel.

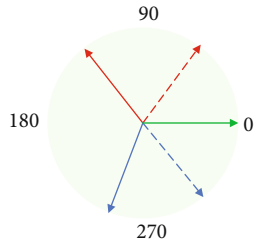
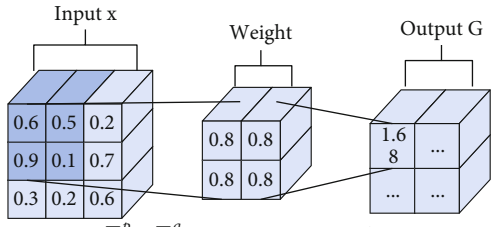
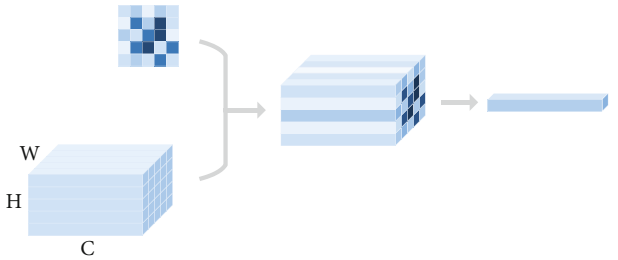
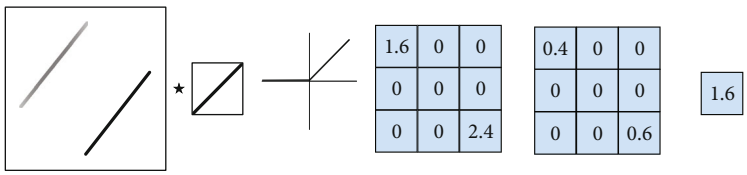
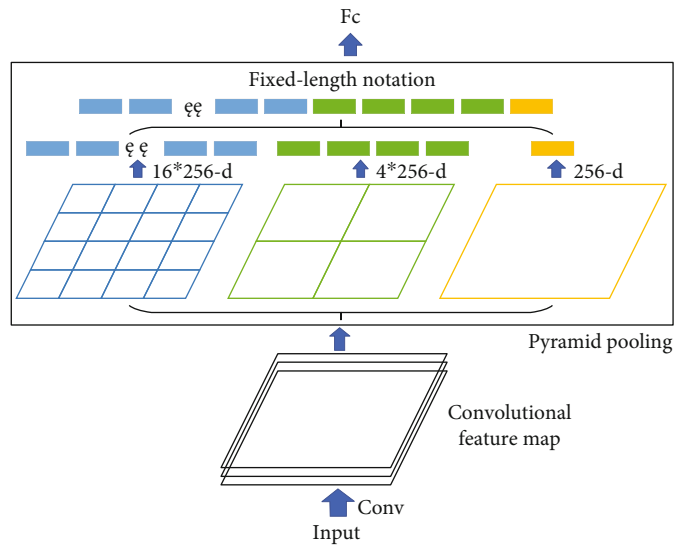
Pooling method	Characteristic		Sketch map
Generalized max pooling [24]	Generalized max pooling operation balances the influence of frequent pixels and rare data and improves the ability to extract fine-grained data		
Parameter pooling [25]	Parameter pooling operation converts the correlation operation into an interpretable pooling operation, which retains information and reduces errors		 $Y_c(i,j) = \sum_{m=0}^{p}\sum_{n=0}^{q} F_c(i+m, j+n) \cdot \beta(W_c(m,n))$
LP pooling [26]	LP pooling operation balances the effects of max pooling and average pooling.		$Y = \left(\sum\sum F(i,j)^P \times G(i,j)\right)^{1/P}$
Spatial attention pooling [27]	Spatial attention pooling operation mitigates the effects of distracting factors and focuses on meaningful parts of the image		
Rank-based pooling [28]	Rank-based average pooling	Alleviate the problem of loss of information in max pooling and loss of discriminative information in average pooling	$y_j = (1/t)\sum_{i\in f_j, r_i \le t} x_i$
	Rank-based weighted pooling	Assigning a weight value to each pixel in the pooling domain can improve performance	$p_r = \alpha(1-\alpha)^{r-1}$ $y_j = \sum_{i\in f_j} p_i x_i$
	Rank-based stochastic pooling	Alleviates the problem that random pooling is limited to nonnegative values and reduces overfitting	$p_r = \alpha(1-\alpha)^{r-1}$ $y_j = x_l$
Stochastic pooling [32]	Stochastic pooling operation is simple and has a strong generalization ability		 $p_i = x_i / \sum_{k\in f_j} x_k$ $y_j = x_l$, where $l \sim P\left(p_1, \cdots, p_{\lvert R_j \rvert}\right)$.

Table 4: Continued.

Pooling method	Characteristic	Sketch map
Spatial pyramid pooling [33]	Spatial pyramid pooling operation can handle images of different scales, is very flexible to use, and can effectively prevent overfitting. However, there are also practical constraints to consider	

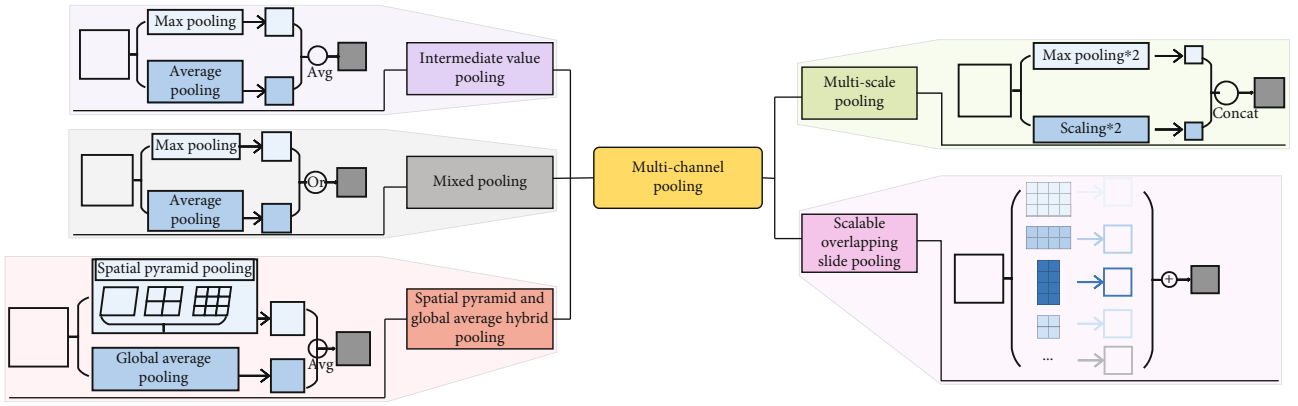


Figure 9: The pooling of invariable "+" variable.

It can also be called multichannel pooling. Multichannel pooling operation means that there are two or more pooling methods in pooling methods to compensate for the defects of a single pooling operation. However, at the same time, it will increase the amount of computation. Multichannel pooling includes intermediate value pooling operation, mixed pooling operation, spatial pyramid and global average hybrid pooling operation, multiscale pooling operation, and scalable overlapping slide pooling operation, as shown in Figure 9.

Firstly, intermediate value pooling [34] operation is used to perform max pooling operation and average pooling operation on the same feature map in parallel. The response values of the two pooling methods are averaged to obtain the response feature map. It combines the characteristics of both and highlights the most discriminative features while preserving the information integrity of the feature map. It is generally suitable for general images, because the error of the model is smaller and the stability is higher. Secondly,

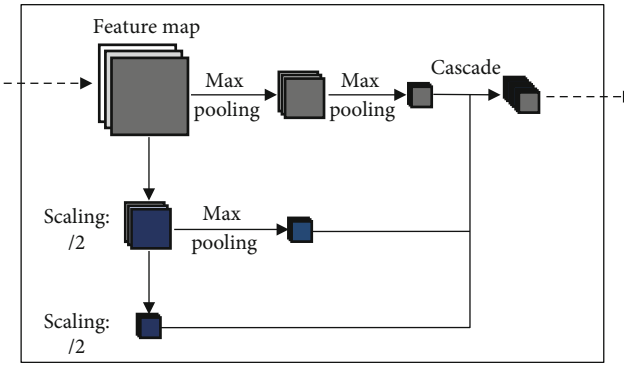
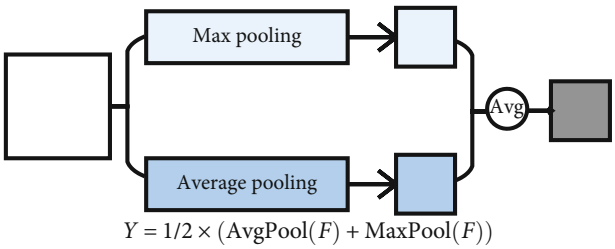
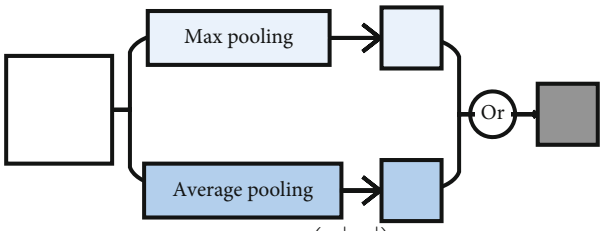
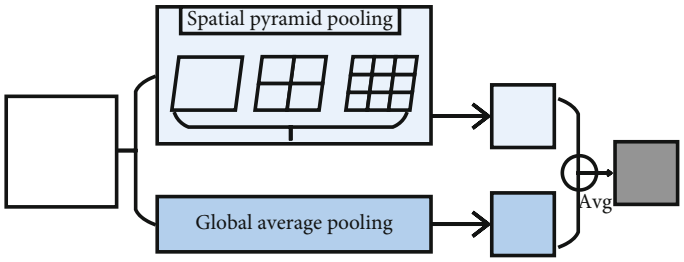
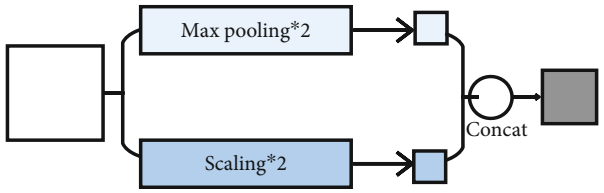
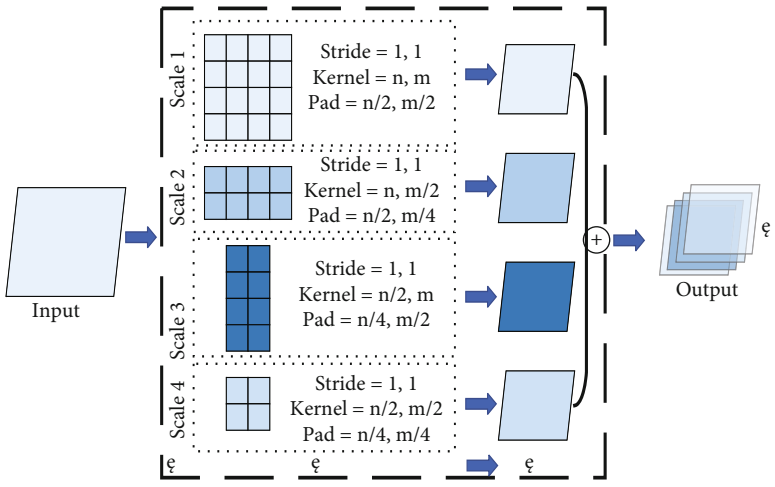


Figure 10: Multiscale pooling.

mixed pooling [35] operation is random using max pooling operation or average pooling operation. Mixed pooling operation utilizes random sampling instead of deterministic pooling operations. To some extent, mixed pooling operation

Table 5: Summary of multichannel pooling.

Pooling method	Characteristic	Sketch map
Intermediate value pooling [34]	It takes into account the average pooling and the max pooling, so that it has smaller model error and higher stability	 $Y = 1/2 \times (\text{AvgPool}(F) + \text{MaxPool}(F))$
Mixed pooling [35]	It solves the problem of which to choose between max pooling and average pooling, but only using one of these pooling methods still has the disadvantage of max pooling or average pooling	 $y_{kij} = \lambda \max\left(x_{kqp}\right) + (1-\lambda)\left(1/\left\|f_{ij}\right\|\right) \cdot \sum x_{kqp}\,(p,q) \in f_{ij}$
Spatial pyramid and global average hybrid pooling [36]	It can extract local and global information, respectively, which learns more information	
Multiscale pooling	It is more flexible and can be used multiple times at the beginning, middle, or end of the network	
Scalable overlapping slide pooling	It considers the saliency of features at different scales and the relationship between adjacent feature elements, so that coarse-grained, medium-grained, and fine-grained multiple features can be extracted	

solves the problem of which to choose between max pooling operation and average pooling operation. It can greatly improve the overall performance of the model when combined with other forms of regularization methods. Thirdly, spatial pyramid and global average hybrid pooling [36] operation uses spatial pyramid pooling operation and global average pooling operation for pooling feature maps in parallel. There are two pooling methods fused to get response values.

Before fully connecting, global average pooling operation can reduce parameters, dimensionality, and overfitting without destroying the spatial characteristics of feature maps. Spatial pyramid pooling operation shows the information of local characteristics. Global average pooling shows the information of global characteristics. There are two pooling methods fused that can learn more comprehensive information. Fourthly, multiscale pooling operation performs max pooling operation twice serially. Then, perform scale operation twice serially. Next, perform max pooling operation and scale operation serially. Finally, concatenate the feature maps obtained by three pooling methods as the response value, as shown in Figure 10. Multiscale pooling operation replaces the max pooling operation in ResNet, which is used to improve the recognition effect of images collected at different resolutions. Compared with the spatial pyramid model, the advantage of multiscale pooling operation is that its substitution position is more flexible and it can be used many times at the beginning, middle, or end of a network. Fifthly, scalable overlapping slide pooling [37] operation pools the feature maps in parallel through pooling kernels of multiple scales. The overlapping slide pooling operation obtains coarse-grained, medium-grained, and fine-grained feature maps. Finally, the multiscale pooling operation response feature maps of multiscale pooling are spliced as the response value. Single-scale pooling operation does not consider the saliency of features in different scales and the relationship between adjacent feature elements, which leads to decreased inaccuracy. Multiscale pooling operation solved this problem. More comprehensive and accurate granular features can be obtained through multiscale overlapping sliding pooling operation. The detection ability of the network for small targets and the overall detection accuracy can be improved. Table 5 shows the summary of multichannel pooling.

7. Conclusions

With the continuous development of multimedia technology, deep learning is widely used in text, speech, image, video, etc. Among them, pooling operation in deep learning has been deeply studied. And pooling operation is iterated from "invariable" to "variable," with different problems and tasks. These methods are constantly optimized due to different problems and tasks. Now, the most common methods of pooling are max pooling operation and average pooling operation. The newly proposed pooling method can improve the efficiency, improve the accuracy, and reduce the degree of overfitting.

In this paper, pooling operation in deep learning is comprehensively summarized. And introduce the pooling steps and their characteristics in detail and standardize the expression of pooling operation. The pooling operation is divided into the pooling domain, pooling kernel, step size, activation value, and response values. The improvement of pooling domain and pooling kernel in pooling operation is summarized as "variable" and "invariable" of pooling operation. Pooling operation can be classified into four categories: invariable of pooling domain, variable of pooling domain, variable of pooling kernel, and the pooling of invariable "+" variable. Invariable of pooling domain refers to the shape of the pooling domain being invariable in pooling operation. There are two types of invariable in pooling domain, including basic pooling methods and improved pooling methods. Variable of pooling domain means that shape and size of pooling domain are changed. Variable of pooling kernel means that the parameters, selection methods, and size of pooling kernel are changed. Pooling of invariable "+" variable means that there are two or more pooling methods in pooling methods. Although pooling operation is constantly being researched, deep learning is developing increasingly better, which has a better role for further research in deep learning.

Data Availability

The data that support the findings of this study are available from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (grant no. 62062003), Key Research and Development Project of Ningxia (grant no. 2020BEB04022), Natural Science Foundation of Ningxia (grant no. 2022AAC03149), and North Minzu University Research Project of Talent Introduction (grant no. 2020KYQD08).

References

[1] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 2006.

[2] C. Le, Y. Boser, and B. Denker, "Backpropagation applied to handwritten zip code recognition," *Neural Computing*, vol. 1, no. 4, pp. 541–551, 1989.

[3] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, "Dense convolutional network and its application in medical image analysis," *BioMed Research International*, vol. 2022, Article ID 2384830, 22 pages, 2022.

[4] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, article 106885, 2021.

[5] Z. Tao, H. Bing-Qiang, L. Huiling, S. Hongbin, Y. Pengfei, and D. Hongsheng, "18F-FDG PETCT whole body imaging lung tumor diagnostic model: an ensemble E-Res Net-NRC with divided sample space," *BioMed Research International*, vol. 2021, 13 pages, 2021.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[7] T. Li, Z. Meng, B. Ni, J. Shen, and M. Wang, "Robust geometric ℓp-norm feature pooling for image classification and action recognition," *Image and Vision Computing*, vol. 55, no. 2, pp. 64–76, 2016.

[8] T. Otsuzuki, H. Hayashi, Y. Zheng, and S. Uchida, "Regularized pooling," in *Artificial Neural Networks and Machine Learning-ICANN 2020*, vol. 12397 of Lecture Notes in Computer Science, pp. 241–254, Springer, Bratislava, Slovakia, 2020.

[9] P. Li, J. Xie, Q. Wang, and Z. Gao, "Towards faster training of global covariance pooling networks by iterative matrix square root normalization," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 947–955, Salt Lake City, USA, 2018.

[10] D. Acharya, Z. Huang, D. P. Paudel, and L. Van Gool, "Covariance pooling for facial expression recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 480–4807, Salt Lake City, UT, USA, 2018.

[11] L. Wanjun, X. Liang, and Q. Haicheng, "Learning performance of CNNs with different pooling models," *Journal of Image and Graphics*, vol. 21, no. 9, pp. 1178–1190, 2016.

[12] O. Morère, J. Lin, A. Veillard, L. Y. Duan, V. Chandrasekhar, and T. Poggio, "Nested invariance pooling and RBM hashing for image instance retrieval," in *The 2017 ACM on International Conference on Multimedia Retrieval*, pp. 260–268, Bucharest, Romania, 2016.

[13] Y. Lou, Y. Bai, J. Lin et al., "Compact deep invariable descriptors for video retrieval," in *Data Compression Conference*, p. 2375, Snowbird, USA, 2017.

[14] C. Zhang, G. Evangelopoulos, S. C. Voinea, L. Rosasco, and T. Poggio, "A deep representation for invariance and music classification," in *Proceedings Icassp IEEE International Conference on Acoustics Speech & Signal Processing*, pp. 6984–6988, Florence, Italy, 2014.

[15] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for CNNs, International Conference on Neural Information Processing Systems," 2015, https://arxiv.org/abs/1506.03767.

[16] Z. Hao and J. Ma, "Hartley spectral pooling for deep learning," *CSIAM Transactions on Applied Mathematics*, vol. 1, no. 3, pp. 518–529, 2020.

[17] Y. Sun, C. Sun, D. Wang, Y. He, and H. Lu, "ROI pooled correlation filters for visual tracking," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5776–5784, Long Beach, CA, USA, 2019.

[18] Y. Zhu, T. Xu, L. Peng et al., "Faster-RCNN based intelligent detection and localization of dental caries," *Displays*, vol. 74, 2022.

[19] F. E. Raymond and B. W. Raymond, "Voluntarily relinquishing private property rights: the existence of risk-pooling wquilibria when facing environmental uncertainty," *Theoretical Economics Letters*, vol. 5, no. 2, pp. 225–237, 2015.

[20] Y. Kaiyu, F. Xu, and J. Yu, "Shallow and wide fractional max-pooling network for image classification," *Neural Computing and Applications*, vol. 31, no. 2, pp. 409–419, 2019.

[21] Q. Hou, L. Zhang, M. M. Cheng, and J. Feng, "Strip pooling: rethinking spatial pooling for scene parsing," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4002–4011, Seattle, WA, USA, 2020.

[22] J. Lin, L. Ma, and C. Jingxia, "A frequency-domain CNN architecture based on the frequency-domain randomized offset rectified linear init and frequency-domain chunk max pooling method," *IEEE Access*, vol. 8, pp. 98126–98155, 2020.

[23] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features," in *The European Conference on Computer Vision (ECCV)*, vol. 8695, pp. 392–407, Zurich, Switzerlan, 2014.

[24] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, "Deep generalized max pooling," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1090–1096, Sydney, Australia, 2019.

[25] J. Zetao, Q. Jiaqi, and Z. Shaoqin, "Parameterized pooling convolution neural network for image classification," *Acta Electronica Sinica*, vol. 48, no. 9, pp. 0372–2112, 2020.

[26] P. Sermanet, S. Chintala, and Y. Lecun, "CNNs applied to house numbers digit classification," in *International Conference on Pattern Recoginition*, pp. 3288–3291, Tsukuba Science City, Japan, 2012.

[27] J. Ma and G. Xiaodong, "Scene image retrieval with siamese spatial attention pooling," *Neurocomputing*, vol. 412, no. 2, pp. 252–261, 2020.

[28] S. Zenglin, Y. Yangdong, and Y. Wu, "Rank-based pooling for deep convolutional neural networks," *Neural Networks*, vol. 83, pp. 21–31, 2016.

[29] S. H. Wang, V. Govindaraj, J. M. Gorriz, X. Zhang, and Y. D. Zhang, "Explainable diagnosis of secondary pulmonary tuberculosis by graph rank-based average pooling neural network," *Ambient Intelligence and Humanized Computing*, 2021.

[30] Y. D. Zhang, S. C. Satapathy, D. Wu, D. S. Guttery, J. M. Górriz, and S. H. Wang, "Improving ductal carcinoma in situ classification by convolutional neural network with exponential linear unit and rank-based weighted pooling," *Complex & Intelligent Systems*, vol. 7, pp. 1295–1310, 2021.

[31] Y. D. Zhang, C. Pan, X. Chen, and F. Wang, "Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling," *Journal of Computational Science*, vol. 27, pp. 57–68, 2018.

[32] Y. D. Zhang, X. Jiang, and S. H. Wang, "Fingerspelling recognition by 12-layer CNN with stochastic pooling," *Mobile Networks and Applications*, 2022.

[33] Z. Ke, C. Le, and Y. Yao, "A multivariate grey incidence model for different scale data based on spatial pyramid pooling," *Journal of Systems Engineering and Electronics*, vol. 31, no. 4, pp. 770–779, 2020.

[34] J. Chang, L. Zhang, N. Gu et al., "A mix-pooling CNN architecture with FCRF for brain tumor segmentation," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 316–322, 2019.

[35] B. A. Skourt, A. E. Hassani, and A. Majda, "Mixed pooling dropout for CNN regularization," *Journal of King Saud University Computer and Information Sciences*, 2021.

[36] Y. Wang, S. Xiaoning, and X. Wu, "Two-stream face spoofing detection network combined with hybrid pooling," *Journal of Image and Graphics*, vol. 25, no. 7, pp. 1408–1420, 2020.

[37] B. Qi, Z. Han, and Q. Kun, "Multi-scale stacking attention pooling for remote sensing scene classification," *Neurocomputing*, vol. 436, no. 12, pp. 147–161, 2021.

Hindawi
BioMed Research International
Volume 2022, Article ID 8265211, 11 pages
https://doi.org/10.1155/2022/8265211

Research Article

Computed Tomography Image Analysis of Body Fat Based on Multi-Image Information

Wei Zang,[1] **Fengrui Zhu**,[2] **and Yang Yu**[3]

[1]*Department of 2nd Clinical College, China Medical University, Shenyang 110004, China*
[2]*University of Surrey, 999020, UK*
[3]*Department of Radiology, Shengjing Hospital of China Medical University, Shenyang 110004, China*

Correspondence should be addressed to Yang Yu; 2019120802@cmu.edu.cn

Received 8 April 2022; Revised 14 May 2022; Accepted 25 May 2022; Published 20 June 2022

Academic Editor: Chen Li

Copyright © 2022 Wei Zang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Body fat assessment is required as part of an objective health assessment, both for nonobese and obese people. Image-based body fat assessment will enable faster diagnosis. Body fat analysis that accounts for age and sex will help in both diagnosis and correlating diseases and fat distribution. After evaluating computed tomography imaging algorithms to identify and segment human abdominal and subcutaneous fat, we present an improved region growing scale-invariant feature transform algorithm. It applies Naive Bayes image thresholding for key point selection and image matching. This method enables rapid and accurate comparison and matching of images from multiple databases and improves the efficiency of image processing.

1. Introduction

Obesity is a global health problem. Globally, 1.5 billion adults are overweight, and of these, 200 million men and 300 million women are obese. Child and adolescent obesity is increasing in both developed and developing countries, negatively impacting physical and mental health. According to the World Health Organization, global child obesity increased from 32 million in 1990 to 41 million in 2016. Obesity leads to metabolic syndrome and other complications, including type 2 diabetes mellitus, nonalcoholic fatty liver disease, hypertension, hyperlipidaemia, chronic kidney disease, cardiovascular disease, obstructive sleep apnoea, osteoarthritis, and malignant tumours (breast, colon, prostate, and others), thus increasing mortality [1].

Body fat imaging, which is crucial for diagnosis and research (Figure 1), has seen substantial advances in terms of field-of-vision and analysis. Reliable fat quantification is important in preventative healthcare and in diagnosing liver steatosis in nonalcoholic fatty liver disease, a risk factor for hepatocellular carcinoma development and progression [2]. Further, visceral obesity is an important predictor of metabolic syndrome, and measuring internal fat is important in understanding metabolic syndrome pathology. Nonetheless, there are as yet no unified diagnostic standards for fat detection.

Fat scanning methods include computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound. Among these, only CT provides an accurate measurement of visceral fat area. Abdominal bioimpedance analysis (BIA), which is as accurate as abdominal CT, can safely and simply detect excessive visceral fat accumulation. Abdominal BIA is more effective than using waist circumference. Abdominal BIA has been used to study the connections between visceral fat area and metabolic syndrome risk factors. The characteristics of subjects with excessive visceral fat revealed via BIA, rather than via waist circumference, have been studied [3–5].

X-ray imaging, which led to unprecedented advances in diagnostic imaging, has a few disadvantages. In X-ray images, the pixel intensity is the sum of the attenuation of all the rays passing through the object, and three-dimensional results are projected on a two-dimensional

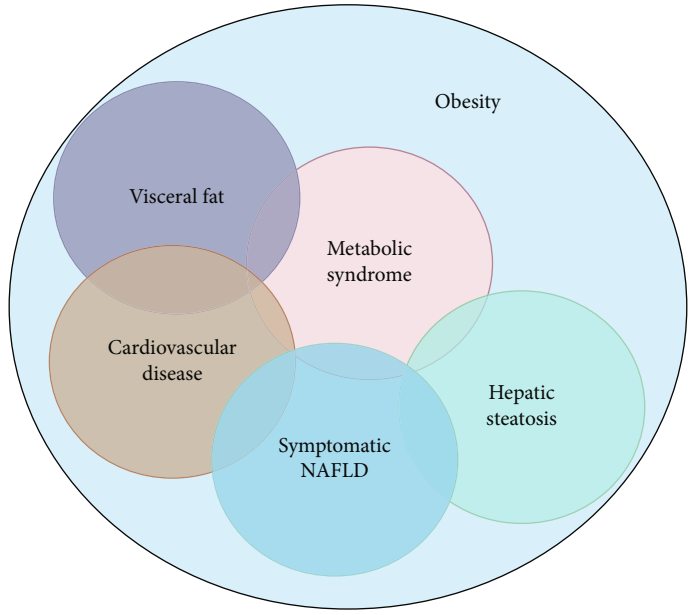


Figure 1: Relationships between obesity, visceral fat, cardiovascular disease, liver steatosis, metabolic syndrome, and nonalcoholic fatty liver disease [6].

plane. This inevitably duplicates information about the target's internal structure, preventing its clear display. CT imaging effectively overcomes this problem by scanning specific layers of the object and measuring the X-ray attenuation coefficient of the object at different angles. CT scanning provides multiangle data, enabling image reconstruction with reduced error. Computer-based reconstruction and algorithms enable the creation of multiple two-dimensional tomographic images.

Human abdominal fat comprises mainly visceral and subcutaneous fat. These two compartments with different molecular, biological, and anatomical compositions have a different meaning and importance. While subcutaneous fat shows a greater activity for long-term energy storage, visceral fat has a greater metabolic and hormonal activity through the release of adipokines [7]. Abdominal fat thickness is a risk indicator for conditions such as heart disease and diabetes. Although abdominal fat content, especially of intra-abdominal fat, is closely related to conditions such as type 2 diabetes, coronary heart disease, insulin resistance, and dyslipidaemia, the correlation between subcutaneous fat and these diseases is not significant [8, 9]. Measuring abdominal fat is more meaningful than measuring entire body fat. The body mass index, waist-to-hip ratio, waist circumference, and BIA are widely used to calculate and measure body fat. Waist circumference is widely accepted and used as an index of visceral fat.

Tetrapolar impedance measurement provides a simple and low-cost alternative method. Focused impedance, another method, provides local information. These two methods have been studied using an intuitive physical model and experimentally using a simulated abdominal subcutaneous fat layer. For accurate subcutaneous fat layer thickness measurement, tetrapolar and focused impedance require different electrode spacings. While small spacings between the current and potential electrodes hardly affect the impedance, larger distances reduce it substantially, until it becomes negligible. However, these methods require further analysis and experimental work [10].

For liver steatosis diagnosis and confirmation, complications are very rare. It is followed by histological classification based on the number of tissue cells with identifiable cytoplasm [11]. Although postbiopsy bleeding is a severe complication, there is a very low risk of massive haemorrhage if it is conducted carefully, and if patients are first screened properly. While such bleeding can occur unpredictably, it is related to age and malignancy [12]. Therefore, despite its advantages for determining visceral fat, the invasiveness and complications of biopsy provide motivation for developing noninvasive methods.

Ultrasound is fast and highly sensitive and has low negative effects on the human body. It is easily performed on comatose patients, which is an advantage with children, or with adults who cannot tolerate MRI, and has no sedative effect [13]. Abdominal obesity can be measured simply and economically via ultrasound and specifically via the newly developed quantitative ultrasound. In liver cirrhosis detection and measurement, quantitative ultrasound attenuation is determined based on the liver attenuation coefficient. It uses nonenhanced CT attenuation as the reference standard to enable noninvasive diagnosis and follow-up. However, ultrasound fat detection is not sufficiently intuitive and its results are difficult to store. The ultrasound device must be configured differently to detect visceral or subcutaneous fat. Because of these limitations associated with ultrasound imaging, CT and MRI are more widely accepted for visceral fat detection.

Compared to CT, MRI detects subcutaneous fat with similarly high consistency but with much greater error in visceral fat segmentation. Historically, nuclear magnetic resonance

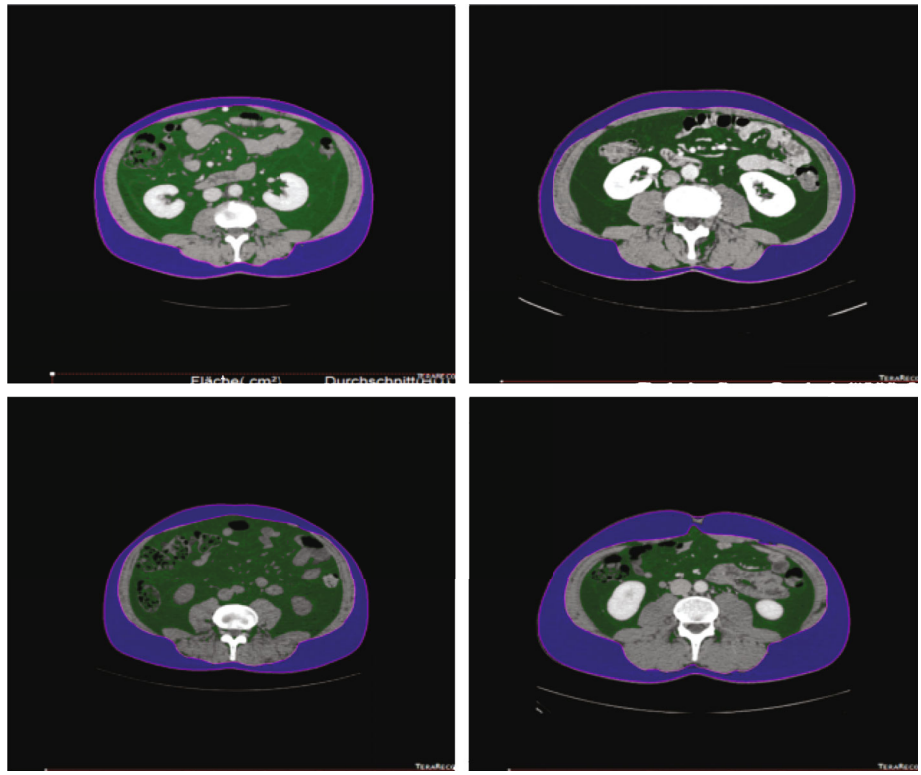


Figure 2: Manual segmentation of human fat CT images.

imaging has often been used to analyse visceral fat. For years, MRI has been considered the optimal noninvasive method for the assessment of fat accumulation [14]. However, it can disrupt cardiac metabolism, leading to complications; has low accuracy [15]; is slow; and is expensive, making it unpopular and uncommon. In human body, the classification of fat is not only subcutaneous and visceral but also white adipose tissue (WAT) and brown adipose tissue (BAT) according to its different functions. BAT is involved in energy dissipation and has been linked to weight loss, insulin sensitivity, and reduced risk of atherosclerotic disease [16]. Therefore, positron emission computed tomography (PET) has some advantages. For instance, 18F-FDG-PET/CT after cold exposure is the most commonly used and mature method for detecting and quantifying activated brown fat in the human body and evaluating its metabolic activity [17]. However, it is considered that PET has no significant effect on fat segmentation. The application of any radiotracer-based molecular imaging study in longitudinal studies with human subjects must take into account radiation dose [18]. So pet is not the main direction of fat segmentation.

CT, which has high repeatability, clinical safety, and convenience, produces high-resolution images and achieves accurate quantitative positioning [19]. Quantitative CT is therefore currently the technology-of-choice for body fat detection and analysis [20]. Nonetheless, it uses ionizing radiation, which limits repeated examination. Further, CT attenuation values based on voxels are confounded by the presence of substances such as glycogen, iron, copper, and iodine [21].

In this study, we analyse abdominal fat images (Figure 2), considering multiple factors, with the objective of elucidating human abdominal fat CT imaging. We examine various image segmentation methods and present an improved intelligent thresholding region growing segmentation method. We present a user-oriented graphical interface and examine image feature extraction and matching via scale-invariant feature transformation.

2. Materials and Methods

2.1. Image Segmentation. Image segmentation, central to image processing and developed over several decades, uses algorithms to decompose an image into several strongly correlated sets, segmenting the image into nonoverlapping areas. As with human visual cells, which separate objects from the background, digital image segmentation separates the objects as distinct areas of the image. Current techniques, using mathematical morphology, machine vision, intelligent algorithms, and other approaches, are based on segmentation thresholds, boundary extraction, and combined approaches.

Image segmentation, formally based on set analysis, strives to identify and extract the focal areas of an image. The entire image region is represented by the set R, which is divided into nonempty subsets $R1, R2, R3, \cdots, Rn$. Assuming that the two regions are critical, the union of Ri and Rj forms a connected region. Ri reflects regional connectivity, where $i = 1, 2, \cdots, N$. Neither the subregions nor regions intersect, and all pixels in the original image are within the union of all subregions. Connecting the pixels in each region in a defined way completes the segmentation process.

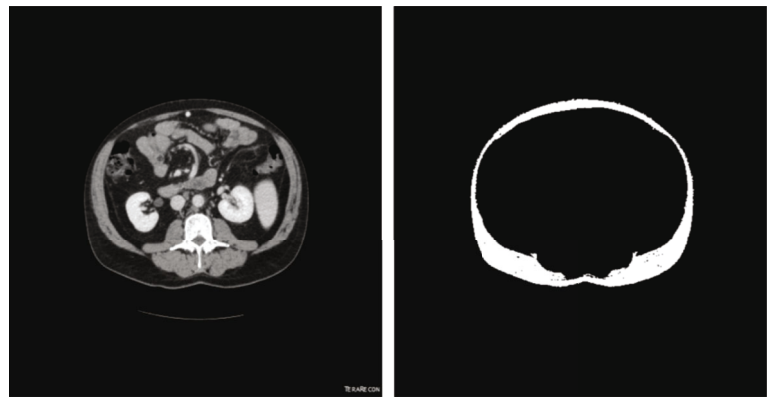


Figure 3: Segmentation based on the original and improved region growing algorithms.

For all i and j, $i \neq j$ and $Ri \cap Rj = \varnothing$ holds; $\varnothing$ represents an empty set, and $\cap$ is the image set intersection. For $i = 1, 2, \cdots, N$, $P(Ri) = \text{true}$ holds; P is a predefined attribute of R. Pixels in the same region should always have the same characteristics. Each subregion has unique characteristics: its elements must have the same grey level or similar characteristics. For $I \neq j$, $P(Ri \cup Rj) = \text{false}$ holds; $\cup$ is the intersection. Pixels in different regions and subregions should have different characteristics, without common elements.

Most image segmentation algorithms are based on identifying discontinuity or similarity (consistency) in grey-scale images. Edge detection, which depends on abrupt changes in grey scale, is a typical example of discontinuity-based segmentation, while threshold segmentation, region growing, and other methods are based primarily on similarity. In edge-based segmentation, local grey-scale discontinuities between the background and boundary are integrated to detect the boundary. Alternatively, subregions can be separated based on background characteristics such as colour, texture, grey scale, or other unique properties. Segmentation should achieve smooth boundaries, and the pixels should be kept as smooth as possible within each region, without gaps. Irregular shapes can be smoothed using filter operators or algorithms that attempt to maintain boundary integrity. Adjacent regions should be effectively contrasted. Image segmentation methods differ in their principles and associated problems, and different methods are required for different fields.

The image sources are as follows: computed tomography (CT) images are stored in Digital Imaging and Communications in Medicine (DICOM) format. The normal images' sizes are 512 by 512. In this step, the DICOM images are converted to grey-scale portable network graphics (PNG) images. All images were standardized so that their pixel values lie in the range of [0,1]. The bed plate interferes with the separation process, so the subcutaneous fat will be automatically segmented after the histogram threshold is selected to remove the bed plate obstacle area.

2.2. Region Growing Segmentation. Region growing combines pixels or subregions into large regions, based on the nature of the problem, applying predefined growth criteria: it requires seed selection, a similarity criterion (growth rule), and a growth stop condition. The seed can be a single pixel or small area of several pixels. This algorithm forms sets of merged pixels with similar properties. For each region, a seed point is designated as the starting point; the surrounding pixels are compared with it, and similar pixels are merged until all similar pixels are merged. The similarity criterion can be based on grey scale, colour, gradient, and texture.

Further, connectivity must be considered in segmentation. Methods based only on texture or other similarity criteria can lead to false fusion or incorrect segmentation that has no research value.

2.2.1. Seed-Point Selection. Seed pixels are selected to obtain seed points for regional growth. Change point selection can be automatic or based on semiautomatic (manual) human–computer interaction. Automatic selection can be targeted to field-specific image-analysis problems.

2.2.2. Improved Region Growing Algorithm. Here, we present an improved region growing algorithm, which differs primarily from the other common region growing algorithms in its seed-point selection. Because fat and surrounding tissue do not differ sufficiently in contrast, it is difficult to select seed points using grey gradient maps. The improved algorithm first applies multidirectional automatic positioning of seed points among subcutaneous fat pixels, thus avoiding manual selection. Seed points are selected based on their grey features, on Euclidean distances between pixels, or via position analysis. In CT images, the subcutaneous fat boundary can usually be clearly identified, occurring between the dermis and the fascia and enclosed by the superficial fascia. Automatic seed-point selection enables the region growing algorithm to proceed effectively.

2.2.3. Image Processing and Analysis. In CT images of body fat, the bed board appears in the lower part of the image, and its influence should be excluded to avoid meaningless segmentation. Because the location of the bed board is uncertain, there are more nonzero pixels in the subject area than in the obstacle area (Figure 3). By setting the threshold values for the nonzero pixels in the horizontal and vertical

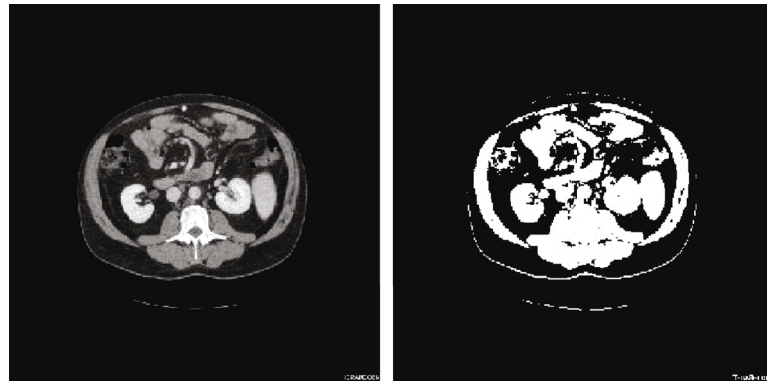


Figure 4: Comparison of segmentation result using the Bayesian algorithm.

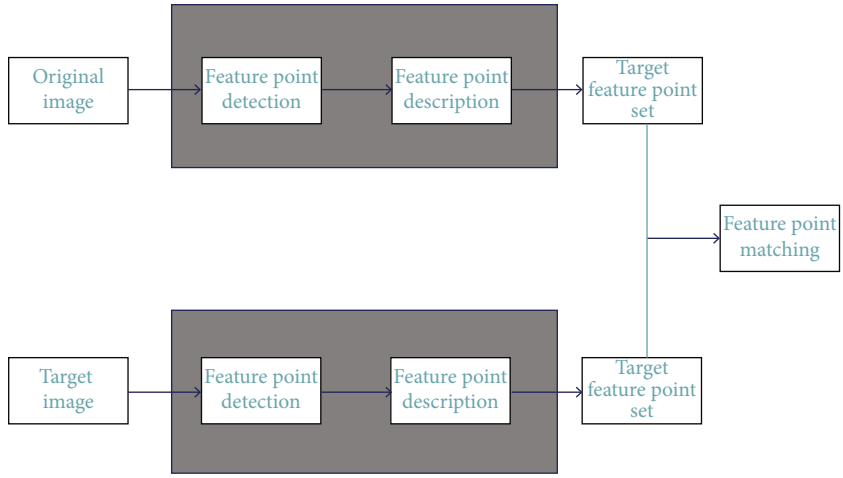


Figure 5: Implementation of the scale invariant feature transform (SIFT) algorithm.

directions, based on analysis of the grey histogram in various directions, the uninteresting areas can be identified and excluded.

When selecting seed points, the improved region growing algorithm passes through eight directions of two main lines in the centre of the image and two diagonal sublines, and the rays from all directions converge at the centre of the image. The centre falls on the first nonzero pixel it meets; the growth thus proceeds several units in this direction into the subcutaneous region, thereby completing the first step of region growing.

Four grey-value ratios are then obtained, based on the grey values of the pixels in the four neighbourhoods around the seed point and on that of the seed point. At ratios greater than the predetermined similarity criterion, the points that meet the criterion can be included in the seed sequence; otherwise, growth is stopped. As the number of seed points increases, the grey value of the seed sequence approaches the average grey value of the subcutaneous tissue. As long as the grey-value ratio of the points in each new neighbourhood to that of the average for the image meets the predetermined conditions, the new neighbourhood is merged into the region. Thus, the average grey level of the sequence is updated and the region is grown. Once almost all of the subcutaneous fat pixels have been evaluated, the set of pixels represents the target fat area.

2.3. Intelligent Image Segmentation. Threshold analysis, which is computationally simple and generates intuitive results, can produce good results for images with large grey-value differences between the target and background. Various intelligent algorithms have been applied in machine vision to classify grey thresholds. Here, we used classical Bayesian classification to select the optimal grey-value threshold for fat images.

2.3.1. Naive Bayes Algorithm. Image segmentation and classification are conventionally performed using the k-nearest neighbour's algorithm, neural networks, and decision trees. Bayesian classification applies Bayes' probability theorem (a class of intelligent algorithms) to categorise objects by their characteristics. Among such classification methods, Naive Bayes is the most efficient and effective, in some cases performing better than neural networks and decision trees.

The Naive Bayes algorithm, which is mathematically derived, is based on clear rules that are easy to understand. Further, it is easier to specify and debug than other methods. It is widely applicable and rapid to implement, accepts most

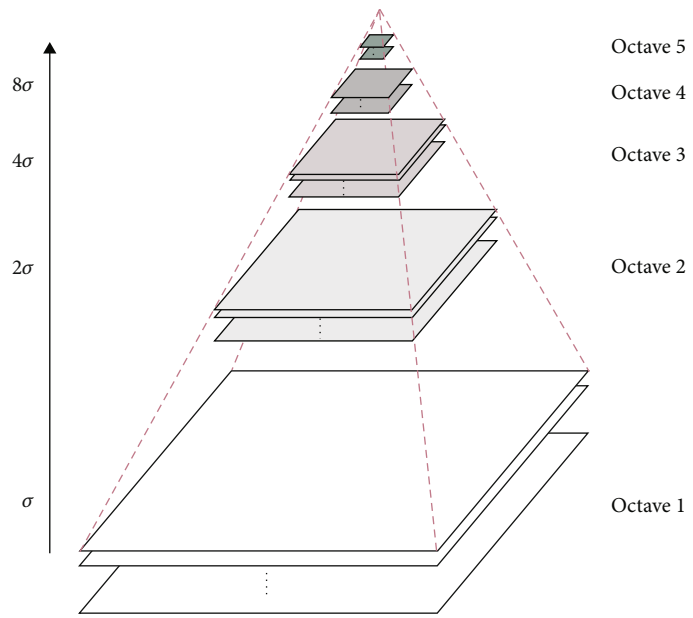


Figure 6: Gaussian difference pyramid model [23].

types of data, and is computationally efficient. In Bayesian classification models, the potential default attribute is independent. While this constrains its use somewhat, it also reduces its complexity and resource requirements.

2.3.2. Segmentation Analysis. Naive Bayes classification classifies items by placing them in the category with highest probability, under this condition. The initial threshold is the average of the minimum and maximum grey levels. The grey values and numbers of the pixels with grey values above or below the initial threshold are then obtained, for all images. The grey values and numbers of pixels on each side of the points between the threshold values are summed to obtain the total grey values above and below the initial threshold values and averages. An iterative threshold optimisation algorithm is then applied; after each iteration, the target threshold value is approached from the direction that increases the existing threshold probability value. The iteration continues until the difference between the front and back threshold values is close to 0, finally returning the optimal threshold value (Figure 4).

Although the optimised segmentation threshold is close to the ideal threshold, many connections between visceral and subcutaneous fat points are broken during the optimization process. Further, the segmentation is affected by the presence of many connections between the external boundary and the background.

2.4. Fat Image Analysis Based on Scale-Invariant Feature Transformation

2.4.1. Scale Invariant Feature Transform (SIFT) Algorithm. The SIFT algorithm (patented by David G. Lowe), which assumes that two photos contain the same content, applies affine transformation based on the degree of illumination,

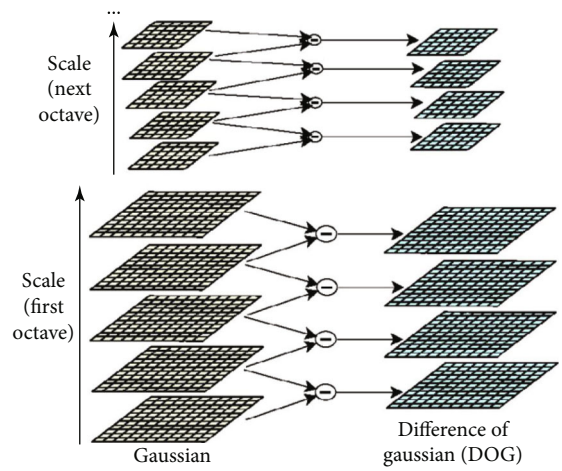


Figure 7: Construction of the Gaussian difference pyramid [23].

image angle, and size. It aims to match two points in the two images.

The SIFT algorithm recognises objects by extracting key points, creating corresponding descriptors, and identifying corresponding feature points via matching, thereby connecting the corresponding points (Figure 5). SIFT features are not affected by ordinary affine transformation and differences in image brightness and are relatively unaffected by noise. SIFT is thus suitable for accurate and fast image matching. Even when there are several objects in the same area, it generates many SIFT feature vectors.

For images of body fat, SIFT can match CT images of the same person at specific points. The ability to match processed body fat CT images to their corresponding database images, via feature point matching, will substantially improve diagnostic efficiency. This method thus contributes to the database and represents an advancement in image processing.

2.4.2. Scale-Space Construction. Before the SIFT algorithm is implemented, the scale space must be constructed (Figure 6); on the left-hand side of the figure is the Gaussian pyramid model, with several explanations. Gaussian convolution is one manifestation of scale space, and Gaussian blur is widely used to analyse scenes where it is necessary to reduce the size. Octaves represent groups of images of the same size. Images are processed by interval within each octave; scales within the scale space are interpreted as groups, and the image blur of each group gradually increases from bottom to top.

A Gaussian kernel, which is required to generate a multi-scale space, simulates clarity for near objects and ambiguity for distant objects [22]. To calculate Gaussian kernels with different variances, the original image is convolved with Gaussian kernels of different sigma, to obtain all the first-group (bottom) layers. The Gaussian pyramid is thus obtained by convolving the original image sequentially, via Gaussian kernels of different scales. The second group of images is obtained by down sampling the first group via interval sampling, which maintains a continuous scale. A single pixel has little influence on the whole image.

The normalised LOG operator is scale-invariant [23]. The LOG operator is related to differences in Gaussian kernel function, giving rise to the Gaussian difference operator. The Gaussian difference pyramid is obtained by subtracting one layer from its adjacent layer in the same group (right side, Figure 7). Gaussian difference calculation requires Gaussian smoothing of adjacent scales.

2.4.3. Accurate Key Point Positioning. After image thresholding, the extreme values must be detected. Local extreme value points in the Gaussian difference space constitute the key points. Key points are generally relatively stable, providing much information, and are used in image matching. They are usually retained even when external conditions change, appearing as bright points in dark areas or dark points in bright areas.

In Figure 8, the X point of the middle layer is compared with eight adjacent points in the same scale space and 18 vertically adjacent points, totalling 26 points. Every pixel is compared in the same way. After finding the extreme value points, they must be accurately located. The extreme value points and scale space are discrete at this stage. Applying ternary second-order Taylor expansion to the detected extreme value points and adjusting their positions improve their sub-pixel location accuracy. The number of iterations is limited. Points with low contrast (values <T/N) are removed; T is generally 0.04. In addition, to ensure feature point stability, the Gaussian difference edge response must be eliminated.

2.4.4. Constructing Key Point Direction. The selected key points, which are relatively accurate, are characterised by their X, Y, and σ (scale) parameters. We first find the image closest to the σ value of the key point scale in the Gaussian pyramid and draw a circle with the point as the centre and radius 1.5 times the Gaussian image scale (Figure 9). Accounting for the gradient direction and gradient amplitude of the pixels in the circle, the whole angle is divided into

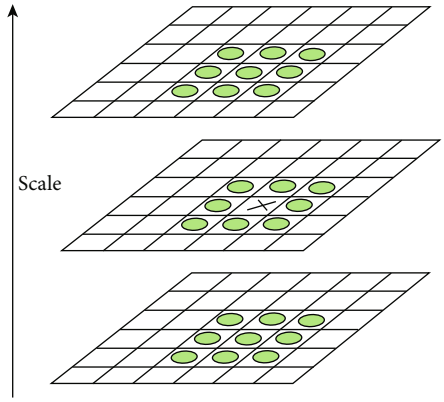


Figure 8: Extreme value point selection [22].

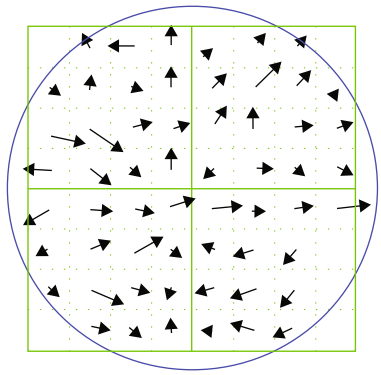


Figure 9: Image gradient of the area around the key points [22].

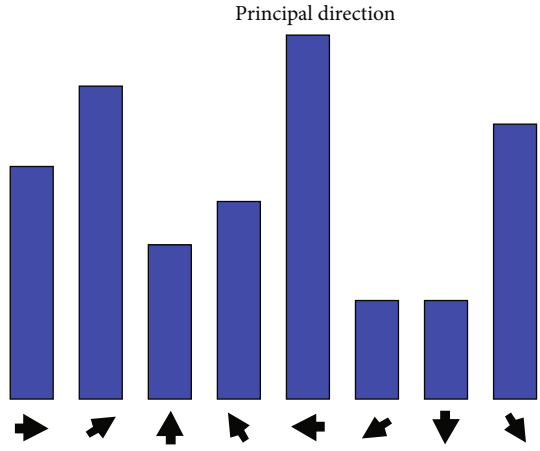


Figure 10: Histogram of key point direction.

eight directions. There are improved methods that select one direction every 10° in a 360° histogram, thus generating more than eight directions. In that case, 1.5σ Gaussian filtering is then applied.

To finalise the key point detection, the most frequent direction from the previous iteration is selected as the primary direction (Figure 10). Feature points have at most one auxiliary direction. A feature point with two feature directions is regarded as two feature points (one main and

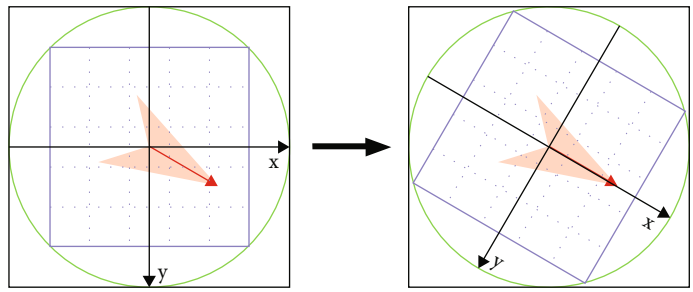


Figure 11: Axis direction transformation [22].

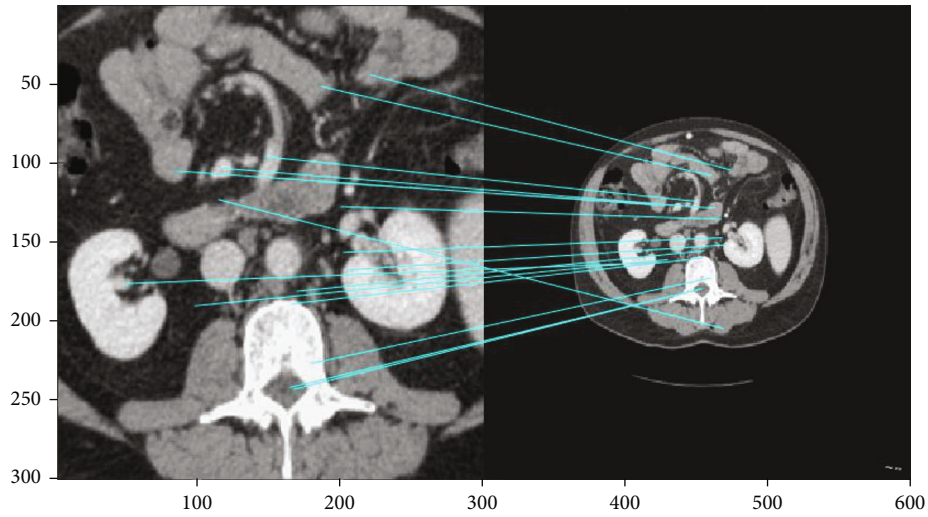


Figure 12: Feature matching between the central and whole region of the human abdominal fat image.

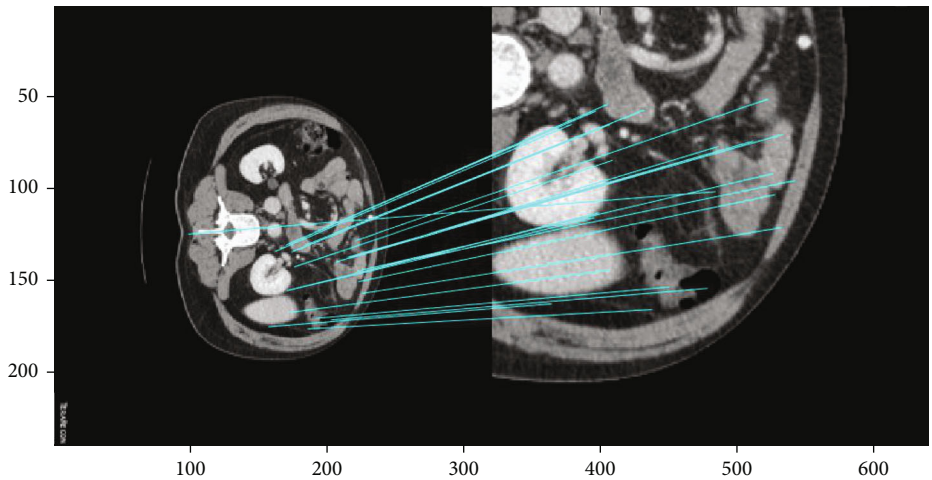


Figure 13: Image registration of the edge area and the whole area of CT images of human abdominal fat.

one auxiliary feature point), with the same position and scale, but different directions.

2.4.5. Key Point Descriptors and Matching. Key points have four parameters, namely, X, Y, σ (scales), and main direction (Figure 11). Pairs of images are matched using a descriptor, which describes the key point using a set of vectors. The descriptor contains both the key point and surrounding pixel points, which may contribute to target matching. The unique vector is generated by dividing the region into blocks and

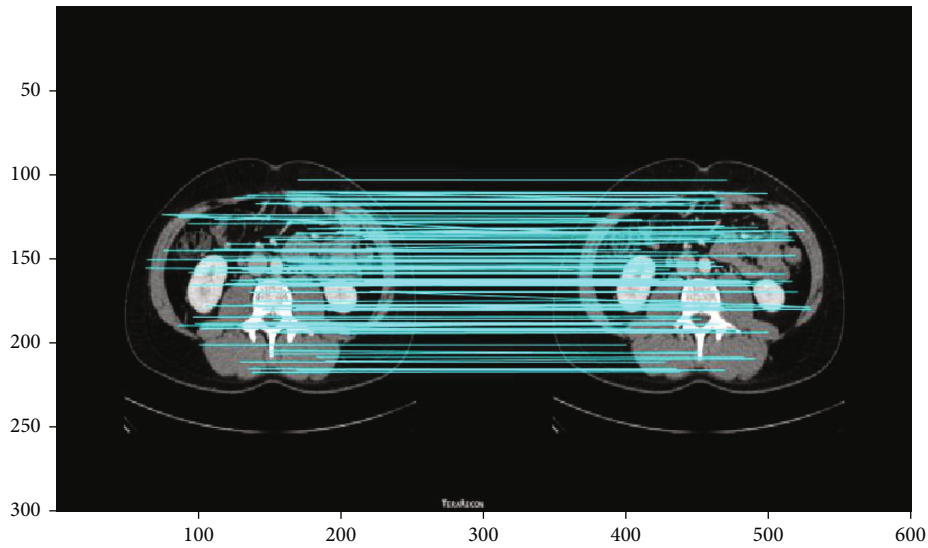


Figure 14: Image matching within the same group.

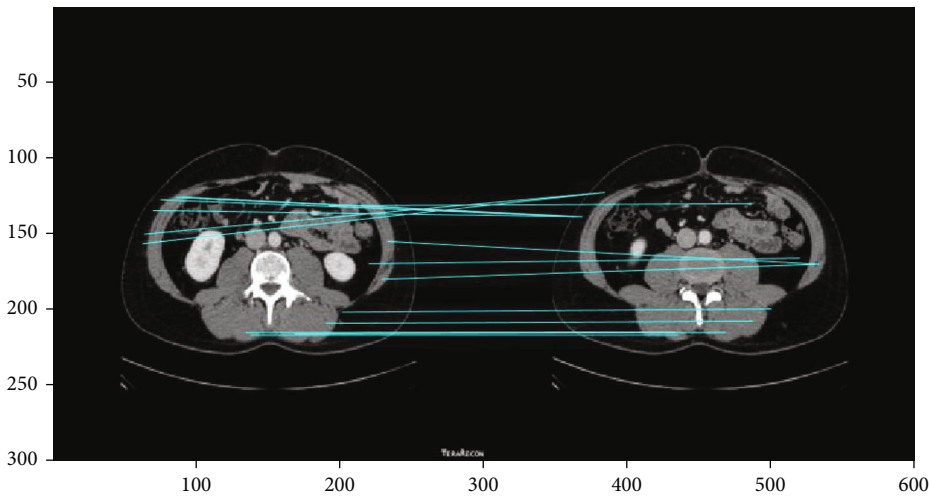


Figure 15: Image matching between human body fat images from different individuals.

calculating the gradient histogram for each block. Typically, 128-dimensional vector representation is the most advantageous.

Once the descriptor area is determined, it is converted to the main direction. Using the KD tree algorithm or the more efficient K nearest-neighbour algorithm, the descriptors of all key points in the two images are extracted, and the two descriptors with the shortest distance are selected. These two descriptors may be for the same key point. The gradients in eight directions are then counted in each subregion.

3. Results and Discussion

3.1. Matching of Parts with the Whole. With the exception of some fuzzy edge points, the key points and image registration were accurate (Figure 12).

There was a high degree of registration between the local image in the edge area and the whole image; further, the part cut from the right part area was consistent with the part of the original image matching the feature points (Figure 13).

Table 1: SIFT registration of the parts and whole of CT images of human abdominal fat from six groups (individuals).

Group	Key points	Matches	Time
1	324	20	0.124
2	428	27	0.232
3	683	33	0.353
4	564	26	0.257
5	477	36	0.296
6	515	38	0.153
Mean	498.5	30	0.235

3.2. Matching within and between Groups. Images in the same group came from one individual. Therefore, we conducted

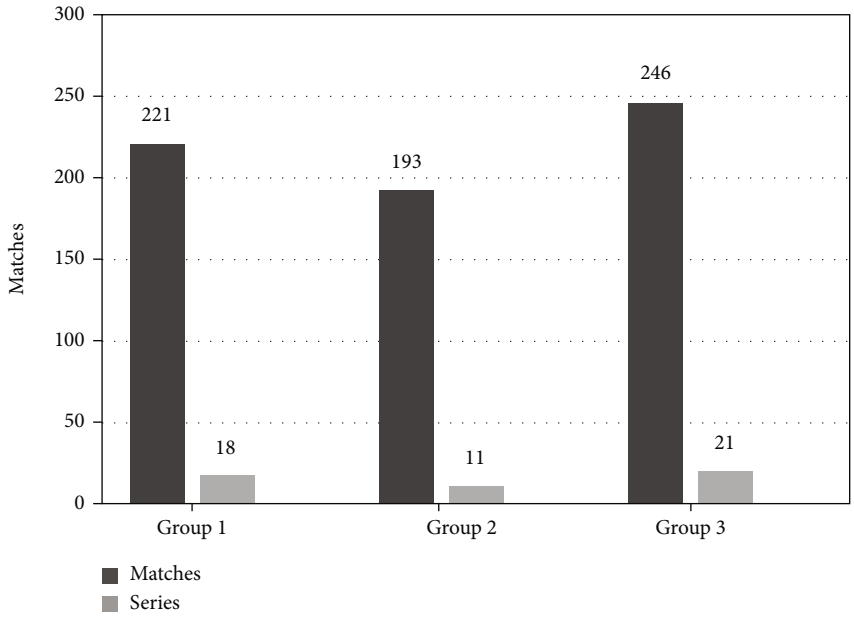


Figure 16: Comparison of the SIFT registration number between groups of CT images of human abdominal fat.

matching within groups to screen feature images for this individual in the database. There were 436 matched features, showing high matching accuracy (Figure 14).

We conducted feature matching of abdominal fat images from different groups (individuals). There were 21 matched features between these different groups (Figure 15), substantially less than the 436 matched features within the same group.

3.3. Data Analysis. Based on SIFT image registration, we conducted partial and overall analyses and intra- and intergroup analyses. The former analyses the key points, matches, and time parameters of the registration between the parts and whole, by list. The latter provides inter- and intragroup comparisons of the three groups. The mean key point value was 498.5, the average number of matches was 30, and the average matching time was 0.235 s (Table 1). This algorithm performs fast, identifying ca. 500 key points in each group and ca. 30 matches. To ensure proper representation, we selected six groups of CT images (that is, from six individuals) and registered and compared the whole and parts.

The left of the histogram (Figure 16) shows the contrast registration number of the same group, and the right shows the configuration number between different groups. We set up three control groups. From the control group histograms, it is clear that image registration was high within groups but low between groups.

The main file generated by the algorithm comprised of the input image Gaussian scale and difference spaces, local extreme value selection, and key point descriptor calculations based on the image gradient.

4. Conclusions

This study presented an improved region growing scale-invariant feature transformation algorithm for human body fat image segmentation and image registration. It used a Naive Bayes classification algorithm for intelligent threshold segmentation. It applied multidirectional convergence and automatic positioning of seed points, advancing semiautomatic manual seed-point positioning. As the seed points expanded, the region grew according to a ratio-similarity growth rule, until the subcutaneous fat region was completely segmented. Given the complexity of visceral tissues, and the abundance of connections between internal fat pixels, we did not examine visceral fat imaging in more detail.

This improved algorithm effectively matched the images within groups (individuals) based on feature similarity, in both the part-to-whole and intragroup registration tests. It has broad operational significance and performs better than other segmentation algorithms. This improved registration of human body fat CT images therefore advances image analysis and makes this algorithm more applicable for research. It provides a conceptual basis for future imaging research.

In the submission, we did not give the corresponding design and application, but in other articles we designed a "one-click segmentation" operation for the quick segmentation of fat images for the convenience of the users. The segmented image is obtained directly after the user interacts with the user interface (UI) component, and there are other useful operations for the user to handle the image.

The "one-click segmentation" operation includes functions such as opening and saving files, setting fonts, printing settings, specifying prompt messages, and entering commands. After the user selects the image in the computer library, the program will read in the image and the user can perform the crop step. The program will eventually show the results of processing the original image for this project and other images in interest. The user can directly see the processing results of various operators through this

detection function, and it is easy for the user to learn and discuss. After use, the user can also save and print the results, which can make the user feel the shortcut of the program, whether it is teaching or research.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] S. A. Polyzos, J. Kountouras, and C. S. Mantzoros, "Obesity and nonalcoholic fatty liver disease: from pathophysiology to therapeutics," *Metabolism*, vol. 92, pp. 82–97, 2019.

[2] I. Molwitz, M. Leiderer, C. Özden, and J. Yamamura, "Dual-energy computed tomography for fat quantification in the liver and bone marrow: a literature review," *Rofo.*, vol. 192, no. 12, pp. 1137–1153, 2020.

[3] P. M. Graffy and P. J. Pickhardt, "Quantification of hepatic and visceral fat by CT and MR imaging: relevance to the obesity epidemic, metabolic syndrome and NAFLD," *The British Journal of Radiology*, vol. 89, no. 1062, p. 20151024, 2016.

[4] B. H. Goodpaster, S. Krishnaswami, T. B. Harris et al., "Obesity, regional body fat distribution, and the metabolic syndrome in older men and women," *Archives of Internal Medicine*, vol. 165, no. 7, pp. 777–783, 2005.

[5] J.-P. Després, "Body fat distribution and risk of cardiovascular disease," *Circulation*, vol. 126, no. 10, pp. 1301–1313, 2012.

[6] K. Shoji, K. Maeda, T. Nakamura, T. Funahashi, Y. Matsuzawa, and I. Shimomura, "Measurement of visceral fat by abdominal bioelectrical impedance analysis is beneficial in medical checkup," *Obesity Research & Clinical Practice*, vol. 2, no. 4, pp. 269–275, 2008.

[7] F. Greco and C. A. Mallio, "Artificial intelligence and abdominal adipose tissue analysis: a literature review," *Quantitative Imaging in Medicine and Surgery*, vol. 11, no. 10, pp. 4461–4474, 2021.

[8] M. Premanath, H. Basavanagowdappa, M. Mahesh, and M. Suresh, "Correlation of abdominal adiposity with components of metabolic syndrome, anthropometric parameters and Insulin resistance, in obese and non obese, diabetics and non diabetics: A cross sectional observational study.(Mysore Visceral Adiposity in Diabetes Study)," *Indian journal of endocrinology and metabolism*, vol. 18, no. 5, pp. 676–682, 2014.

[9] J. P. Després, I. Lemieux, J. Bergeron et al., "Abdominal obesity and the metabolic syndrome: contribution to global cardiometabolic risk," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 28, no. 6, pp. 1039–1049, 2008.

[10] S. Haowlader, T. N. Baig, and K. Siddique-e Rabbani, "Determination of abdominal fat thickness using dual electrode separation in the focused impedance method (FIM)," *Physiological Measurement*, vol. 33, no. 5, pp. 707–718, 2012.

[11] M. Virarkar, J. Szklaruk, C. T. Jensen, M. W. Taggart, and P. Bhosale, "What's new in hepatic steatosis," *Seminars in Ultrasound, CT, and MR*, vol. 42, no. 4, pp. 405–415, 2021.

[12] D. B. McGill, J. Rakela, A. R. Zinsmeister, and B. J. Ott, "A 21-year experience with major hemorrhage after percutaneous liver biopsy," *Gastroenterology*, vol. 99, no. 5, pp. 1396–1400, 1990.

[13] A. Ozturk, J. R. Grajo, M. S. Gee et al., "Quantitative hepatic fat quantification in non-alcoholic fatty liver disease using ultrasound-based techniques: a review of literature and their diagnostic performance," *Ultrasound in Medicine & Biology*, vol. 44, no. 12, pp. 2461–2475, 2018.

[14] Q. H. Zhang, Y. Zhao, S. F. Tian et al., "Hepatic fat quantification of magnetic resonance imaging whole-liver segmentation for assessing the severity of nonalcoholic fatty liver disease: comparison with a region of interest sampling method," *Quantitative Imaging in Medicine and Surgery*, vol. 11, no. 7, pp. 2933–2942, 2021.

[15] I. J. Neeland, S. M. Grundy, X. Li, B. Adams-Huet, and G. L. Vega, "Comparison of visceral fat mass measurement by dual-X-ray absorptiometry and magnetic resonance imaging in a multiethnic cohort: the Dallas heart study," *Nutrition and Diabetes*, vol. 6, no. 7, p. e221, 2016.

[16] P. Santhanam, S. P. Rowe, L. B. Solnes, B. Quainoo, and R. S. Ahima, "A systematic review of imaging studies of human brown adipose tissue," *Annals of the New York Academy of Sciences*, vol. 1495, no. 1, pp. 5–23, 2021.

[17] M. Chondronikola, S. C. Beeman, and R. L. Wahl, "Non-invasive methods for the assessment of brown adipose tissue in humans," *The Journal of Physiology*, vol. 596, no. 3, pp. 363–378, 2018.

[18] J. Frankl, A. Sherwood, D. J. Clegg, P. E. Scherer, and O. K. Öz, "Imaging metabolically active fat: a literature review and mechanistic insights," *International Journal of Molecular Sciences*, vol. 20, no. 21, p. 5509, 2019.

[19] D. Berker, S. Koparal, S. Isik et al., "Compatibility of different methods for the measurement of visceral fat in different body mass index strata," *Diagnostic and Interventional Radiology*, vol. 16, no. 2, pp. 99–105, 2009.

[20] M. Cerit, H. N. Şendur, E. Cindil et al., "Quantification of liver fat content with ultrasonographic attenuation measurement function: correlation with unenhanced multidimensional computerized tomography," *Clinical Imaging*, vol. 65, pp. 85–93, 2020.

[21] J. Starekova and S. B. Reeder, "Liver fat quantification: where do we stand?," *Abdominal Radiology*, vol. 45, no. 11, pp. 3386–3399, 2020.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] T. Lindeberg, "Scale-space theory: a basic tool for analyzing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 1-2, pp. 225–270, 1994.

Hindawi
BioMed Research International
Volume 2022, Article ID 7512445, 15 pages
https://doi.org/10.1155/2022/7512445

Research Article

Region-Based Convolutional Neural Network-Based Spine Model Positioning of X-Ray Images

Le Zhang[1], **Jiabao Zhang**,[2] **and Song Gao**[3]

[1]*Department of Radiology, Qingdao Municipal Hospital, Qingdao University, Qingdao, China*
[2]*University College London, London, UK*
[3]*Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China*

Correspondence should be addressed to Song Gao; gaos@sj-hospital.org

Received 23 April 2022; Revised 11 May 2022; Accepted 31 May 2022; Published 17 June 2022

Academic Editor: Chen Li

Copyright © 2022 Le Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Idiopathic scoliosis accounts for over 80% of all cases of scoliosis but has an unclear pathogenic mechanism. Many studies have introduced conventional image processing methods, but the results often fail to meet expectations. With the improvement and evolution of research in neural networks in the field of deep learning, many research efforts related to spinal reconstruction using the convolutional neural network (CNN) architecture of deep learning have shown promise. *Purpose*. To investigate the use of CNN for spine modeling. *Methods*. The primary technique used in this study involves Mask Region-based CNN (R-CNN) image segmentation and object detection methods as applied to spine model positioning of radiographs. The methods were evaluated based on common evaluation criteria for vertebral segmentation and object detection. Evaluations were performed using the loss function, mask loss function, classification loss function, target box loss function, average accuracy, and average recall. *Results*. Many bony structures were directly identified in one step, including the lumbar spine (L1-L5) and thoracic spine (T1-T12) in frontal and lateral radiographs, thereby achieving initial positioning of the statistical spine model to provide spine model positioning for future reconstruction and classification prediction. An average detection box accuracy of 97.4% and an average segmentation accuracy of 96.8% were achieved for the prediction efficacy of frontal images, with good image visualization. Moreover, the results for lateral images were satisfactory considering the evaluation parameters and image visualization. *Conclusion*. Mask R-CNN can be used for effective positioning in spine model studies for future reconstruction and classification prediction.

1. Introduction

Scoliosis is defined as a C- or S-shaped lateral spine curvature greater than 10°. Adolescent scoliosis is a common three-dimensional deformity of the spine that manifests as a disease with abnormal spinal sequences in the coronal, sagittal, and transverse planes. Scoliosis is classified as idiopathic, congenital, or neuromuscular, depending on its etiology.

Cases of scoliosis where the pathogenic mechanism is unclear are classified as idiopathic scoliosis, currently accounting for over 80% of cases. In the United States, the prevalence of adolescent idiopathic scoliosis is 1-3% [1]. Although the incidence of congenital scoliosis is much lower than that of idiopathic scoliosis, congenital scoliosis has a more significant impact on the physical and mental health of patients because it is accompanied by structural deformities such as vertebral hypoplasia [2]. Without timely intervention and treatment during spinal growth and development, the angle of scoliosis gradually worsens, leading to spine deformity, ribcage deformity, and possibly pain and even nerve compression in the shoulder, back, and thoracic and lumbar spine and respiratory distress [3, 4]. Given the high incidence and impact on the physical and mental health of adolescents, scoliosis has been listed as one of the four major diseases of adolescents and

has received widespread attention from all segments of society. Therefore, the classification and prediction of the tendency of scoliosis through technology are of great interest.

Three-dimensional reconstruction of the spine is important for future scoliosis staging and outcome prediction. Further, the classification of scoliosis is an important foundation for determining an appropriate treatment plan. Currently, the classification of scoliosis and treatment guidelines are based on 2D spine radiographs. However, scoliosis manifests as a disease with abnormal spinal sequences in the coronal, sagittal, and transverse planes and the three-dimensional spatial structure of the abnormal vertebrae. Consequently, the diversity in curve types and morphology cannot be appreciated through 2D radiographs alone, possibly leading to misinterpretation of different scoliosis deformities with similar 2D parameters. Artificial intelligence technology can be used for objective measurement and analysis of the characteristic parameters of scoliosis and to achieve accurate classification based on the characteristic parameters. However, the progression and prognosis of scoliosis are difficult to determine even after accurate and objective scoliosis classification owing to vertebral growth disorders causing imbalances in spinal growth and other causes. Therefore, spine model positioning is important for subsequent studies requiring 3D reconstruction of the spine for scoliosis classification and treatment.

In this paper, we propose a training network model based on the Mask Region-based convolutional neural network (Mask R-CNN) architecture for one-step direct identification of spine models in human frontal and lateral X-ray images, including the lumbar spine (L1-L5) and thoracic spine (T1-T12). In addition, we provide the initial positioning of statistical spine models, thus providing spine model positioning for future reconstruction and classification prediction. Figure 1 shows a flow diagram of this study.

2. Related Work

Many studies have introduced conventional image processing algorithms for spine model positioning, including using conventional segmentation algorithms such as the Canny edge detector, the active contour (snake) algorithm [5], and the generalized Hough transform for automatic segmentation of vertebral projections. Subsequently, in conjunction with a generic 3D model of the spine and segmented vertebral radiograph images, 3D/2D alignment is used to separately estimate the position of each vertebra. However, the results of these attempts have predominantly been inadequate. Difficulties such as low contrast in fluoroscopic images and overlapping or concealed tissue structure make conventional segmentation algorithms ineffective. In addition, applying 3D/2D alignment to each vertebra is time-consuming, and the local alignment results are incompatible with the overall structure of the spine.

With the continued improvement and evolution of research in neural networks, encouraging research results have been obtained in academic studies. Many research studies related to spinal reconstruction have used the CNN architecture. Arif et al. employed a CNN-based segmentation model to segment the spine in lateral radiographs and obtained ideal experimental results with high accuracy [6]. However, because of the single output of the CNN, the boundaries of overlapping parts are difficult to determine. The individual optimization of CNNs in multiple fragmented processes increases algorithm complexity and makes quality control difficult. In contrast, the primary technique used in our study is Mask R-CNN-based image segmentation and object detection methods as applied to spine model positioning in X-ray images.

2.1. Developments in 3D Spine Reconstruction. The evolution of 3D spine reconstruction algorithms has been following the technical route of gradually increasing the fine-tuning parameters of 3D spine models while reducing manually controllable parameters during the reconstruction process. Commonly used spine models are typically obtained by generating 3D models from boundaries outlined in sectioned computed tomography or magnetic resonance imaging images [7]. Owing to differences in algorithms and use cases, the generated 3D models are of three primary types: point cloud, mesh, and nonuniform rational B-splines [7–10]. Earlier reconstruction algorithms required manually locating 0-30 bony marker points for each vertebra in projection images to predict its 3D coordinates and generate a 3D mesh model for morphological adjustments and eventually complete the reconstruction [11]. Various statistical models based on standard a priori databases have since emerged to reduce the processing time. Semiautomatic methods using these models require only a small number of parameter inputs to predict the geometric parameters of a 3D model of the spine. There are two types of models: parametric dimensionality reduction models and parametric spine models. The parametric dimensionality reduction model is a principal component analysis model or a multiple linear regression model aimed at reducing the parameter inputs [12, 13].

Early parametric spine models generally utilized six marker points per vertebra, which were sufficient to generate the entire point cloud model of the spine. However, such models lacked parametric degrees of freedom and could not fully represent a fine-grained 3D mesh model [12, 13]. To address this issue, Humbert et al. made modifications and proposed a more mature and easy-to-use semiautomatic reconstruction algorithm [14]. Their method accounts for both the longitudinal and transverse geometric morphologies of the spine. The longitudinal morphology is determined by a line through the center of the vertebral body, whereas the transverse morphology is determined according to a simplified parametric model of the local deformation of the vertebrae. This two-level parametric description is still used today. However, this semiautomatic method has different sensitivities to the location of individual marker points for manual input. Operator interaction with the iterative input process is often nonintuitive and time-consuming, at approximately 11.5 min on average [15].

Recently, Parent et al. proposed a CNN-based architecture for bony anatomical landmark detection in frontal and lateral images. These include the vertebral arch and the vertebral endplate, which allow for finer adjustments in the reconstructed model [16]. Aubert et al. conducted a study based on bony landmark detection and achieved rapid automated reconstruction within 2 minutes [17]. However, individual optimization of CNNs in multiple fragmented processes increases algorithm complexity and makes quality control difficult; in addition,

Figure 1: Flow diagram of our Mask R-CNN-based spine model positioning study.

such algorithms still suffer from low contrast caused by overlapping tissue structures.

2.2. Developments in Spine Model Positioning in Spinal Reconstruction. Before spinal reconstruction, the coordinate system and geometric scale relationships must be defined based on the accurate positioning of bony structures such as the femoral head, sacroiliac joint, and sacral endplate. In semiautomatic reconstruction methods, these data are manually marked. In a recent automatic reconstruction study, Aubert et al. used a multilevel, mutually independent CNN model to determine these bony structures separately to improve the accuracy of parameter estimates through multiple adjustments [17]. However, training multiple CNNs separately in this manner makes it difficult to achieve global cooperative training optimization and poses problems for quality control of the algorithm in clinical applications.

In recent years, the accuracy and speed of multitarget recognition frameworks in the field of computer vision have reached unprecedented levels owing to the development of deep learning [18]. In contrast to conventional CNN, the R-CNN architecture can classify multiple regions of interest in the input image based on a feature map obtained from the shared CNN training, eventually resulting in multitarget recognition and positioning. Mask R-CNN can further segment the object boundaries to facilitate higher-order in-depth analysis. Therefore, we aimed to use the R-CNN architecture to train the network model to directly identify many bony structures, including lumbar and thoracic vertebrae in frontal and lateral X-ray images in a single step, thereby achieving the initial positioning of the statistical spine model.

Thus, the present study differs from recent studies in that the Mask R-CNN architecture is used to train the network model to directly identify many bony structures, including the thoracic spine and lumbar spine, in frontal and lateral radiographs in a single step, while providing the initial positioning of the statistical spine model, thereby simplifying the positioning process.

3. Materials and Methods

3.1. Preprocessing and Serial R-CNN Models

3.1.1. Preprocessing. In this study, we preprocessed the human X-ray images acquired before putting the dataset into a standard format. The X-ray images used were anonymized X-ray images in DCM format. First, ImageJ was used to convert the X-ray images from DCM to JPG format.

3.1.2. R-CNN. R-CNN uses an object recognition method on regions based on sliding windows. First, category-independent sets of candidate regions are generated on a given image, and each candidate region set is labeled with a category and a bounding box. Next, for each candidate region, a CNN is used to extract a fixed-length eigenvector. A support vector machine (SVM) classifier is trained for each category to determine the probability that the eigenvectors belong to each category. The eigenvectors output from the CNN are fed into the SVM classifier to predict the probability that the objects in the candidate regions belong to each category. Figure 2 depicts the R-CNN model.

3.1.3. Faster R-CNN. Faster R-CNN improves upon R-CNN by adding a region proposal network (RPN) for detecting target regions instead of using selective search to extract candidate regions. The RPN is a fully convolutional network that can exclude meaningless regions not belonging to any target category, reduce the number of proposed regions generated, and increase the speed of candidate region generation while ensuring precise object detection. Figure 3 illustrates the Faster R-CNN model.

3.1.4. Mask R-CNN. Mask R-CNN adds a feature pyramid network (FPN) to overcome the limitations of region transformation, improve algorithm efficiency, and effectively use the pixel-level location information of the training data image labels. Introducing a residual network as deep semantic information prevents the vanishing gradient problem, thereby improving model accuracy. Figure 4 illustrates the Mask R-CNN model algorithm.

Mask R-CNN uses a residual network to better utilize the features extracted from deep convolution. The RoIPool is replaced with a region of interest alignment layer using a bilinear difference method instead of roughly selecting a value to substitute for the value of the entire region because each region also has a size gap. The bilinear interpolation method is extended by calculating the linear difference in each of the two directions and is more appropriate for mapping the region of interest to the feature map by performing bilinear interpolation for each point. The output of the region of interest alignment layer contains the same shape as the feature map and can be used for mask classification prediction and the bounding-box shape of the region of interest in all regions of interest. In addition, the pixel-level mask position of the target can be predicted using the full CNN. Improving the RPN to FPN connects different convolutional layers to better identify features at different scales.

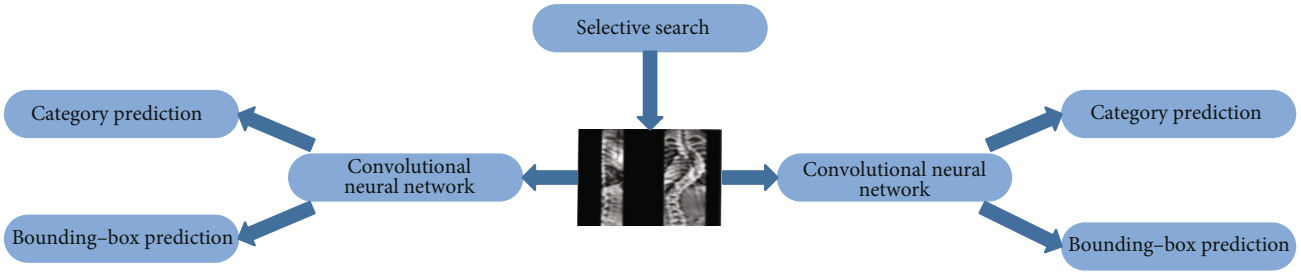


Figure 2: Schematic of R-CNN model.

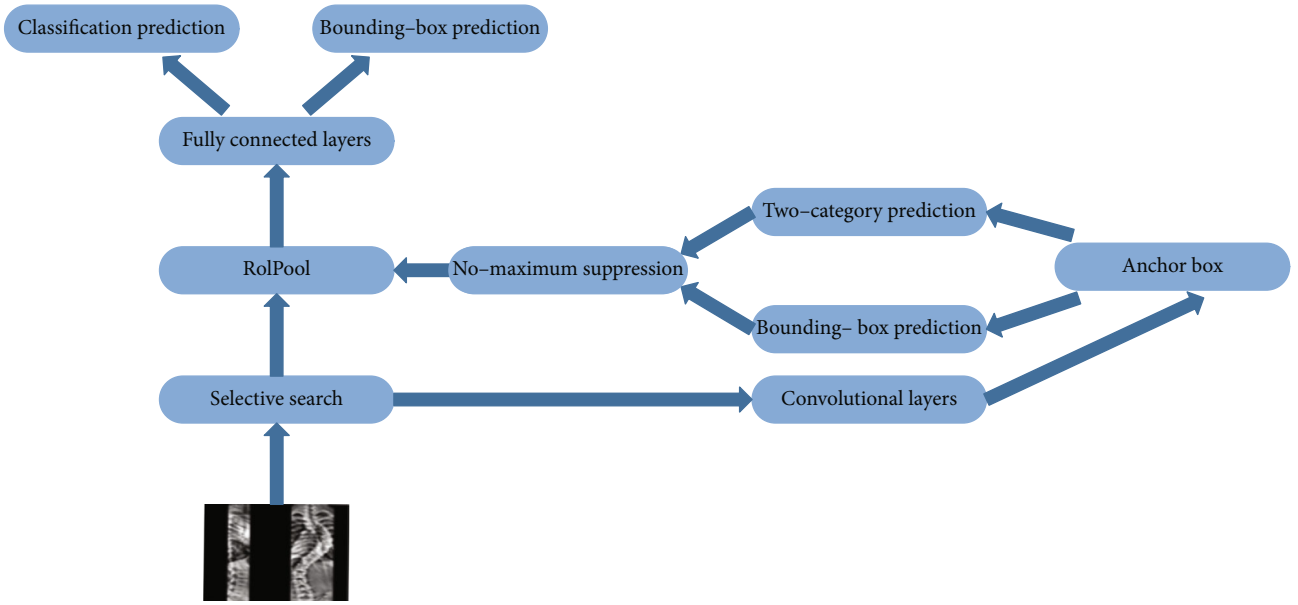


Figure 3: Schematic of the Faster R-CNN model.

3.2. Basic Principles of the Mask R-CNN Model

3.2.1. Spine Positioning Model Design. As shown in Figure 5, the spine X-ray image positioning network structurally contains modules such as a CNN backbone, feature map processing, mask calculation, bounding-box prediction, and category prediction. First, spine radiographs are input into the Mask R-CNN, and then, the convolutional layer of the CNN backbone extracts features from the input image. Next, the candidate regions of the vertebrae are extracted using the RPN, with feature maps of different scales generated by the FPN. Finally, the feature map set is passed into the head of the network, comprising three branches: mask branch that calculates the 2D mask of vertebra instances in each region of interest, box regression branch to determine the extent of the vertebra box, and classification prediction branch, for instance, classification prediction.

3.2.2. Convolutional Computation. At the base of the model is a convolutional backbone that extracts the input image feature map through layer-by-layer convolutional computation. The entire spine model localization process must undergo convolutional computation before subsequent classification, vertebral mask generation, and vertebral box computation can be performed. Convolutional computation often requires improvement. Simply increasing the network depth results in the vanishing gradient problem, causing the accuracy of results to gradually tend toward fitting after increasing to some value, and the accuracy will instead keep decreasing as the network depth increases.

In deep convolutional networks, although the number of convolutional layers and the feature expression capability increase with increased convolutional computation, the vanishing gradient problem often remains. The deeper the layers of an ordinary neural network, the closer the initialization parameters are to zero. As the training of neural networks usually involves a backpropagation algorithm for chain product derivation, the more the gradient of the shallow layer tends to zero as the information is propagated forward when the parameters of the shallow layer are updated, the more the gradient eventually disappears. In other words, the accuracy tends to fit after the depth of the model increases to a certain limit, and the accuracy decreases if the convolution depth of the network continues to increase. Therefore, the Mask R-CNN model introduces a residual convolutional network (ResNet) to effectively avoid the vanishing gradient problem and improve model accuracy [19–21].

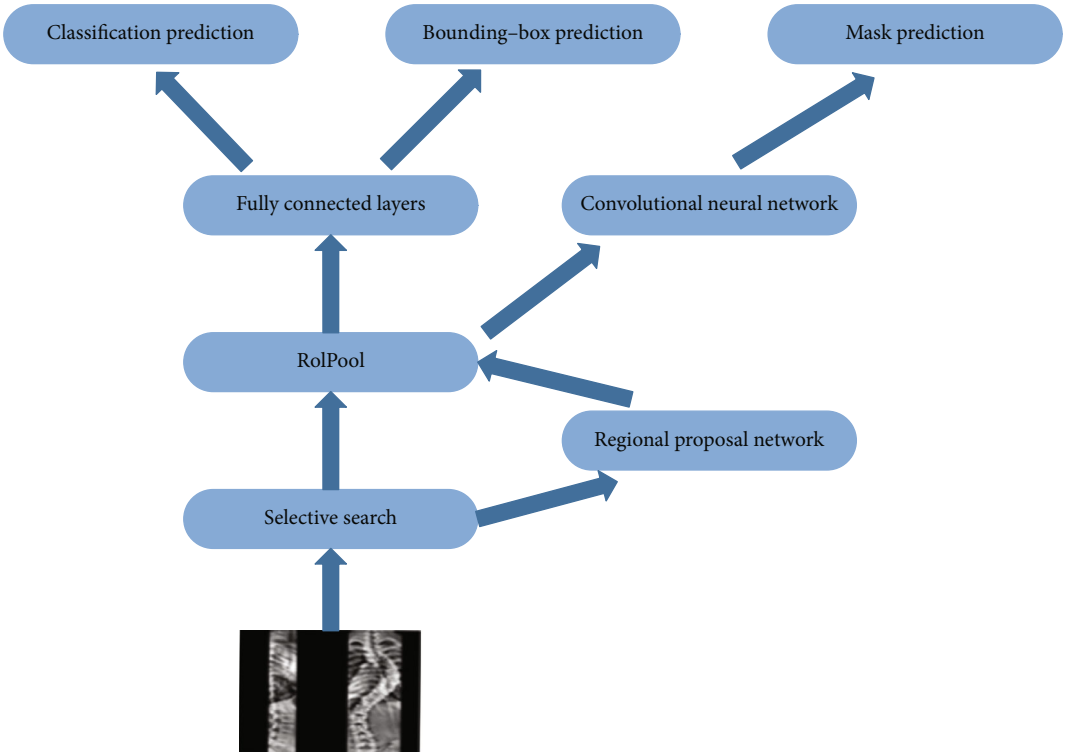


Figure 4: Schematic of the Mask R-CNN model.

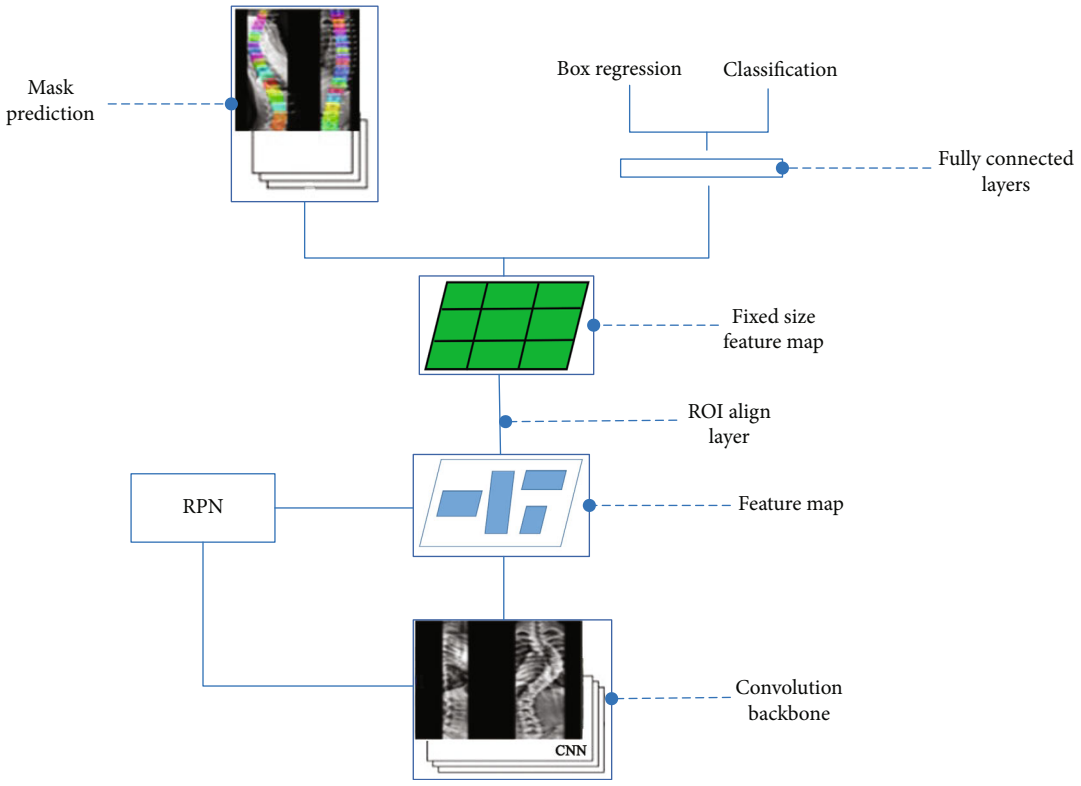


Figure 5: Schematic of spine model positioning network.

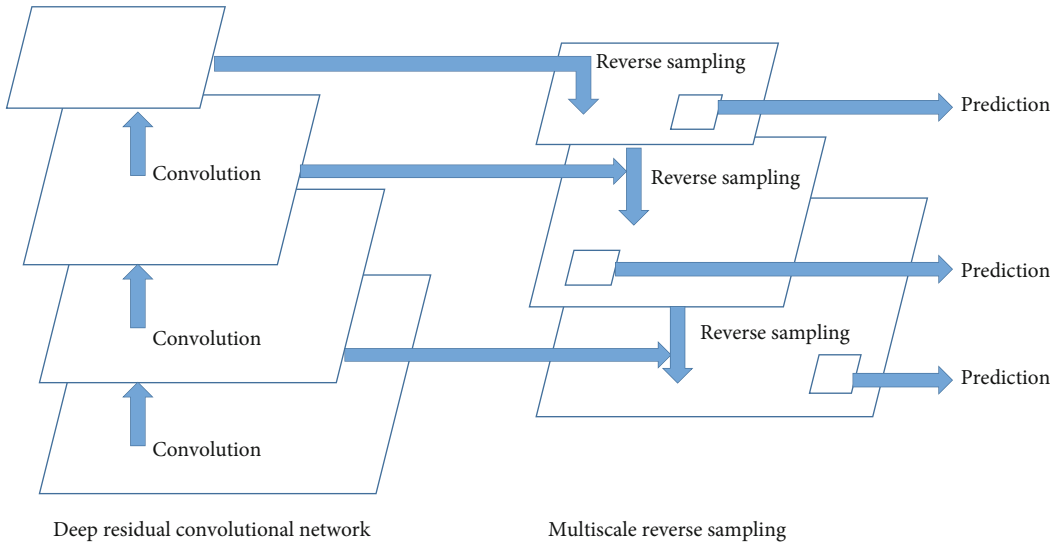


Figure 6: Schematic of the feature pyramid network (FPN).

Equation (1) is the mathematical representation of the residual module. The input data x_i of the residual block are identity mapped to $w_i x_i$ ($W = 1$ without dimensional conversion) through a shortcut connection, and x_i is convolved and activated by the activation function G_i to output the residual value $F(x_i, G_i)$.

$$x_{i+1} = F(x_i, G_i) + W x_i. \tag{1}$$

Here, x_i is the data input of the residual block at layer i, G_i is the activation function, $F(x_i, G_i)$ is the residual value, W is the constant mapping parameter (usually 1), and x_{i+1} is the input of the residual block at layer ☒ + 1.

Any layer x_k deeper than layer x_i can be represented as x_i (Equation (2)). The gradient dl/dx_i of the loss function for x_i is found using Equation (3) with W as one.

$$x_k = \sum_{j=i}^{k-1} F(x_i, G_i) + x_i, \tag{2}$$

$$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial x_k}\frac{\partial x_k}{\partial \mathrm{x_i}}. \tag{3}$$

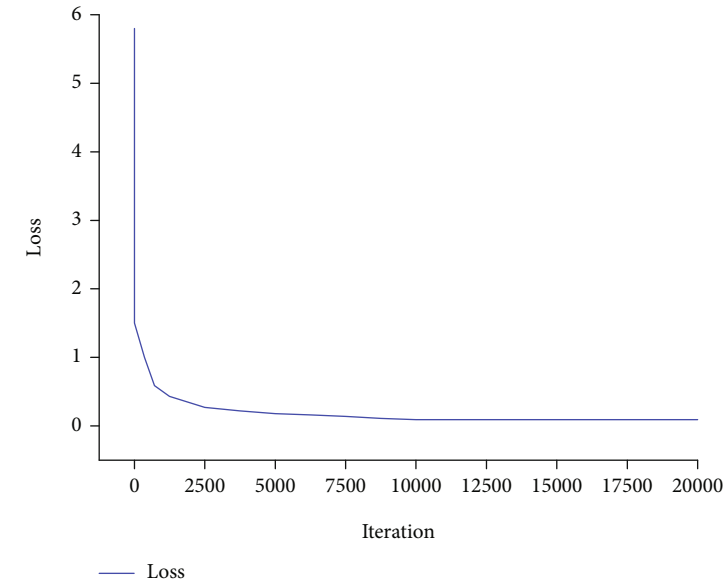
Here, x_k is the residual block input value of layer $k(k > i)$, x_j is the residual value from layer j to layer $j - 1$, G_j is the activation function from layer i to layer $k - 1$, and $F(x_i, G_i)$ is the residual value from layer i to layer $k - 1$.

Equation (2) is substituted into Equation (3) to obtain Equation (4). Equation (4) shows that the gradient of the deeper layer x_k can be transferred to any layer x_i that is shallower than x_k. In addition, as the constant one in the parentheses and the product calculation in the conduction become a summation, its gradient never disappears regardless of the depth of the network layers.

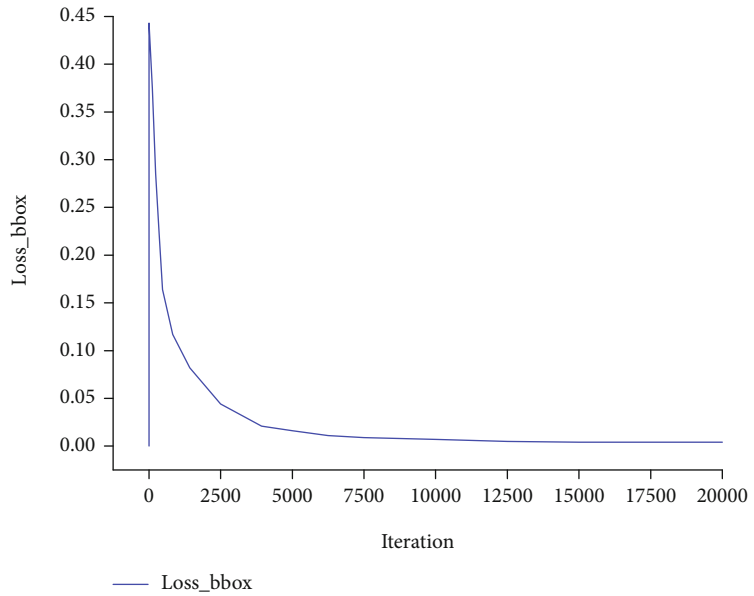
$$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial x_k}\frac{\partial\left(\sum_{j=i}^{k-1} F(x_i, G_i) + x_i\right)}{\partial x_i} = \frac{\partial l}{\partial x_k}\left(1 + \frac{\partial \sum_{j=i}^{k-1} F(x_i, G_i)}{\partial x_i}\right). \tag{4}$$

3.2.3. Multiscale Processing. Usually, image targets have different scale sizes at the time of acquisition, and the objects in the dataset have different scales and sizes. A general dataset does not capture all image attributes, with objects of sizes smaller than the step size of the convolutional network not recognized effectively because they are ignored; this phenomenon is known as the multiscale problem. Vertebrae exist in different sizes, thus causing multiscale problems in training and recognition scenarios. FPNs using Fast R-CNN can effectively solve the multiscale problem [22–24]. As shown in Figure 6, FPNs employ a double pyramidal structure where they use lateral connections and top-down paths together through a simple merging layer, which is highly effective for feature processing.

The left pyramid is a conventional feature map CNN, and the feature map shrinks layer by layer with the number of convolutions. The classification task cannot be completed because the initial semantic information and features obtained from the bottom feature map after convolutional computation feature extraction in the initial convolutional layer are not sufficiently strong. The deep feature map has stronger semantic information and features; FPN exploits this key feature to capture stronger semantic information from the deep feature map and use it for the classification task through the path connections of the pyramid. However, the middle pyramid reverse samples the convolved image from top to bottom and enhances the expression of spatial and semantic information at different scales for each layer on the right side of the pyramid to detect small-scale targets by scaling up the feature map and then summing the elements with each pixel point of the feature map on the left side of the next layer without increasing the number of operations. The FPN takes the feature maps of images to the deepest layer of the network through a top-down connection path.

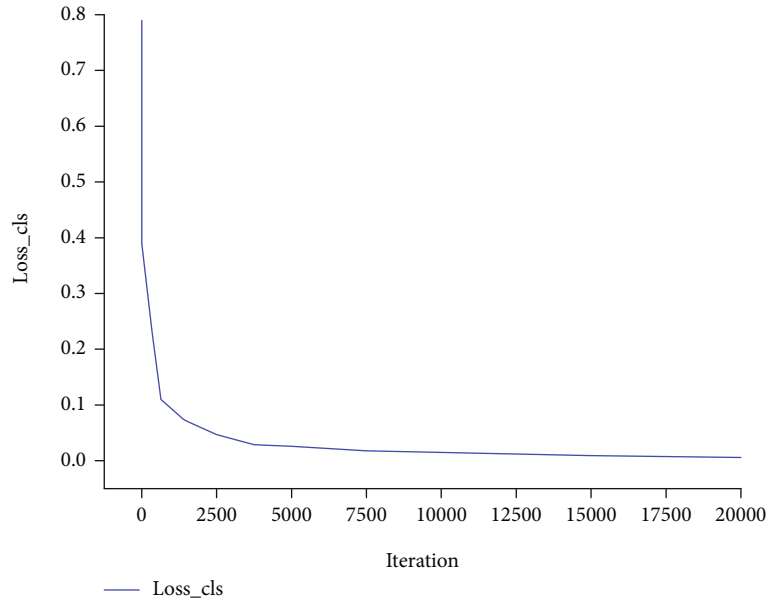


(a) Loss function value

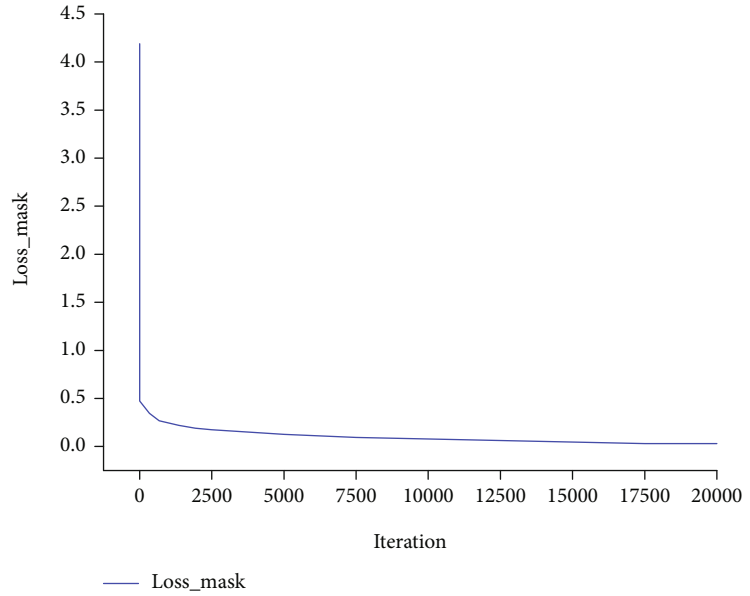


(b) Regression loss function

Figure 7: Continued.



(c) Classification loss function



(d) Mask loss function

Figure 7: Change in loss values for frontal image training set.

The FPN architecture cleverly combines lateral and top-down connectivity paths, allowing the extraction of deeper feature semantic information from images, feature fusion, and multiscale feature output, preventing the loss of semantic information caused by the limitations of existing processing. In scenarios where vertebral bone recognition is performed, the FPN model can exploit its multiscale recognition and increase the probability of detecting small-scale vertebral bone objects in images.

3.2.4. Loss Calculations. After feature region extraction and screening, the network head separately calculates the mask, box, and classification loss values for the feature regions. The classification loss value is the logarithmic loss that divides the object and nonobject classes after binary discrimination of the target. The segmentation error is the loss calculated for each class of the mask. The innovation of the segmentation error is that if the region of interest is detected to belong to a certain category, the relative entropy error of that category is used for the calculation. Kaiming et al. empirically demonstrated that selecting the sum of the classification, box, and mask loss values (Equation (5)) as the overall loss value in model training yields good results for multiple datasets [25].

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}}. \tag{5}$$

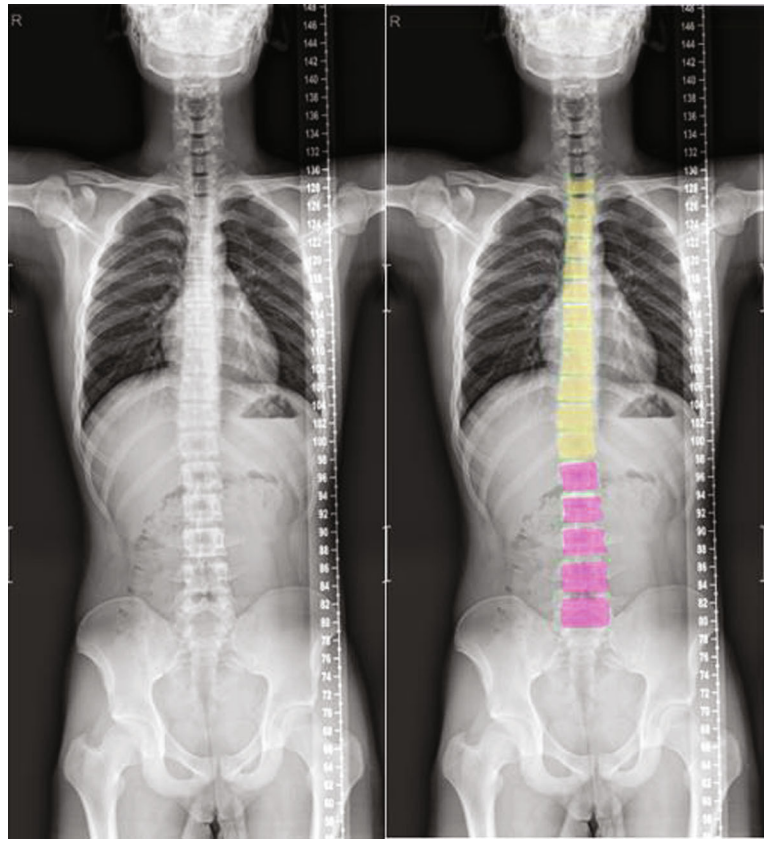


Figure 8: Example of test results of frontal image semantic segmentation.

Here, L_{cls} is the loss value calculated using classification, L_{box} is the loss value calculated using the bounding-box, L_{mask} is the loss value calculated using the mask, and L is the overall loss value of the network.

4. Results and Discussion

4.1. Experimental Environment

4.1.1. Database Construction. The human spine images acquired for the experiments were anonymized and divided into frontal and lateral views. One category comprised human X-ray images converted from DCM to JPG format using ImageJ; these images were used to study the images to be processed. The other category comprised ground-truth images obtained by converting DCM to JPG format using ImageJ, adding mask annotation using LabelMe image annotation software with alignment using ImageJ, and generating mask images and mask annotated images required for model training. The images of the 12 thoracic vertebrae and the 5 lumbar vertebrae were labeled, with different vertebrae being labeled with different colors. Labeling was performed based on vertebra identification according to human anatomical morphology, starting from T1 and labeling the thoracic vertebrae (T1-T12) and the lumbar vertebrae (L1-L5).

4.1.2. Data Settings and Augmentation. In this study, there were 250 frontal and lateral X-ray images of the human spine. In order to obtain standard X-ray images, we took the following measures. First, we minimized the overlap between the bone and soft tissue of the shoulder joints and the upper thoracic spine by standardizing the position and method of radiography. Second, we carefully observed each image and reject images with artifacts. Finally, we adjusted the contrast of the images to obtain the best experimental results.

ImageJ was used for grayscale adjustment to facilitate annotation. The data were created in the COCO dataset format, and the dataset was divided into 170 images for the training set in the frontal and lateral position, 6 images for the validation set in the frontal position, 5 images for the validation set in the lateral position, and 70 images for the test set.

4.1.3. Data Settings and Augmentation Experimental Platform Construction and Environment Configuration. Python programming language and Linux operating system were used in this study. The environment was configured to use Python, PyTorch, and CUDA to build the Mask R-CNN models using a ResNet-50 convolutional backbone. Two GPUs were used for training.

4.2. Analysis of Experimental Results

4.2.1. Frontal Spine Radiograph Experiments. In this study, we first trained and tested frontal spine radiographs, and the model was trained with the Mask R-CNN on the existing training set. Figure 7 shows the changes in loss values.

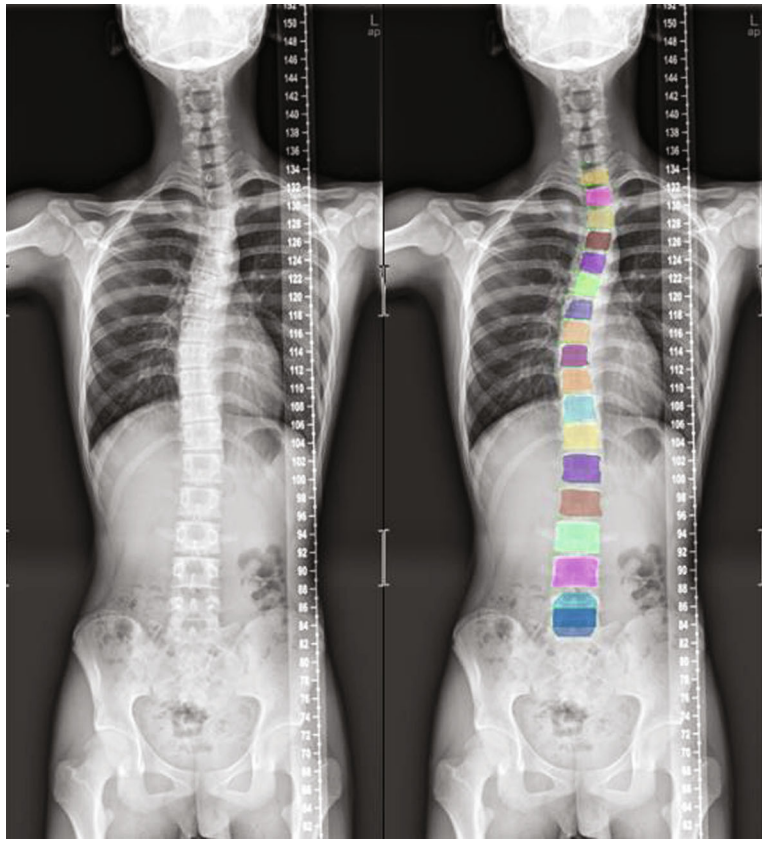


Figure 9: Example of test results of frontal image instance segmentation.

Table 1: Parameters for evaluating the experimental results of the spine model positioning (frontal images).

	Accuracy	IoU = 0.5 Average precision (AP50)	Average recall (AR)
Spine segmentation	99.8%	96.8%	62.2%
Detection box	99.8%	97.4%	70.5%

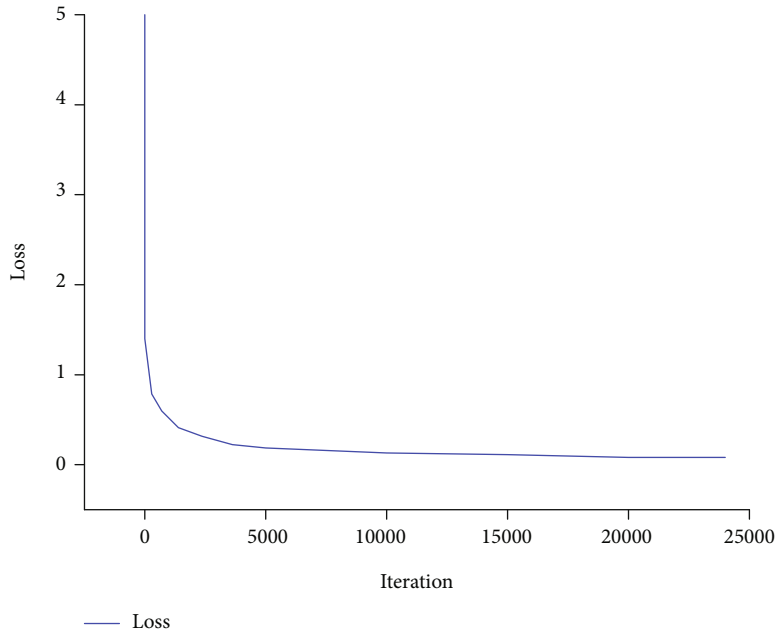
Note: average recall (AR) represents the average IoU in the range of 0.5-0.95 at intervals of 0.05.

The experimental results show that the loss function, mask loss function, regression loss function, and classification loss function converge after approximately 10,000 iterations of training on frontal radiographs. After model training, frontal spine radiographs were tested, the test set was selected, and the results were obtained, as shown in Figures 8 and 9. For spine model positioning, the vertebrae were identified during object detection, with semantic and instance segmentation of the vertebrae also achieved at the thoracic and lumbar levels. As shown in Figure 8, the thoracic vertebrae (T1-T12) and the lumbar vertebrae (L1-L5) were individually segmented in the semantic segmentation, and the detection box was distinguished between the thoracic vertebrae and lumbar vertebrae. In the instance segmentation in Figure 9, the vertebrae are segmented into different colors, and the detection box distinguishes between the thoracic vertebrae and lumbar vertebrae.

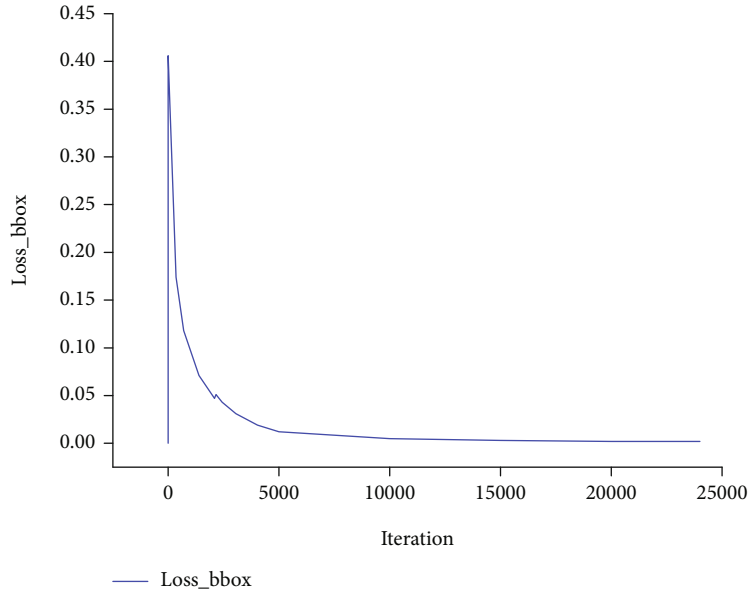
Table 1 shows the values of the average recall (AR, the average value at a threshold of 0.5-0.95 as the IoU), the average precision (AP) 50 (at an IoU of 0.50), and accuracy for the segmentation and object detection box using the ResNet-50 CNN. The table indicates good test results with the frontal radiographs, with the accuracy of both vertebral segmentation and vertebral object detection boxes reaching 99.8%. The average accuracy of vertebral segmentation was 96.8%, and the AR was 62.2%. The average accuracy of the object detection box was 97.4%, and the AR was 70.5%. As the vertebrae in the frontal human spine radiographs are relatively clearly visible and virtually completely unobstructed by tissue or bone, it had good performance in both the labeling and training tests.

Spine model positioning of frontal spine radiographs based on the Mask R-CNN architecture provides good performance for recognition of vertebrae, including detection of frontal spine radiographs by segmentation and the object detection box.

4.2.2. Lateral Spine Radiograph Experiments. In this study, lateral spine radiographs continued to be used as test objects. While labeling the lateral spine radiographs, we found that some of the vertebral structures were almost invisible because they were obscured by the body, especially in the upper thoracic vertebrae, where the vertebral column was nearly invisible

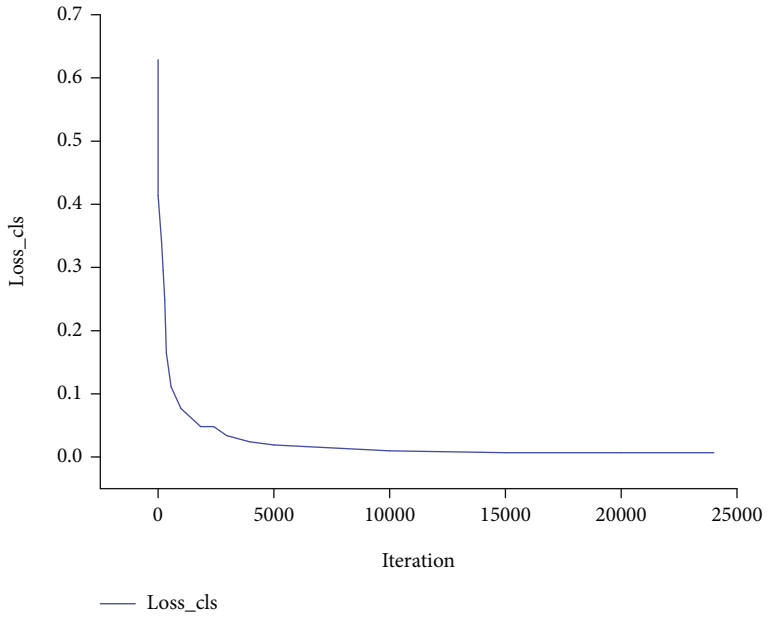


(a) Loss function value

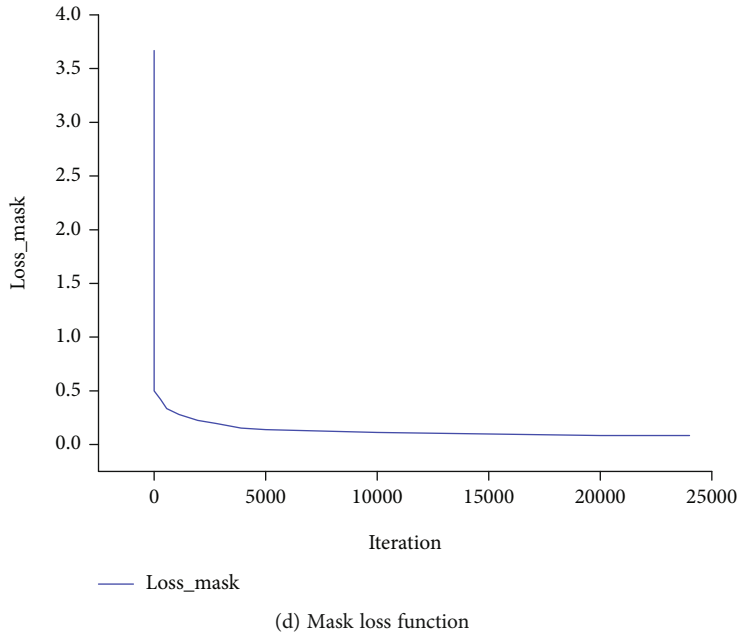


(b) Regression loss function

Figure 10: Continued.



(c) Classification loss function



(d) Mask loss function

Figure 10: Change in loss values for lateral image training set.

during labeling and could only be judged by the scale bar. This made it difficult to label, train, and make predictions using lateral radiographs, thus increasing the difficulty of the experiment.

The model was trained with the Mask R-CNN on the existing training set of lateral radiographs. Figure 10 shows the changes in loss values.

After 24000 iterations of model training, the test set of lateral spine slices was tested; the results obtained are shown in Figures 11 and 12. The test results of spine model positioning based on the Mask R-CNN architecture, including detection of lateral spine radiographs by segmentation and the object detection box, were unsatisfactory. Although the vertebrae can be identified in object detection and semantic and instance segmentation of the thoracic and lumbar vertebrae was performed, there were cases of incomplete image segmentation. In addition, although the object prediction box predicted the vertebra position correctly, the classification of thoracic and lumbar vertebrae was incorrectly predicted.

Table 2 shows the values of the AR (the average value at a threshold of 0.5-0.95 as the IoU), the AP75 (at an IoU of 0.75), and accuracy for the segmentation and object detection box using the ResNet-50 CNN. The table indicates satisfactory test result parameters with the lateral radiographs, with an average accuracy of vertebral segmentation of 81.6% and an AR of 63.8%. The average accuracy of the object detection box was 90.1%, and the AR was 90.0%.

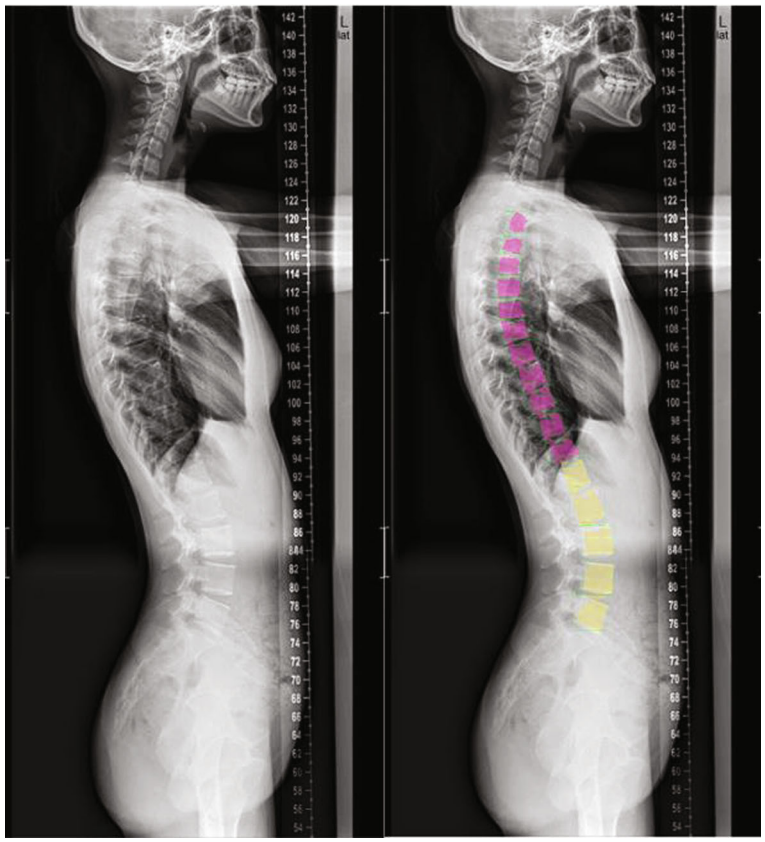


Figure 11: Example of test results of lateral image semantic segmentation.

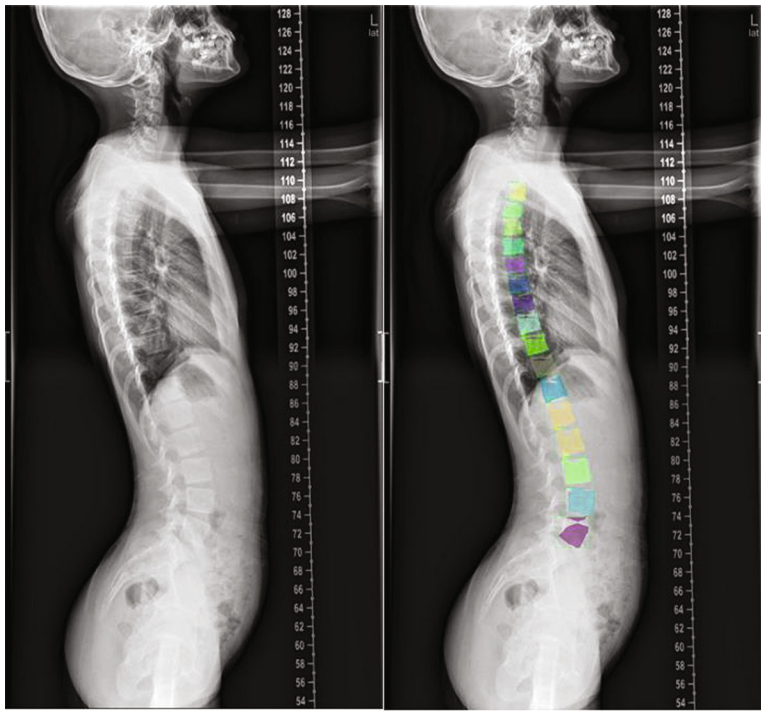


Figure 12: Example of test results of lateral image instance segmentation.

Table 2: Parameters for evaluating the experimental results of spine model positioning (lateral images).

	Accuracy	IoU = 0.75 Average precision (AP75)	Average recall (AR)
Spine segmentation	99.8%	81.6%	63.8%
Bounding box	99.8%	90.1%	90.0%

Note: average recall (AR) represents the average IoU in the range of 0.5-0.95 at intervals of 0.05.

The experimental results of spinal model positioning of lateral images are analyzed as follows:

(1) In clinical work, because the first to third thoracic vertebrae overlap with the shoulder joints on both sides in the lateral radiographs, sometimes, part of the thoracic vertebrae cannot be clearly displayed. In this case, it is difficult for even experienced radiologists to accurately label this part of the thoracic vertebrae on the lateral radiographs. Consequently, there were difficulties in labeling, training, and prediction in the experiments. The loss was still approximately 0.08 after training for 24000 iterations, and the rate of decline was also slow. This made the final prediction results not reach the best state, resulting in inaccurate segmentation and classification reversal errors

(2) In this study, the training of lateral images required twice as many iterations as frontal images to achieve a similar loss. The loss function was not yet optimal, and the existing experimental results could not yield completely correct results for the positioning of the lateral image spine model. The prediction results have some incomplete skeletal segmentation, and the reason for the reversed thoracic and lumbar classification prediction warrants further investigation. To solve the problems, more training or comparison iterations or a more efficient training backbone network may be required

(3) The dataset partitioning performed in this study still needs to be improved as the excessive number of training sets without a reasonable proportion of datasets may lead to errors in the evaluation parameters. In future studies, a more reasonable dataset division for training, testing, and validation should be performed. In addition, constructing a larger database will also help improve the accuracy of experimental results

5. Conclusions

In this study, we performed human X-ray image spine model positioning based on an R-CNN. We focused on object detection and image segmentation and the effect of Mask R-CNN on human radiograph spine model positioning, detection, and segmentation. Considering application directions, we focused on radiographic spine model positioning.

Python language was used to configure Mask R-CNN, including FPN with ResNet-50 and other network architectures, to conduct extensive training on a large amount of data from frontal and lateral slices and testing on spine radiographs. The segmentation and detection results of many bony structures, including L1-L5 in the lumbar spine and T1-T12 in the thoracic spine, were obtained in a single step. Considering the experimental results, the prediction efficacy for frontal images and image visualization was good, with an average detection box accuracy of 97.4% and an average segmentation accuracy of 96.8%. The results for lateral images were satisfactory considering the evaluation parameters and image visualization, but the problem of classification and prediction errors warrants further attention. Overall, the study provides a basis for spine reconstruction, scoliosis classification, and scoliosis prediction and adds value to the research on automatic spine model positioning.

Based on the characteristics of the vertebrae in lateral radiographs of the spine, there are still many shortcomings in this study that must be addressed. In this study, the lateral images were often not as easily observed because the body blocked the thoracic spine, leading to inaccurate labeling, which affected training and testing. We reduced errors by standardizing the photographic position, adjusting the contrast of the images, and using the scale bar. There were still shortcomings in the positioning and segmentation of the lateral images in this study. In the future, more network architectures such as ResNet-101 and VGG neural networks should be used as convolutional backbones for comparison, with the network parameters adjusted more carefully to select the optimal network architecture to solve the problems of incorrect classification and segmentation of lateral radiographs.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] D. C. Grossman, S. J. Curry, D. K. Owens et al., "Screening for adolescent idiopathic scoliosis: US preventive services task force recommendation statement," *JAMA The Journal of the American Medical Association*, vol. 319, no. 2, pp. 165–172, 2018.

[2] C. E. Mackel, A. Jada, A. F. Samdani et al., "A comprehensive review of the diagnosis and management of congenital scoliosis," *Child's Nervous System*, vol. 34, no. 11, pp. 2155–2171, 2018.

[3] J. M. Pahys and J. T. Guille, "What's new in congenital scoliosis?," *Journal of Pediatric Orthopedics*, vol. 38, no. 3, pp. e172–e179, 2018.

[4] D. Hedequist and J. Emans, "Congenital scoliosis," *Journal of Pediatric Orthopedics*, vol. 27, no. 1, pp. 106–116, 2007.

[5] S. Kumar, K. P. Nayak, and K. S. Hareesha, "Combined SCP and geometric reconstruction of spine from biplanar radiographs," in *Proceedings of Fifth International Conference on Soft Computing for Problem Solving*, vol. 437 of *Advances in Intelligent Systems and Computing*, pp. 215–225, Singapore, 2016.

[6] S. M. R. A. Arif, K. Knapp, and G. Slabaugh, "Fully automatic cervical vertebrae segmentation framework for X-ray images," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 95–111, 2018.

[7] T. L. Bredbenner, T. D. Eliason, F. W. Loren, J. M. McFarland, A. C. Merkle, and D. P. Nicolella, "Development and validation of a statistical shape modeling-based finite element model of the cervical spine under low-level multiple direction loading conditions," *Frontiers in Bioengineering & Biotechnology*, vol. 2, p. 58, 2014.

[8] T. Cresson, B. Godbout, D. Branchaud, R. Chav, P. Gravel, and J. A. De Guise, "Surface reconstruction from planar X-ray images using moving least squares," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3967–3970, Vancouver, BC, Canada, 2008.

[9] E. J. Bekkers and C. A. Taylor, "Multiscale vascular surface model generation from medical imaging data using hierarchical features," *IEEE Transactions on Medical Imaging*, vol. 27, no. 3, pp. 331–341, 2008.

[10] D. Fazanaro, P. Amorim, T. Moraes, J. Silva, and H. Pedrini, "NURBS parameterization for medical surface reconstruction," *Applied Mathematics*, vol. 7, no. 2, pp. 137–144, 2016.

[11] A. Mitulescu, W. Skalli, D. Mitton, and J. de Guise, "Three-dimensional surface rendering reconstruction of scoliotic vertebrae using a non stereo-corresponding points technique," *European Spine Journal*, vol. 11, no. 4, pp. 344–352, 2002.

[12] J. Boisvert and D. C. Moura, "Interactive 3D reconstruction of the spine from radiographs using a statistical shape model and second-order cone programming," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5726–5729, Boston, MA, USA, 2011.

[13] D. C. Moura, J. Boisvert, J. G. Barbosa, H. Labelle, and J. M. R. S. Tavares, "Fast 3D reconstruction of the spine from biplanar radiographs using a deformable articulated model," *Medical Engineering & Physics*, vol. 33, no. 8, pp. 924–933, 2011.

[14] L. Humbert, J. A. D. Guise, B. Aubert, B. Godbout, and W. Skalli, "3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences," *Medical Engineering & Physics*, vol. 31, no. 6, pp. 681–687, 2009.

[15] B. Ilharreborde, J. S. Steffen, E. Nectoux et al., "Angle measurement reproducibility using EOSthree-dimensional reconstructions in adolescent idiopathic scoliosis treated by posterior instrumentation," *Spine*, vol. 36, no. 20, pp. E1306–E1313, 2011.

[16] C. Bakhous, B. Aubert, C. Vazquez, T. Cresson, S. Parent, and J. De Guise, "Automatic pedicles detection using convolutional neural network in a 3D spine reconstruction from biplanar radiographs," *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575, 2018.

[17] B. Aubert, C. Vazquez, T. Cresson, S. Parent, and J. A. de Guise, "Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting," *IEEE Transactions on Medical Imaging.*, vol. 38, no. 12, pp. 2796–2806, 2019.

[18] L. Liu, W. Ouyang, X. Wang et al., "Deep learning for generic object detection: a survey," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, 2020.

[19] K. He, Z. Xiangyu, R. Shaoqing, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, 2016.

[20] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Thirty-first AAAI Conference on Artificial Intelligence*, pp. 4278–4284, San Francisco, California, USA, 2017.

[21] W. Zifeng, S. Chunhua, and A. V. D. Hengel, "Wider or deeper: revisiting the ResNet model for visual recognition," *Pattern Recognition*, vol. 90, pp. 119–133, 2019.

[22] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, Hawaii, USA, 2017.

[23] Z. Liang, J. Shao, D. Zhang, and L. Gao, "Small object detection using deep feature pyramid networks," in *Pacific Rim Conference on Multimedia*, pp. 554–564, Hefei, China, 2018.

[24] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6399–6408, California, USA, 2019.

[25] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2017.

Hindawi
BioMed Research International
Volume 2022, Article ID 1485584, 8 pages
https://doi.org/10.1155/2022/1485584

Research Article

A Heart Segmentation Algorithm Based on Dynamic Ultrasound

Mingjun Tian and Minjuan Zheng

Department of Ultrasound, Xijing Hospital, Fourth Military Medical University, Xi'an 710032, China

Correspondence should be addressed to Minjuan Zheng; zhengmj@fmmu.edu.cn

Received 26 April 2022; Revised 23 May 2022; Accepted 7 June 2022; Published 17 June 2022

Academic Editor: Chen Li

Copyright © 2022 Mingjun Tian and Minjuan Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The heart is one of the most important organs of the human body. The role of the heart is to promote blood flow and provide sufficient blood flow to organs and tissues. The research on the heart has important theoretical and clinical significance. Because of the noninvasive and intuitive display of ultrasound image, it can dynamically obtain the heart state and has become the main means to detect the heart dynamics. We analyze the characteristics of cardiac ultrasound image from the medical point of view and signal processing. The heart movement is periodic and rhythmic. The image signal can be decomposed. Firstly, the image is decomposed into high- and low-frequency signals to highlight different dimensional information. Then, the attention model was introduced, focusing on the heart region. Finally, the multidimensional network carrying model was established to achieve cardiac segmentation. The experimental results show that the AOM of the algorithm proposed in this paper reaches 92%, which has a certain degree of advancement and can assist doctors to make accurate diagnosis.

1. Introduction

The heart is an important organ of the human body. Its main function is to provide pressure for blood flow and run the blood to all parts of the body. Its efficiency is related to the shape and size of the heart [1]. Therefore, it is very important to extract the region of the heart and carry out research. Tang et al. [2] proposed a clustering algorithm for heart sound segmentation. Icardo [3] established a model from morphology to analyze the heart. Ahn et al. [4] focused on the region of coronary artery through CT and carried out research. Mythili et al. [5] used SVM to predict heart disease. Pedrosa et al. [6] built a model to analyze the voice of young children. Methaila et al. [7] used the big data technology to analyze early heart disease. Pace et al. [8] implemented the whole heart segmentation interactively. Saquib et al. [9] diagnosed heart disease by calculating the volume ratio. Xiong et al. [10] combined CT images with clinical data for analysis. Wolterink et al. [11] used dilated convolutional neural networks to segment MR images. Arabasadi et al. [12] used a hybrid neural network genetic algorithm to predict heart disease. Tong et al. [13] constructed 3D deep supervised U-Net to segment all hearts. Ahmed et al. [14] built a deep network to segment the heart. Gao and Lu [15] focused on fetal baseline to realize classification and extraction. Xu et al. [16] combined a deep learning network and graph matching to realize whole heart segmentation. de Albuquerque et al. [17] proposed fast heart fat segmentation based on CT data set. Naseer et al. [18] constructed fuzzy sets to diagnose heart diseases. Yoshida et al. [19] used U-Net to automatically segment the heart based on CT data. Banerjee et al. [20] reconstructed the heart in 3D from 2D data. Diniz et al. [21] built Concat-U-Net to realize automatic heart extraction. Diniz et al. [22] built U-Net++ to realize heart segmentation. Liu et al. [23] proposed automatic segmentation algorithm using attentional convolutional network. Chen et al. [24] constructed 3D filter to suppress ultrasonic image noise. Song et al. [25] proposed deep networks for heart segmentation and explained the significance and challenges of heart segmentation.

Through the above introduction, the current analysis of the heart is focused on mature images with high imaging quality such as CT [26] and MR [27]. However, CT and MR mostly present static images, which cannot meet the situation of dynamic analysis. Therefore, medical staff often use dynamic echocardiography in analyzing [28] and has achieved good results. Echocardiography is the most widely used cardiac examination method in clinic, which can dynamically evaluate the structure and function of the heart. The ventricular wall motion of the heart is an important driving force to maintain the function of the heart pump. It is very important for doctors to identify whether the ventricular wall motion is normal based on ultrasonic images. The current ventricular wall motion evaluation is mainly by eye or manual trace for motion amplitude, which is very dependent on the operator's experience and time-consuming. If the segmental recognition method of ventricular wall can be established to quickly identify the abnormal motion of myocardium, it would be very helpful for the diagnosis of heart disease.

However, due to the large noise interference of ultrasonic image and the lower imaging quality than CT and MR, it is difficult to put forward computer-aided segmentation. In general, main problems are as follows: (1) low image quality leads to limited access to information. (2) The heart only occupies a limited area in the image, which is inefficient to process the image pixels in a unified scale. (3) The constructed cardiac ultrasound image feature network has limited carrying capacity and insufficient representation.

In view of the above shortcomings, we propose a new segmentation algorithm based on dynamic echocardiography: (1) Octconv is proposed from the perspective of signal composition, and the signal is decomposed into high- and low-frequency parts to extract different frequency information; (2) to simulate the doctor's diagnosis process and build an attention mechanism to focus on the region of the heart; and (3) from the perspective of deep learning, a multidimensional information carrying network is constructed to fuse low-frequency and high-frequency signals to achieve heart segmentation.

2. Algorithm

The research of echocardiography algorithm is based on the current situation of low echocardiography quality and the actual situation of echocardiography clinical diagnosis. A dynamic echocardiography segmentation algorithm based on deep learning is constructed, and the specific block diagram is shown in Figure 1. Firstly, Octconv is proposed to decompose the image into low-frequency and high-frequency signals by simulating the principle of visual perception. Then, an attention mechanism is built to focus on the region where the heart is located. Finally, a multidimensional information carrying network is built to realize heart segmentation.

2.1. Image Preprocessing. Evaluating the motion of ventricular wall is an important content of echocardiography. The left ventricle is the main chamber responsible for pumping blood. Usually, the left ventricular wall is divided into 16 or 17 segments (including the apex) according American Society of Echocardiography (ASE) guidelines. Each segment was dominated by its corresponding coronary artery branches for blood supply. Segmental wall motion abnormalities can occur in coronary heart disease or other myocardial diseases (diabetes, hypertension, amyloidosis, etc.); thus, each LV segment is needed to be evaluated. This cardiac segmentation can help doctors quickly identify the region of ventricular wall with abnormal motion, so as to accurately locate the location of impaired myocardium.

We used video sequences to do our research. Therefore, the detection region is extracted from the pixel variation of the whole video sequence, and the sector region can be preliminarily determined. To this end, we build the maximum density projection mechanism to obtain the information of the sequence image and focus on the image region, as shown in Figure 2.

The projection of maximum density is the maximum value of the coordinates at (x, y) and can be used to calculate the area of concern.

$$M(x, y) = \max\left(I_1(x, y), \cdots I_n(x, y)\right). \tag{1}$$

After the maximum density projection of the image, the threshold method was used to determine detection area, and then, the ultrasonic image region was obtained according to mathematical morphology operation.

2.2. Octconv. Information can be transmitted at different frequencies, and the characteristic diagram output by the convolution layer can also be regarded as the combination of information at different frequencies. Therefore, scholars have proposed Octconv to store and process feature graphs with low spatial resolution and slow spatial change. Octconv has orthogonality and complementarity and can establish a better topology to reduce the spatial redundancy caused by low-frequency information in deep convolution network. The specific structure is shown in Figure 3. α_{in} is the proportion of input low-frequency channel in the total channels. The corresponding input size is $c_{\text{in}} \times w \times h$, and the output size is $c_{\text{out}} \times w \times h$. W is the convolution kernel, and the corresponding parameter is $c_{\text{in}} \times c_{\text{out}} \times k \times k$, stride = 1, and padding = same. The most available calculation amount is $(c_{\text{in}} \times k \times k) \times (c_{\text{out}} \times w \times h)$. Compared with ordinary convolution, Octconv contains a certain proportion of low-frequency channels, and the complexity is improved. But the performance is better than ordinary convolution.

Octconv uses the coefficient α to append the feature graph into high-resolution component (X^{H}) and low-resolution component (X^{L}). X^{H} saves the edge, contour, and other details of the images with a large amount of information data. X^{L} saves the abstract information of the images with a small amount of information data. The multifrequency feature representation method is constructed, and the smooth low-frequency mapping is stored in the low-frequency tensor to reduce spatial redundancy. The specified

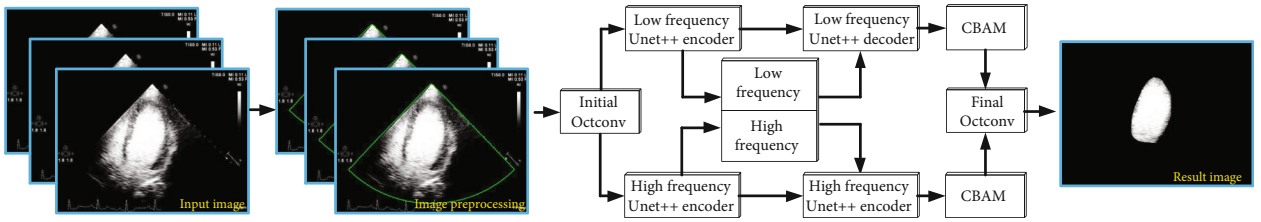


Figure 1: Algorithm flowchart.

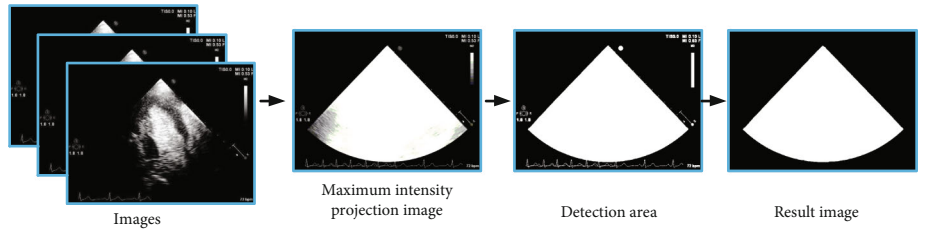


Figure 2: Image preprocessing flowchart.

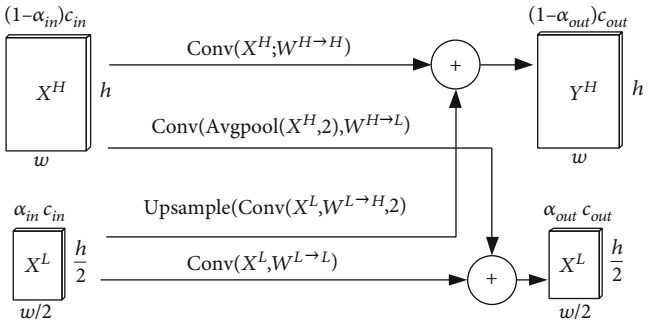


Figure 3: Octconv composition.

convolution kernel w is used to convolute x to obtain the corresponding component:

$$\begin{aligned} Y^{H\longrightarrow H} &= \mathrm{Conv}\left(X^H, W^{H\longrightarrow H}\right), \\ Y^{H\longrightarrow L} &= \mathrm{Conv}\left(\mathrm{Avgpool}\left(X^H, 2\right), W^{H\longrightarrow L}\right), \\ Y^{L\longrightarrow H} &= \mathrm{upsample}\left(\mathrm{Conv}\left(X^L, W^{L\longrightarrow H}\right), 2\right), \\ Y^{L\longrightarrow L} &= \mathrm{Conv}\left(X^L, W^{L\longrightarrow L}\right). \end{aligned} \tag{2}$$

High-frequency output is $Y^H = Y^{H\longrightarrow H} + Y^{L\longrightarrow H}$, and low-frequency output is $Y^L = Y^{H\longrightarrow L} + Y^{L\longrightarrow L}$. The amount of Octconv convolution calculation is only 1/4 of that of ordinary conv.

2.3. CBAM Attention Module. Due to the increase of the length of neural network, the long-distance information will be weakened, resulting in the loss of important information after information transmission. To reduce the risk of losing important information, we focus our limited attention on important information to build attention modules. Due to the convolutional block attention module (CBAM) combining channel and space, the channel attention module and spatial attention module are connected in series. According to the importance of features, they focus on the area of interest step by step and extract key information through corresponding spatial transformation. The structure is shown in Figure 4.

(1) Channel attention module. Input feature $F(H\times W\times C)$ and two channel features $(1\times 1\times C)$ are obtained through Maxpool and Avgpool to connect MLP (multilayer perceptron). The corresponding number of neurons in the first layer is C/r, and the number of neurons in the second layer is c. The weight coefficient of the activation function is obtained as follows:

$$M_c(F) = \mathrm{Sigmoid}(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \tag{3}$$

The corresponding new feature is

$$F' = M_C(F)F. \tag{4}$$

(2) Spatial attention module. Maxpool and Avgpool are performed on F' to get the channel description, and the dimension is $H\times W\times 1$. Splice the two descriptions according to the channel, and pass through 7×7 convolution, and the weight function is

$$M_s\left(F'\right) = \mathrm{Sigmoid}\left\{f\left(\mathrm{AvgPool}\left(F'\right), \mathrm{MaxPool}\left(F'\right)\right)\right\} \tag{5}$$

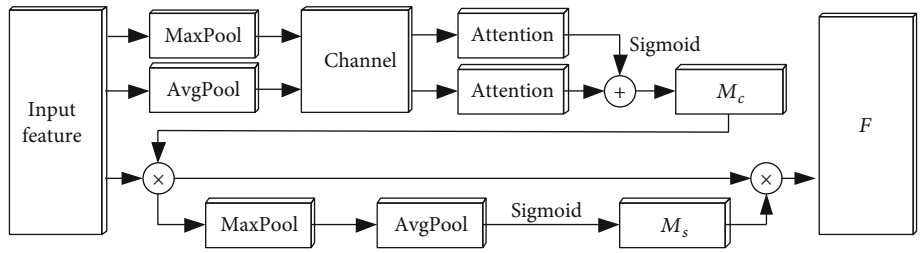


Figure 4: CBAM structure.

The corresponding new feature is

$$F'' = M_s\left(F'\right)F', \tag{6}$$

where M_c and M_s form a complementary relationship. The use of Maxpool and Avgpool increases more diversified information, resulting in less computation and stable performance improvement of CBAM.

2.4. Network Model. The main advantage of U-Net++ is that it connects each branch U-Net to share a coding layer and allows the middle part of the model to participate in training, so that the information loss in the encoder process can be repaired to some extent. The performance is improved on the basis of limited increase of parameters. Since each branch U-Net shares an encoder path, the same information is lost in the downsampling process. U-Net++ obtains U-Net features with different depths in the decoder process, but this feature difference is obtained after each branch U-Net performs its own upsampling, so the information supplement of U-Net++ in the feature recovery process is limited and has lack of pertinence.

Through the above analysis, there are five layers in total based on the U-Net++ framework. Octconv is used to replace the traditional Conv2D. The whole framework is divided into two U-Net++ structure branches of high-frequency and low-frequency synchronous parallel. Octconv is used in each layer of coding-decoding to exchange low-frequency and high-frequency information. It reduces the model parameters, reduces the information loss introduced by each coding layer in downsampling, and makes the whole network obtain more abundant information. In the decoding stage, bilinear interpolation sampling is used to realize image restoration. After each sampling, it is spliced with the features of the same layer and the same scale and then connected with CBAM to strengthen the attention of convolution operation in the model to the target area and realize more accurate pixel category classification. The problem of offset of restoration features is avoided as much as possible in order to make the network more robust.

Init-Octconv and input high frequency obtain characteristics through downsampling, and then, Octconv is performed to obtain high-frequency and low-frequency output. COM-Octconv is input by high-frequency and low-frequency characteristics, and Octconv is directly carried out through hyperparameter α to control the proportion of high- and low-frequency channels of output. Fin-Octconv is input by high-frequency and low-frequency characteristics. After convolution, the low-frequency upsampling is added to the high frequency to output the high frequency.

We introduce the objective function. Tversky coefficient is the generalized coefficient of Dice coefficient and Jaccard coefficient:

$$TI(X, Y) = \frac{X \cap Y}{X \cup Y + \beta|X - Y| + \alpha|X + Y|}, \tag{7}$$

where X represents the true value and Y represents the predicted value. In medical image segmentation, Dice Loss ($\alpha = \beta = 0.5$) is often used in small lesion segmentation as the objective function. It has good performance in the case of extremely unbalanced samples, but in general, its use will have an adverse impact on back propagation and make the training unstable. As a result, the effect of Dice Loss is not ideal and the fluctuation range is large.

Therefore, we construct a new objective function to reduce the weight of simple samples and increase the weight of difficult samples. γ coefficient is introduced to learn difficult samples with small sample regions of interest:

$$FTL = \sum_c (1 - TI_c)^\gamma, \gamma \in [1, 3]. \tag{8}$$

3. Experiment and Result Analysis

The experiment is conducted with 50 groups of image sequences collected by the hospital, and the data resolution is 600×800, and the equipment is Philips Epiq 7c. As shown in Figure 5, the red box is the continuous image sequence. The area where the heart is located is marked at the pixel level by a professional doctor. We adopt double-blind method, which is a common algorithm in annotation field. The gold standard was determined by combining the labeling results of two physicians. The programming environment is Linux and Python. We adopted the deep learning network. So the input of the deep learning network we designed is 600×800. For other resolutions, normalize to 600×800. We built 1 : 1 ratio of training samples to test samples. The cross-validation method is adopted.

3.1. Octconv Performance. Based on the principle of signal processing, we decompose signals into low-frequency and high-frequency components to construct Octconv. To verify Octconv performance, we build a unified network model

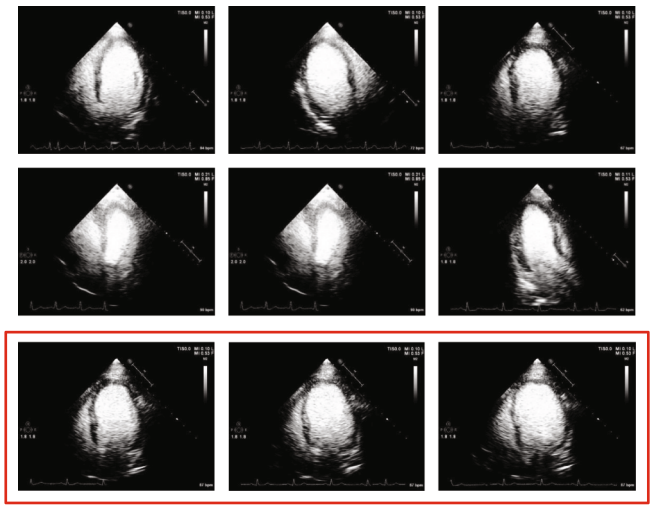


Figure 5: Data display.

and compare the convergence curves of Octconv and ordinary convolution to obtain the performance. AOM is introduced to measure the convergence effect.

$$\mathrm{AOM} = \frac{A \cap B}{A \cup B}. \tag{9}$$

As shown in Figure 6, Octconv has better iteration times than traditional convolution when the curve converges. Due to the orthogonality and complementarity of Octconv, signals are decomposed into low-frequency and high-frequency components, which is in line with the principle of visual sensing and has low redundancy. However, the traditional convolution only considers the relationship between parts, resulting in a large amount of redundancy in the parameters and slow convergence speed.

3.2. Segmentation Effect Comparison. The loss function, crossentropy loss function, Dice Loss, and Tversky Loss are proposed as the objective function for experiments. We introduce AVM, AUM, and CM to measure the accuracy of the algorithm:

$$\begin{cases} \mathrm{AVM} = \dfrac{A - B}{A}, \\ \mathrm{AUM} = \dfrac{B - A}{B}, \\ \mathrm{CM} = \dfrac{1}{3}\{\mathrm{AOM} + (1 - \mathrm{AVM}) + (1 - \mathrm{AUM})\}, \end{cases} \tag{10}$$

where A is the segmentation result marked by the doctor and B is the segmentation result of the algorithm, in which AOM and CM are proportional to the segmentation result and AVM and AUM are inversely proportional to the segmentation result. The performance is shown in Table 1. Crossentropy loss is used in multitarget segmentation and analyzed from the perspective of energy, but the signal-to-noise ratio of ultrasonic image is low and the energy is not concentrated, resulting in poor performance. Dice Loss is often used in medical small target segmentation loss function, but the image area occupied by heart data is large, which has the risk of overfitting, resulting in inaccurate boundary focusing. Tversky comprehensively considers the differences between simple samples and complex samples and constructs weights, which has better improved the case of fuzzy boundary and reduced the probability of pixel misclassification.

In order to verify the performance of different network algorithms, Sen and Spe are introduced to measure:

$$\begin{cases} \mathrm{Sen} = \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \\ \mathrm{Spe} = \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \end{cases} \tag{11}$$

where TP represents the number of pixels correctly predicted as positive samples and TN represents the number of pixels correctly predicted as negative samples. FP represents the number of pixels incorrectly predicted as positive samples, and FN represents the number of pixels incorrectly predicted as negative samples. The results are shown in Table 2, and the corresponding ROC curve is shown in Figure 7.

To visually display the segmentation effect, as shown in Figure 8, the pretreatment results are shown in green, and the obtained sector area is the detection area.

The heart is periodic diastolic and contractile, which can be determined according to the correlation of motion. It is on this basis that the significance region is determined to achieve auxiliary segmentation. Traditional U-Net can inevitably lose details after downsampling. Encoder-decoder structure can reduce the loss, but after the image restoration by upper sampling, it is difficult to pay attention to the details and determine the contour limits of the target category under the influence of multinoise. The proposed attention model pays more attention to the target region, highlights the difference

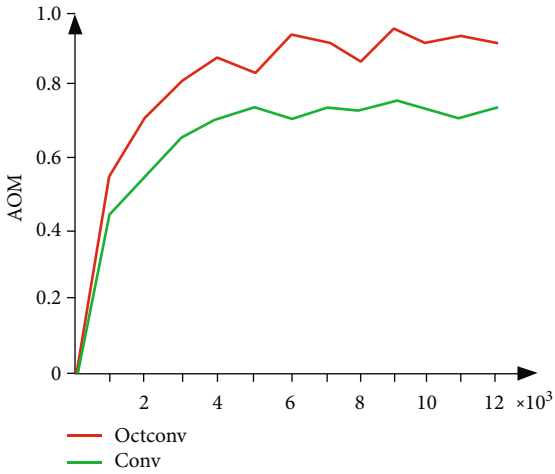


Figure 6: Algorithm convergence curve.

Table 1: Algorithm performance.

Algorithm	AOM	AVM	AUM	CM
Cross entropy	0.72	0.31	0.35	0.69
Dice	0.78	0.27	0.33	0.73
Tversky	0.92	0.18	0.21	0.84

Table 2: Algorithm performance.

Algorithm	Sen (%)	Spe (%)
U-Net	82	79
U-Net++	87	82
Ours	94	85

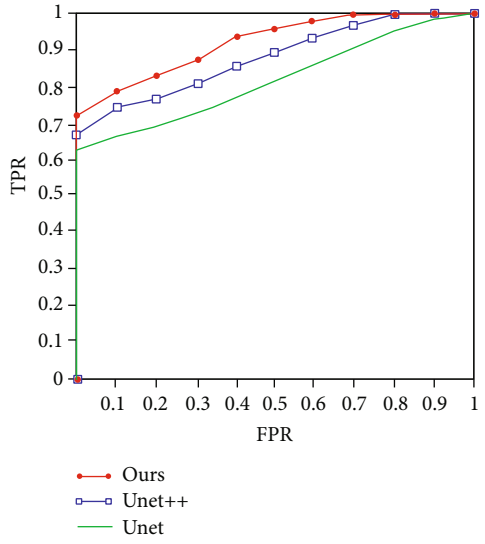


Figure 7: ROC curve.

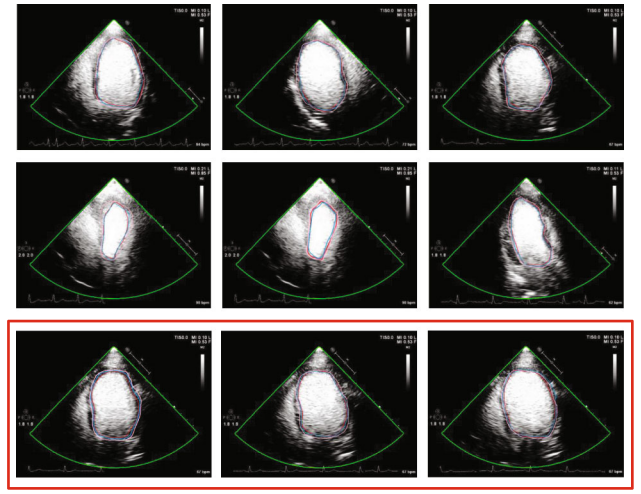


Figure 8: Segmentation effect.

between foreground and background, and realizes accurate segmentation.

In this paper, the deep learning network is adopted, and the results are related to the input image. Therefore, if the acquisition equipment is changed, the image quality will be different and the algorithm should be retrained, but the overall framework remains unchanged, so the algorithm should have certain effects.

4. Conclusion

The heart is the human body's important organ and has very important sense to the dynamic monitoring. Based on dynamic echocardiographic noninvasive, can display the superiority, is currently the main observation way of the heart. However, it is limited by the mechanism of ultrasonic imaging; imaging noise is large and easy to produce motion blur. We propose a new heart segmentation algorithm, which decomposes the image into low-frequency and high-frequency signals according to the principle of signal decomposition to obtain different frequency information. According to the principle of cardiac dynamic contraction, the attention model was introduced to simulate the diagnosis process of doctors and focus on the region of the heart. The network model will be built to realize multiscale information carrying and finally realize heart segmentation. Subsequent studies on myocardial extraction and analysis will be carried out to study the lightweight network, improve the calculation efficiency of the algorithm, and quickly assist doctors to make accurate diagnosis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by Key project of Shaanxi International Science and Technology Cooperation Foundation (2022kwz-19) and Xi'an Science and Technology Project (21YXYJ0105).

References

[1] J. T. Des Jardin, J. Chikwe, R. T. Hahn, J. W. Hung, and F. N. Delling, "Sex differences and similarities in valvular heart disease," *Circulation Research*, vol. 130, no. 4, pp. 455–473, 2022.

[2] H. Tang, T. Li, T. Qiu, and Y. Park, "Segmentation of heart sounds based on dynamic clustering," *Biomedical Signal Processing and Control*, vol. 7, no. 5, pp. 509–516, 2012.

[3] J. M. Icardo, "The teleost heart: a morphological approach," *Ontogeny and Phylogeny of the Vertebrate Heart*, pp. 35–53, 2012.

[4] S. J. Ahn, D. K. Kang, J. S. Sun, and M. H. Yoon, "Accuracy and predictive value of coronary computed tomography angiography for the detection of obstructive coronary heart disease in patients with an Agatston calcium score above 400," *Journal of Computer Assisted Tomography*, vol. 37, no. 3, pp. 387–394, 2013.

[5] T. Mythili, D. Mukherji, N. Padalia, and A. Naidu, "A heart disease prediction model using SVM-decision trees-logistic regression (SDL)," *International Journal of Computer Applications*, vol. 68, no. 16, 2013.

[6] J. Pedrosa, A. Castro, and T. T. Vinhoza, "Automatic heart sound segmentation and murmur detection in pediatric phonocardiograms," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2294–2297, Chicago, IL, USA, August 2014.

[7] A. Methaila, P. Kansal, H. Arya, and P. Kumar, "Early heart disease prediction using data mining techniques," *Computer Science & Information Technology Journal*, vol. 24, pp. 53–59, 2014.

[8] D. F. Pace, A. V. Dalca, T. Geva, A. J. Powell, M. H. Moghari, and P. Golland, "Interactive whole-heart segmentation in congenital heart disease," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., vol. 9351 of Lecture Notes in Computer Science, pp. 80–88, Springer, Cham, 2015.

[9] N. Saquib, M. T. I. Papon, I. Ahmad, and A. Rahman, "Measurement of heart rate using photoplethysmography," in *2015 International Conference on Networking Systems and Security (NSysS)*, pp. 1–6, Dhaka, Bangladesh, January 2015.

[10] G. Xiong, P. Sun, H. Zhou et al., "Comprehensive modeling and visualization of cardiac anatomy and physiology from CT imaging and computer simulations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 2, pp. 1014–1028, 2017.

[11] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease," in *Reconstruction, Segmentation, and Analysis of Medical Images*, pp. 95–102, Springer, Cham, 2016.

[12] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," *Computer Methods and Programs in Biomedicine*, vol. 141, pp. 19–26, 2017.

[13] Q. Tong, M. Ning, W. Si, X. Liao, and J. Qin, "3D deeply-supervised U-Net based whole heart segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 224–232, Springer, Cham, 2018.

[14] R. S. Ahmed, J. Liu, Z. Fei, and M. Zahid, "Automated segmentation of whole cardiac CT images based on deep learning," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 4, pp. 1–8, 2018.

[15] W. Gao and Y. Lu, "Fetal heart baseline extraction and classification based on deep learning," in *2019 International Conference on Information Technology and Computer Application (ITCA)*, pp. 211–216, Guangzhou, China, December 2019.

[16] X. Xu, T. Wang, Y. Shi et al., "Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 477–485, Cham, October 2019.

[17] V. H. C. de Albuquerque, D. D. A. Rodrigues, R. F. Ivo et al., "Fast fully automatic heart fat segmentation in computed tomography datasets," *Computerized Medical Imaging and Graphics*, vol. 80, p. 101674, 2020.

[18] I. Naseer, B. S. Khan, S. Saqib, S. N. Tahir, S. Tariq, and M. S. Akhter, "Diagnosis heart disease using Mamdani fuzzy inference expert system," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 26, 2020.

[19] A. Yoshida, Y. Lee, N. Yoshimura, T. Kuramoto, A. Hasegawa, and T. Kanazawa, "Automated heart segmentation using U-Net in pediatric cardiac CT," *Measurement: Sensors*, vol. 18, p. 100127, 2021.

[20] A. Banerjee, J. Camps, E. Zacur et al., "A completely automated pipeline for 3D reconstruction of human heart from 2D cine magnetic resonance slices," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2212, p. 20200257, 2021.

[21] J. O. B. Diniz, J. L. Ferreira, O. A. C. Cortes, A. C. Silva, and A. C. de Paiva, "An automatic approach for heart segmentation in CT scans through image processing techniques and Concat-U-Net," *Expert Systems with Applications*, vol. 196, p. 116632, 2022.

[22] J. O. B. Diniz, D. A. Dias Júnior, L. B. da Cruz et al., "Heart segmentation in planning CT using 2.5 D U-Net++ with attention gate," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–9, 2022.

[23] H. Liu, W. Chu, and H. Wang, "Automatic segmentation algorithm of ultrasound heart image based on convolutional neural network and image saliency," *IEEE Access*, vol. 8, pp. 104445–104457, 2020.

[24] H. Chen, H. Xu, P. Shi et al., "3-D Gabor-based anisotropic diffusion for speckle noise suppression in dynamic ultrasound images," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 1, pp. 207–219, 2021.

[25] Y. Song, S. Ren, Y. Lu, X. Fu, and K. K. Wong, "Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge," *Computer Methods and Programs in Biomedicine*, vol. 220, p. 106821, 2022.

[26] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[27] N. Oyama-Manabe, T. Aikawa, S. Tsuneta, and O. Manabe, "Clinical applications of 4D flow MR imaging in aortic valvular and congenital heart disease," *Magnetic Resonance in Medical Sciences*, vol. 21, no. 2, pp. 319–326, 2022.

[28] M. Lisi, M. Cameli, G. E. Mandoli et al., "Detection of myocardial fibrosis by speckle-tracking echocardiography: from prediction to clinical applications," *Heart Failure Reviews*, pp. 1–11, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 7552881, 15 pages
https://doi.org/10.1155/2022/7552881

Research Article

Accurate Adiabatic and Diabatic Potential Energy Surfaces for the Reaction of He + H_2

Jing Cao,[1] Nan Gao,[2] Yuxuan Bai,[3] Dequan Wang,[1] Ming Wang,[1] Shaokang Shi,[1] Xinyu Yang,[1] and Xuri Huang[1]

[1]*Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, College of Chemistry, Jilin University, Changchun, China*
[2]*Department of Thoracic Surgery, China-Japan Union Hospital of Jilin University, Changchun, China*
[3]*College of Life Science, Sichuan Agricultural University, Ya'an, China*

Correspondence should be addressed to Dequan Wang; dequan_wang@jlu.edu.cn and Xuri Huang; huangxr@jlu.edu.cn

Received 10 April 2022; Accepted 12 May 2022; Published 16 June 2022

Academic Editor: Chen Li

Copyright © 2022 Jing Cao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The accurate adiabatic and diabatic potential energy surfaces, which are for the two lowest states of He + H_2, are presented in this study. The Molpro 2012 software package is used, and the large basis sets (aug-cc-pV5Z) are selected. The high-level MCSCF/MRCI method is employed to calculate the adiabatic potential energy points of the title reaction system. The triatomic reaction system is described by Jacobi coordinates, and the adiabatic potential energy surfaces are fitted accurately using the B-spline method. The equilibrium structures and electronic energies for the H_2 are provided, and the corresponding different levels of vibrational energies of the ground state are deduced. To better express the diabatic process of the whole reaction, avoid crossing points being calculated and conical intersection also being optimized. Meanwhile, the diabatic potential energy surfaces of the reaction process are constructed. This study will be helpful for the analysis of histopathology and for the study in biological and medical mechanisms.

1. Introduction

Helium and hydrogen atoms were the most abundant substances in the universe [1]. The heating led to the rotation and vibration of the H_2 molecule in the universe when helium collided with H_2 in the molecular cloud. In the interstellar medium, the formation of stars was quite closely linked to their interactions, which suggested that the interactions between helium and hydrogen atoms were of particular interesting. Therefore, the reaction between He and H_2 was necessary to know the star formation in physics and chemistry, which was essential to understand the properties of many astrophysical objects. In addition, many researchers had constructed diabatic potential energy surfaces [2–16], so the construction of the global potential energy surfaces was an essential prerequisite for the title reaction. Which was an essential prerequisite for the study of molecular reaction kinetics.

Many years ago, the He + H_2 reaction system was studied by experimental [17–21] and theoretical [1, 22–35] researchers. In 1963, Roberts [36] obtained the interaction energy of the HeH_2 system at a relatively large center of charge separation. Subsequently, the interaction potential between He and H_2 was studied by Krauss and Mies [23], who got conclusion that the forces are comparable for collisions along the centerline and along the molecular parallels. Henry and co-works [37] constructed and calculated each atom in the HeH_2 system separately using Gaussian basis functions. The calculated 416 configurations were used to predict the van der Waals interaction between the lowest state of the helium atom and the hydrogen molecule. At the same time, Gengenbach and Hahn [20] constructed the earliest empirical potential energy surface for He-H_2. Wilson's work [26] proposed a universal expression for the HeH_2 potential energy surfaces with a straightforward function. Soon after that, Dove and Raynor [38] improved and

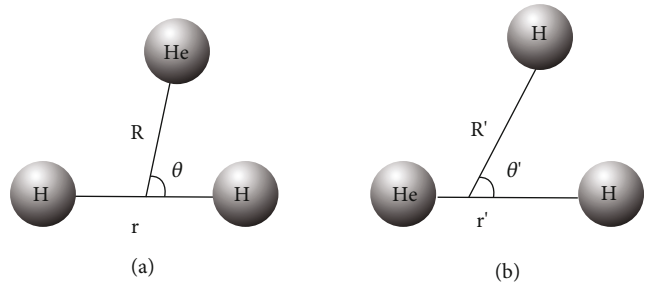


Figure 1: (a) Reactant Jacobi coordinate. (b) Product Jacobi coordinate.

Table 1: Calculation range of reactant part.

Calculation range of reactant part	
R (He-HH, Å)	0.4, 0.6, 0.8, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.2, 5.4, 5.6, 5.8, 6.0
r (HH, Å)	0.4, 0.6, 0.7, 0.72, 0.74, 0.75, 0.76, 0.8, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0
θ (degree)	0, 10, 20, 30, 40, 50, 60, 70, 80, 90

R (He-HH, Å): bond length from helium to the center of mass between two hydrogens, the unit is Å. r (HH, Å): bond length between two hydrogen atoms, the unit is Å. θ (degree): the angle between R and r, the unit is degree.

Table 2: Calculation range of products part.

Calculation range of products part	
R' (H-HeH, Å)	0.4, 0.6, 0.8, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.2, 5.4, 5.6, 5.8, 6.0
r' (HeH, Å)	0.4, 0.6, 0.7, 0.8, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0
θ' (degree)	0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180

R' (H-HeH, Å): bond length from the hydrogen atom to the center of mass between helium and hydrogen, the unit is Å. r' (HeH, Å): bond length between helium and hydrogen atoms, the unit is Å. θ' (degree): the angle between R' and r', the unit is degree.

optimized these potential energy surfaces. Meanwhile, they studied the collisional dissociation of He to H_2. The ground and excited state potential energy surfaces were constructed for the HeH_2 system by Stavros [36], who used the *ab initio* MRD CI calculation method. Up to 2003, Arnold and Peter [1] employed the *ab initio* method and calculated 20203 energy points in the interaction region of HeH_2, which had an improvement and enhancement compared with previous scholars. Cavity-enhanced spectra of He-H_2 were discussed by Słowiński et al. [39], and the results also showed that *ab initio* calculations could provide reference data for atmospheric spectroscopy studies. In the laboratory, the collision of hydrogen molecular and helium atoms would create ultracold molecules. Hence, the interaction between He and H_2 was also an essential factor in determining the properties of hydrogen clusters within helium.

In order to getting the precisely potential energy surfaces, the large range is used to scan. the large basis sets (aug-cc-pV5Z) were selected, and the three energy states of the He + H_2 system were calculated, using high-level MCSCF/MRCI. Secondly, the energy points were fitted using the B-spline method. Finally, as electrons would transfer near the interaction area, the conical intersection was optimized, and the avoided crossing points were studied. Meanwhile, the adiabatic and diabatic potential energy surfaces were constructed for this reaction system.

The study was organized as follows: in the next section, a brief overview of the theoretical calculation method; the adiabatic and diabatic potential energy surfaces of He + H_2 were presented in the third chapter. And the simple summary and the conclusion would be shown in the fourth chapter.

2. Computational Methods

2.1. Ab Initio Calculations. The MOLPRO 2012 [40] package was used to perform all *ab initio* single-point energies at the high-level method (MCSCF/MRCI) with large basis sets (aug-cc-pV5Z) [41]. The three lowest state adiabatic potential energies were calculated in the He + H_2 reaction system. There were 4 electrons in the orbital, including 15 active orbitals and 229 external orbitals ($141A' + 88A''$). The Jacobi (r, R, θ) coordinates were selected for this triatomic system, which was widely used in our former studies [42–49]. The configuration coordinates (Figure 1) were illustrated to describe the reaction of the triatomic molecules.

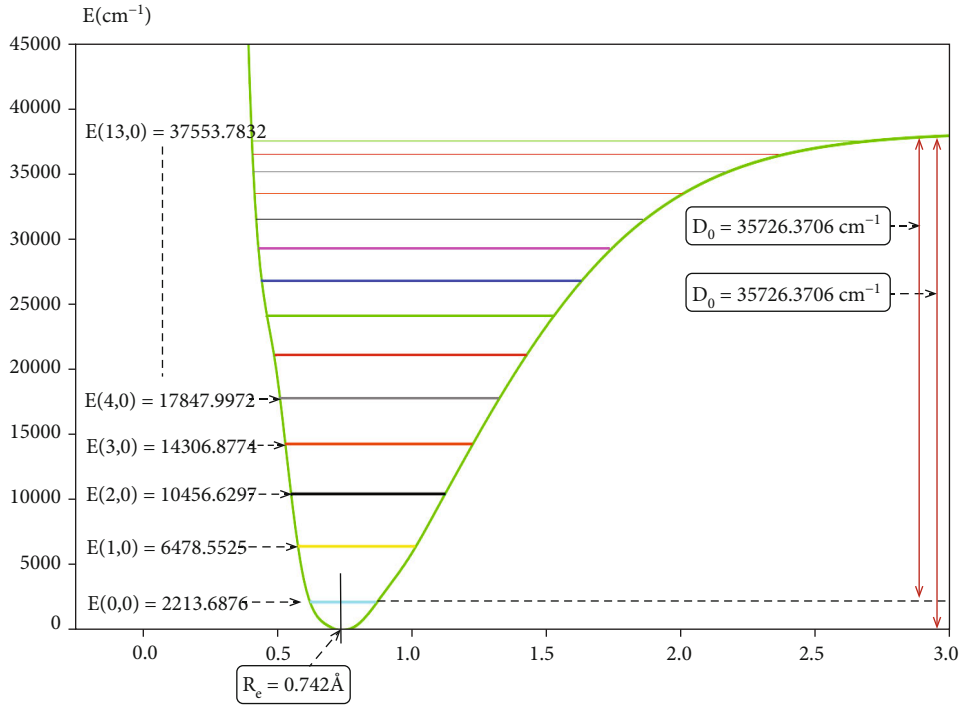


Figure 2: The one-dimensional potential energy surfaces for H_2 molecule as a function of distances $r_{H\text{-}H}$ (in Å) and its different vibrational state energies $E(v, j{=}0)$ (in cm^{-1}).

Table 3: Spectroscopic constants for H_2.

H_2 ($X^1\Sigma_g^+$)	R_e (Å)	D_0 (cm^{-1})	D_e (cm^{-1})
This work	0.742	35726.3706	37940.0582
Lee's work [49]	0.743	35687	37868
Yuan's work [50]	0.7414		38313.2
He's work [51]	0.7418		38186.2
Expt.	0.742 [52]	36118.06 [53]	38288 [52]

R_e (Å): the equilibrium bond length of the H_2, the unit is Å. D_0 (cm^{-1}): the experimental dissociation energy of the H_2, the unit is cm^{-1}. D_e (cm^{-1}): the equilibrium dissociation energy of the H_2, the unit is cm^{-1}.

The angle was determined from 0° to 90° with the grids at 10° for the reactants and field, and 12960 adiabatic potential energy points were calculated independently for each electronic state energy. For the products region, 21888 points were scanned at an angle of 0° to 180°. For each electronic state energy in the reactants and products, 34848 adiabatic potential energy points were measured, which guaranteed that the adiabatic potential energy surfaces were reliable. Tables 1 and 2 indicate the scan ranges of reactants and products, respectively. The B-spline method was used to fit the potential energy surface for this process. This method may be useful for the analysis of histopathology and for the study in biological and medical mechanisms.

2.2. Mixing Angle (α) Calculation. The wave functions of the three lowest states were ψ_1^a, ψ_2^a, and ψ_3^arespectively, and the adiabatic wave functions for the two lowest states were expressed as:

$$\begin{pmatrix} \Psi_1^d \\ \Psi_2^d \end{pmatrix} = \begin{pmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{pmatrix} \begin{pmatrix} \varphi_1^a \\ \varphi_2^a \end{pmatrix}, \tag{1}$$

where α was denoted as the mixing angles, and the energy of the diabatic potential energy (H_{ii}^d) was obtained by fitting the adiabatic potential energy (E_i^a):

$$H_{11}^d = \cos^2 a E_1^a + \sin^2 a E_2^a, \tag{2}$$

$$H_{22}^d = \sin^2 a E_1^a + \cos^2 a E_2^a, \tag{3}$$

$$H_{12}^d = H_{21}^d = \sin a \cos a (E_2^a - E_1^a), \tag{4}$$

$$H_{12}^d = H_{21}^d. \tag{5}$$

H_{11}^d and H_{21}^d were the coupling energy between the two states. The $\hat{P}_z$ was the Z component of the dipole moment operator. The following equation was obtained:

$$\langle \Psi_3^a | \hat{P}_z | \Psi_1^a \rangle = \sin a \langle \Psi_3^a | \hat{P}_z \left| \varphi_2^d \right\rangle + \cos a \langle \Psi_3^a | \hat{P}_z \left| \varphi_1^d \right\rangle, \tag{6}$$

$$\langle \Psi_3^a | \hat{P}_z | \Psi_2^a \rangle = -\sin a \langle \Psi_3^a | \hat{P}_z \left| \varphi_1^d \right\rangle + \cos a \langle \Psi_3^a | \hat{P}_z \left| \varphi_2^d \right\rangle. \tag{7}$$

$\langle \Psi_3^a | \hat{P}_z | \varphi_1^d \rangle$ and $\langle \Psi_3^a | \hat{P}_z | \varphi_2^d \rangle$ were both highly symmetric

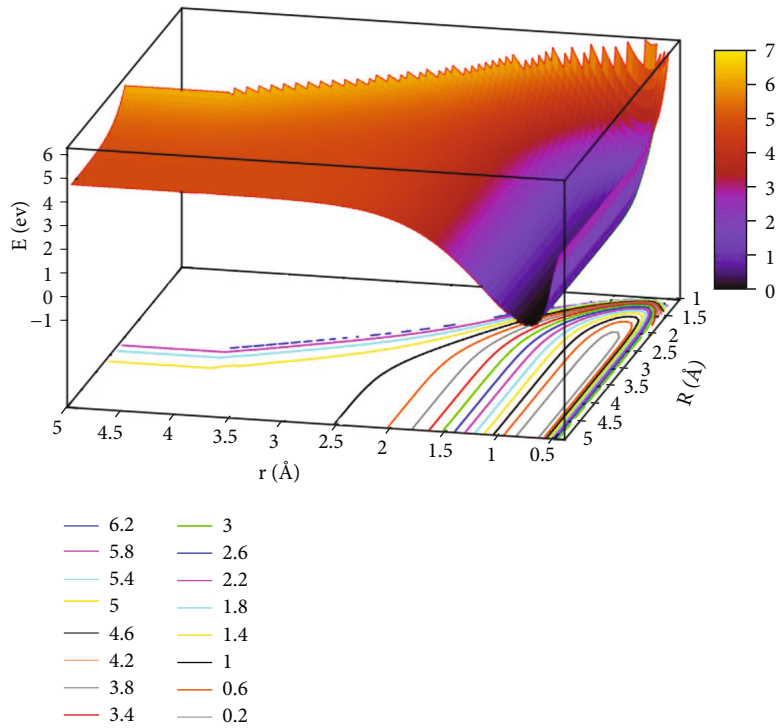


Figure 3: The lowest state potential energy surface (in eV) for the reaction of He + H_2 at θ=0.0° in Jacobi coordinate.

matrices:

$$\langle \Psi_3^a | \hat{P}_z | \varphi_1^a \rangle = 0, \tag{8}$$

$$\langle \Psi_3^a | \hat{P}_z | \varphi_2^a \rangle = 1. \tag{9}$$

Thus, the mixing angles were obtained as:

$$a = \arctan\left(\left|\frac{\langle \Psi_3^a | \hat{P}_z | \Psi_1^a \rangle}{\langle \Psi_3^a | \hat{P}_z | \Psi_2^a \rangle}\right|\right). \tag{10}$$

The main idea of the method for the mixing angle calculating can be used for histopathology after optimizing.

3. Results and Discussion

The lowest ground state and the first excited state potential energy surfaces were constructed in this reaction system with angles (including 0°, 30°, 50°, 60°, and 90°), sequentially. To more clearly and specifically display the potential energy surfaces, the lowest state energy of the He + H_2 was shifted at 0.00 eV, and the results were analyzed as follows.

3.1. One-Dimensional Diatomic Adiabatic Potential Surface. Figure 2 shows the equilibrium bond length (R_e) of the H_2 molecule was 0.742 Å, and the dissociation energy (D_e) was 37940.0582 cm^{-1}. From the image, it was clearly found that a total of 13 vibrational states in the one-dimensional potential energy surface, and the zero point energy of the diatomic molecule was $E(0,0)$ =2213.6876 cm^{-1}, and the highest vibrational energy was $E(13,0)$ =37553.7832 cm^{-1}.

The spectroscopic constants for H_2 are presented in Table 3. The data were found to be in good agreement with the theoretical and experimental researchers, such as Lee, Yuan, and He, which indicated that the calculations were more specific in this study.

3.2. Two-Dimensional Adiabatic Potential Energy Surfaces. The lowest potential energy for the He + H_2 reaction system at θ =0.0° is seen in Figure 3. All the potential energy surface curves were very smooth. The self-variable for x, y, and z were represented by r (in Å), R (in Å), and E (in eV). The contour intervals of the potential energy surface were 0.40 eV. When r >2.0 Å and R >3.0 Å, the distance between contours was relatively sparse, which implies that the energy was comparable. When 1.0 Å > r >0.5 Å and$R$ >2.5 Å was the entrance area in the reaction. The entrance part of the reaction had no reaction barrier, the potential energy surface had no minimum geometries, and there was no stable configuration. To more intuitively reflect the lowest ground state and the first excited state potential energy surfaces, we plotted the first excited state potential energy surfaces.

Figure 4 depicts the first excited state potential energy surface at θ =0.0°. When the He atom collision with H_2 along a straight line, the geometry crossing a transition state TS_1 (Panel B) would form the M_1 (Panel A). The M_1 configuration was r =1.15 Å and R =1.44 Å, and the energy was 11.870 eV. The energy of the TS_1 (Panel B) was about 12.037 eV, when the distance was R =1.85 Å, r =1.06. It indicated that when He attacked with H_2, it possibly generated a stable configuration (M_1). Similarly, the complex (M_1) overcame the TS_2 (Panel C) which would form the M_2, the complex (M_2) with r =2.66 Å and R =2.21 Å, and the energy value

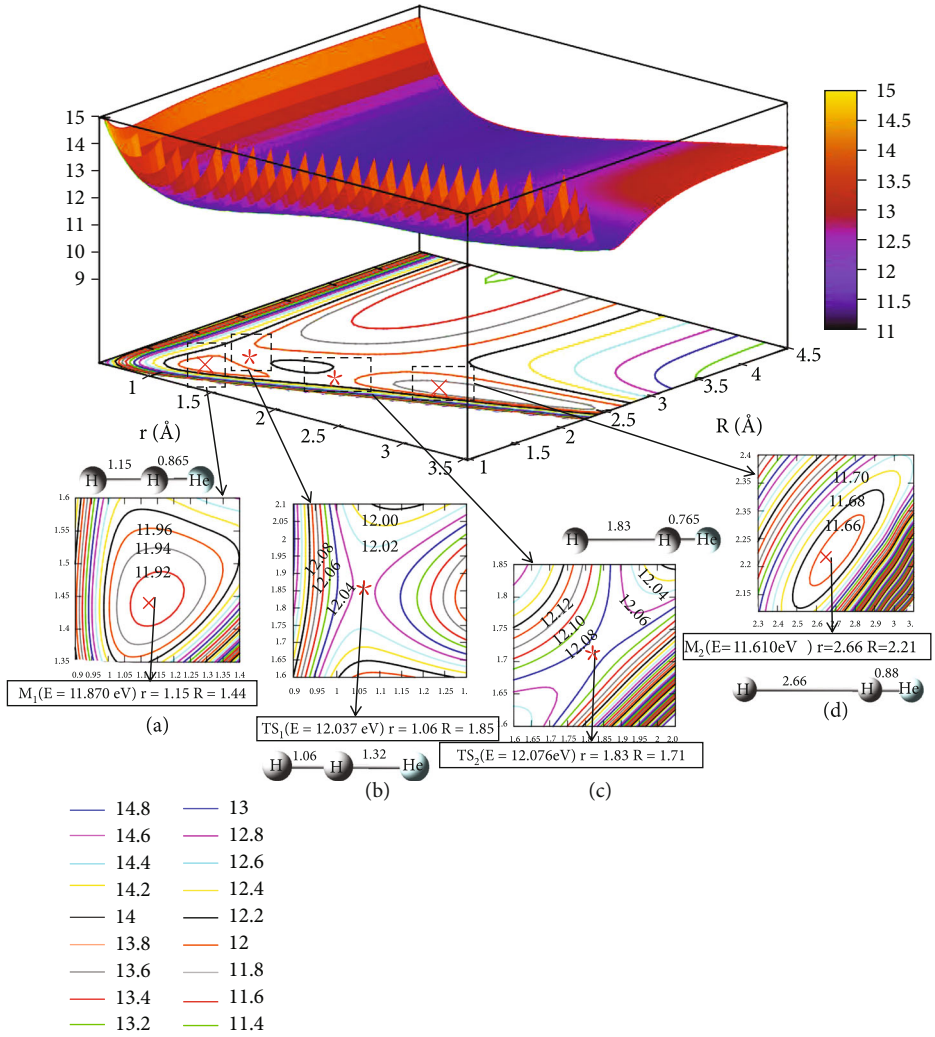


FIGURE 4: The first excited state potential energy surface (in eV) for the reaction of He + H_2 at $\theta = 0.0^{\circ}$ in Jacobi coordinate.

was 11.610 eV. To sum up, the energy of M_2 was 11.610 eV, which was slightly lower than the M_1.

The character of Figure 5 exhibited the lowest state for potential energy surfaces when the angle was 30.0°. From the figure, the energy between adjacent contours was 0.40 eV in the contours plot, and the potential energy values became smaller as the value of r and R decreased. When R >2.5 Å and 0.5 Å < r <1.0 Å, which was the entrance part of the reactants. There were no obvious minimum at the lowest potential energy surface and no stable configuration in this region. In order to express this process more clearly and distinctly, we also plotted the first excited state potential energy surface. As shown in Figure 6, when r was in the range of 1.0 Å to 1.5 Å and R >3.0 Å, which was the entrance region for the reactants. As R gradually became smaller and r became larger, the potential energy surface appeared at a saddle point (TS_3) at r =1.25 Å and R =1.73 Å, and the energy was 11.776 eV. Crossing the saddle point, a minimum (M_3) would form gradually. The configuration of M_3 was r equal to 2.65 Å and R equal to 0.88 Å, and the minimum energy was 10.147 eV.

When the angle was 50.0°, Figure 7 depicts the image of the potential energy surfaces of the ground state. In the region of 0.5 Å < r <1.0 Å and R >4.0 Å, which was the reaction entrance of the lowest state potential energy surface, and there was no obvious minimum and stable configuration. To better illustrate this process, we also showed the potential energy surfaces of excited state for the He + H_2 reaction system, which was described in Figure 8. It is clear that for 1.0 Å < r <1.5 Å and R >2.5 Å, the energy hardly changes with R, which indicated the entrance part of the reactant. As the helium atom gradually approached the center of mass of the two hydrogen atoms, the minimum (M_4) was formed, where at r =1.90 Å and R =0.84 Å, and the energy was 9.616 eV, which was the global minimum of the first excited state.

The lowest state potential energy surface and contour plot at θ =60.0° are shown in Figure 9. There was no minimum in the lowest state potential energy surface. In the range of 0.5 Å < r <1.0 Å and R >4.0 Å, the potential energy hardly changes with increasing R, so the region was the entrance region of the reactant. Similarly, we plotted the first excited state potential energy surface and contours plot in Figure 10.

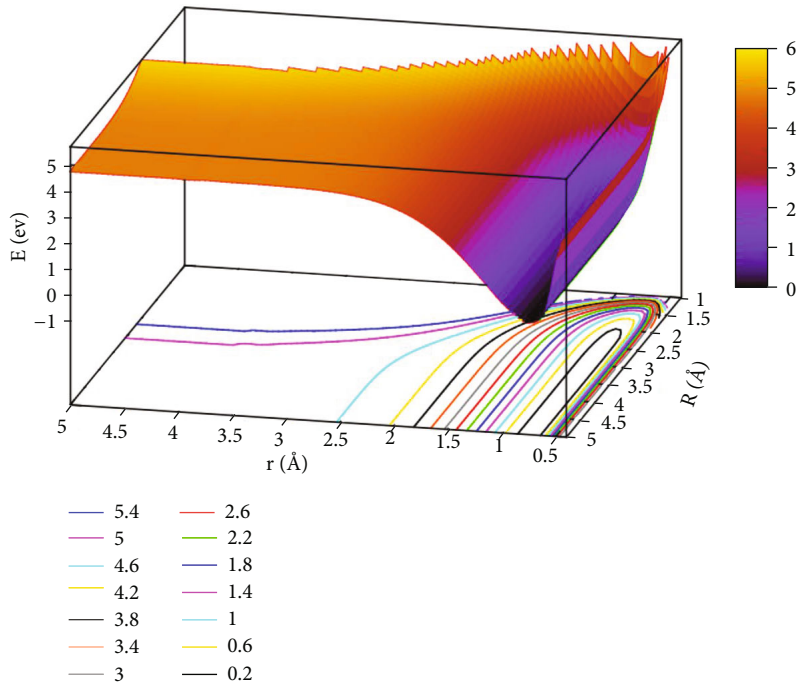


Figure 5: The lowest state potential energy surface (in eV) for the reaction of He + H_2 at θ=30.0° in Jacobi coordinate.

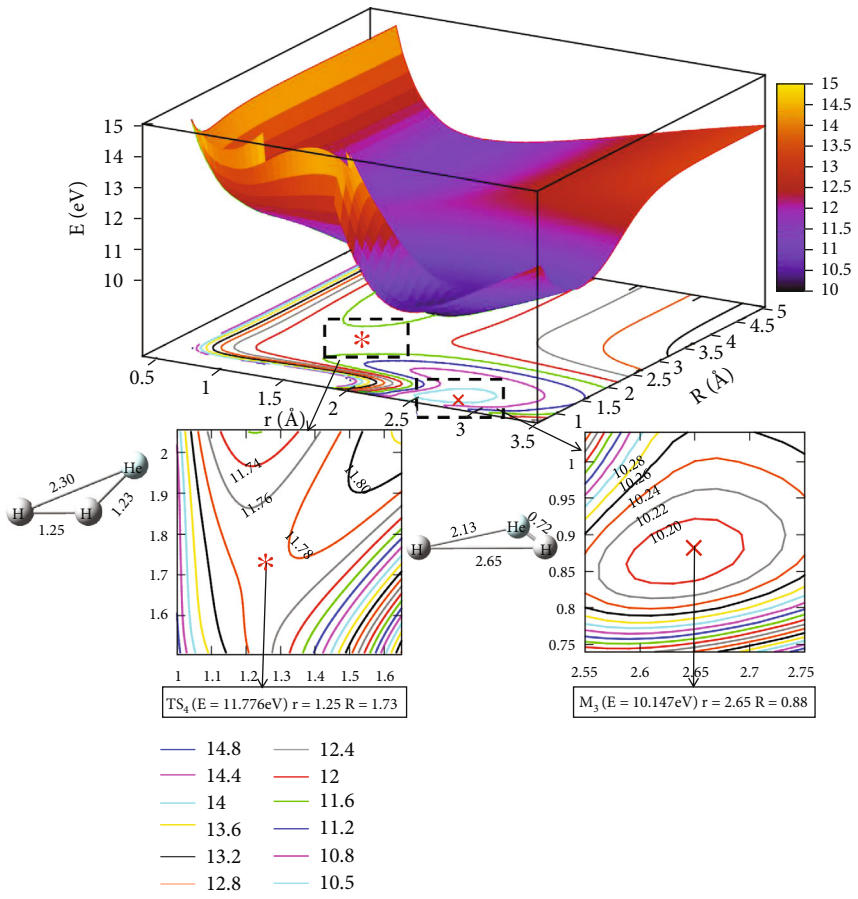


Figure 6: The first excited state potential energy surface (in eV) for the reaction of He + H_2 at θ=30.0° in Jacobi coordinate.

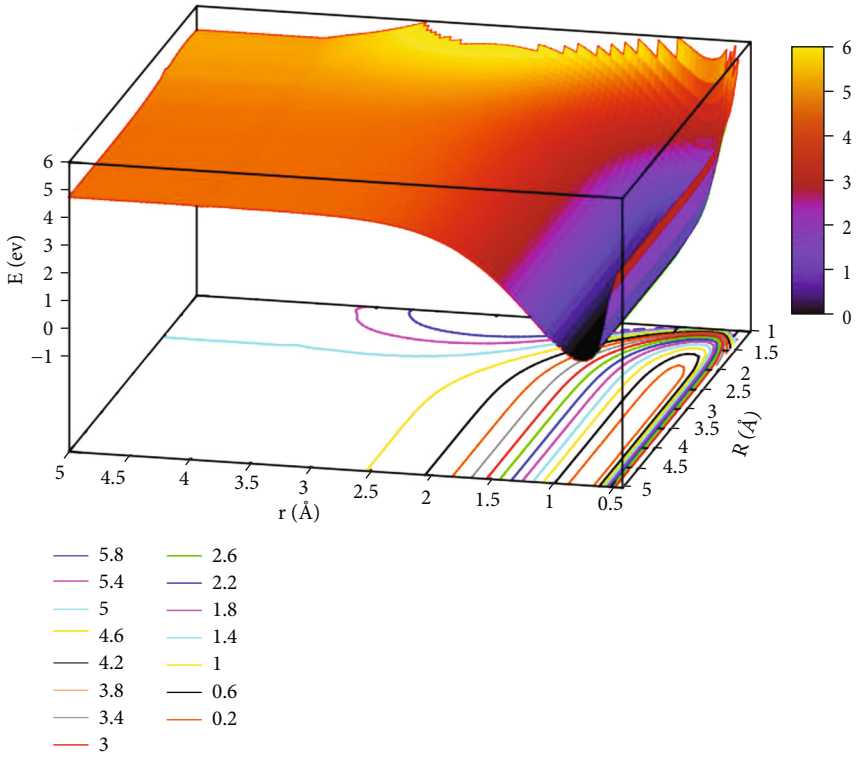


Figure 7: The lowest state potential energy surface (in eV) for the reaction of He + H_2 at θ=50.0° in Jacobi coordinate.

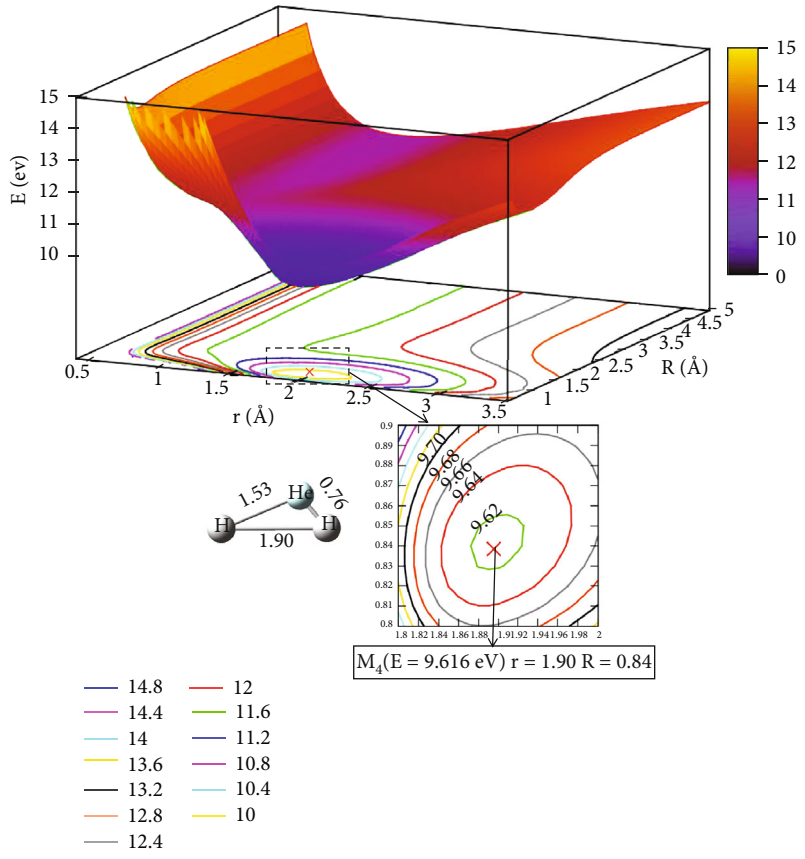


Figure 8: The first excited state potential energy surface (in eV) for the reaction of He + H_2 at θ = 50.0° in Jacobi coordinate.

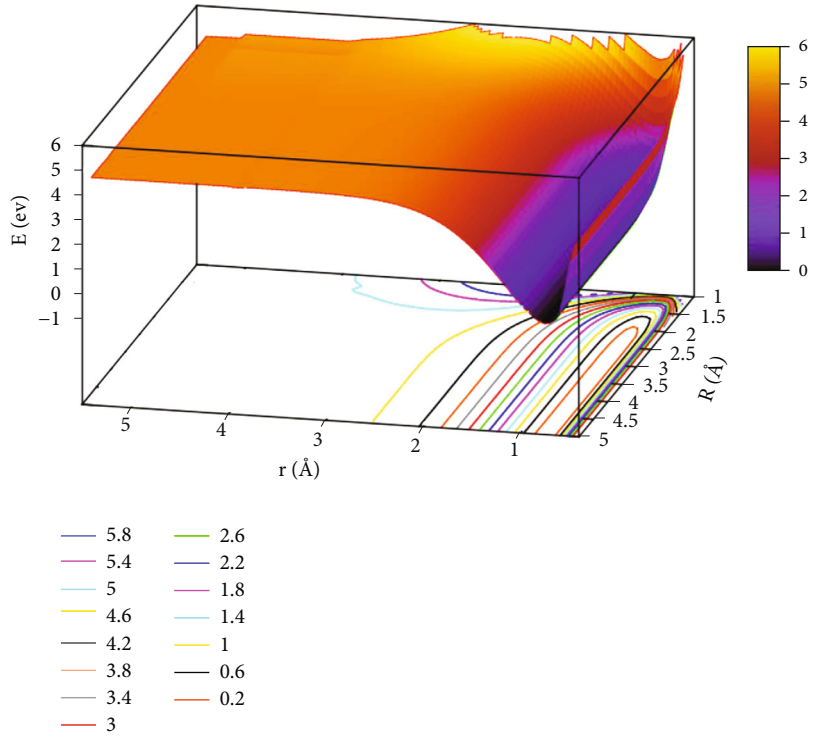


Figure 9: The lowest state potential energy surface (in eV) for He + H_2 at θ = 60.0° in Jacobi coordinate.

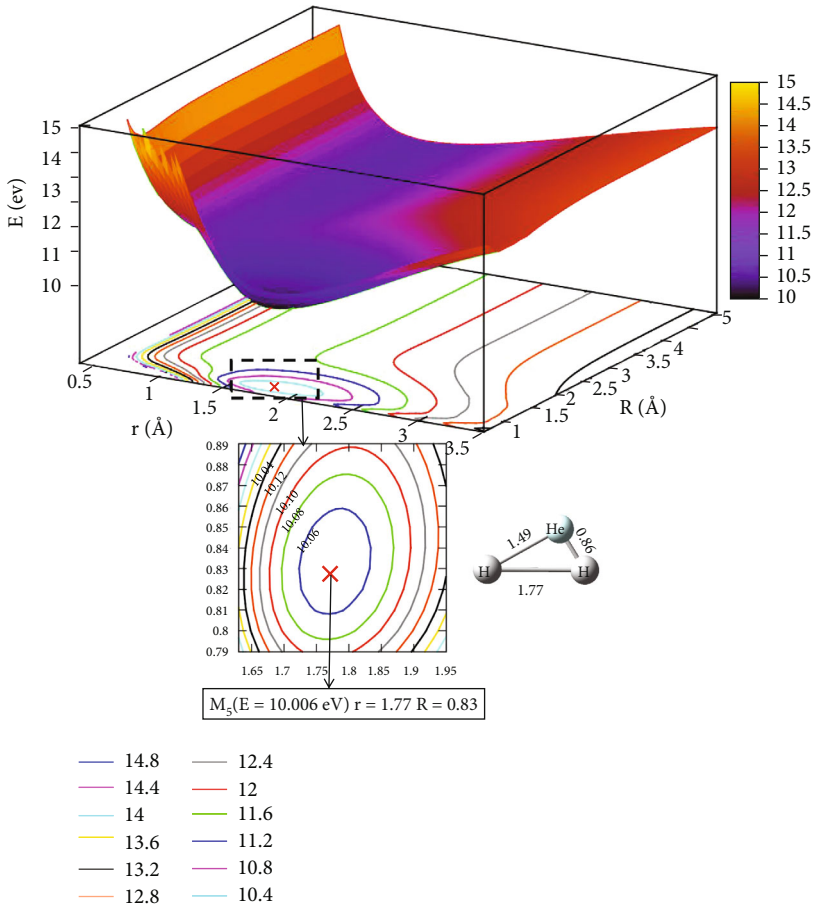


Figure 10: The first excited state potential energy surface (in eV) for the reaction of the He + H_2 at θ=60.0° in Jacobi coordinate.

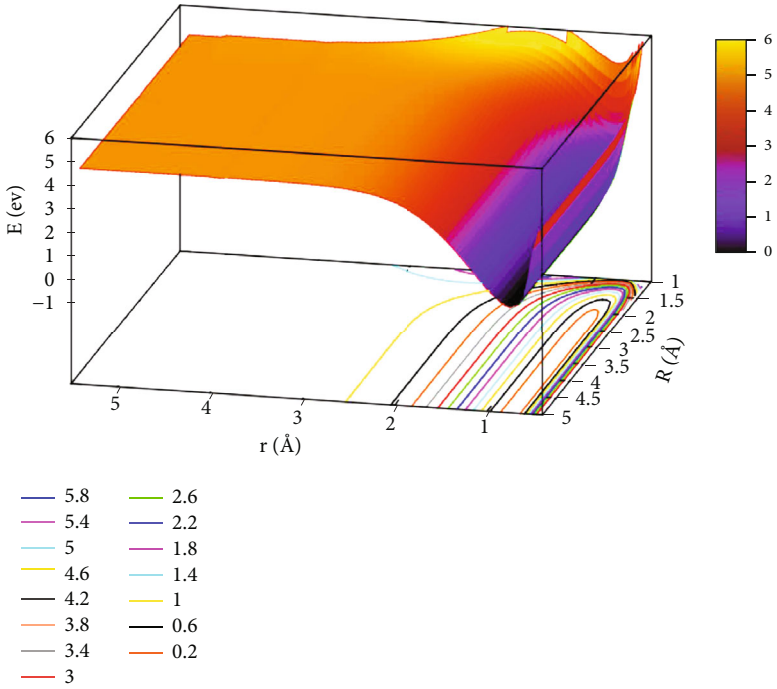


Figure 11: The lowest state potential energy surface (in eV) for He + H_2 at θ = 90.0° in Jacobi coordinate.

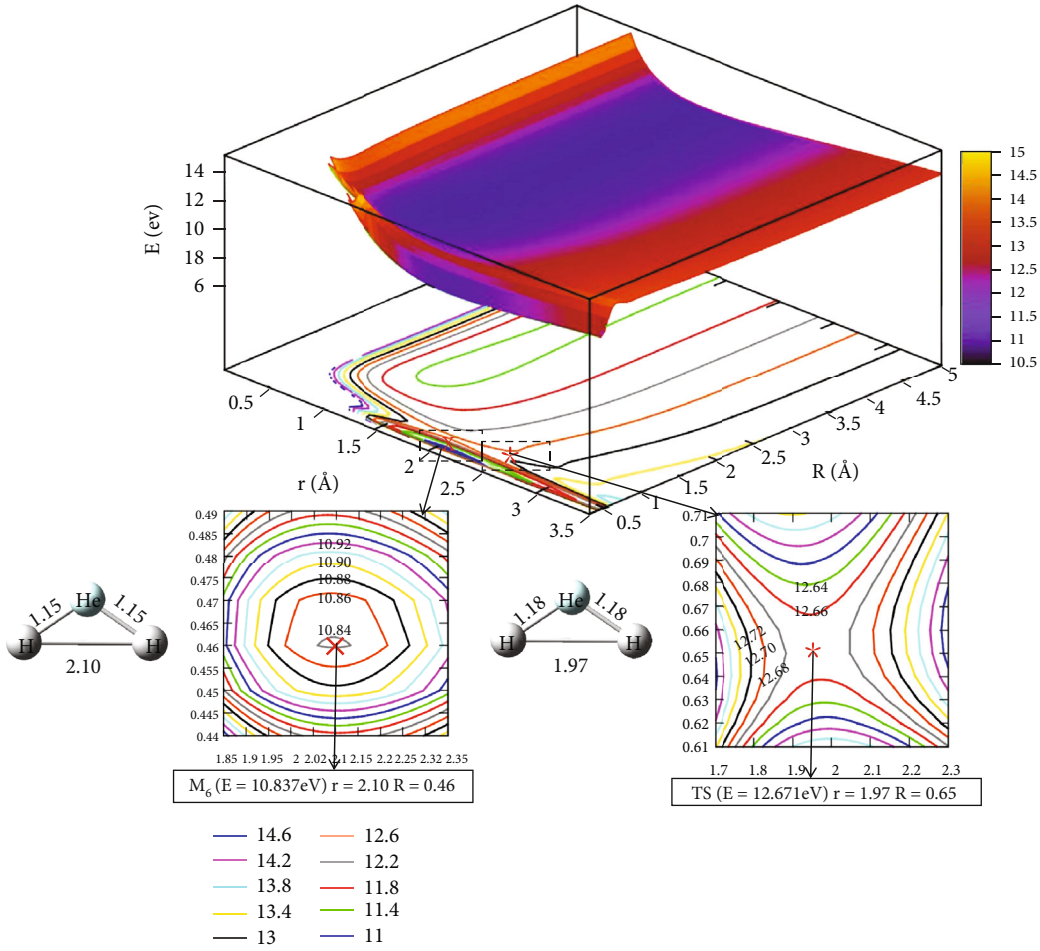


Figure 12: The first excited state potential energy surface (in eV) for the reaction of He + H_2 at θ=90.0° in Jacobi coordinate.

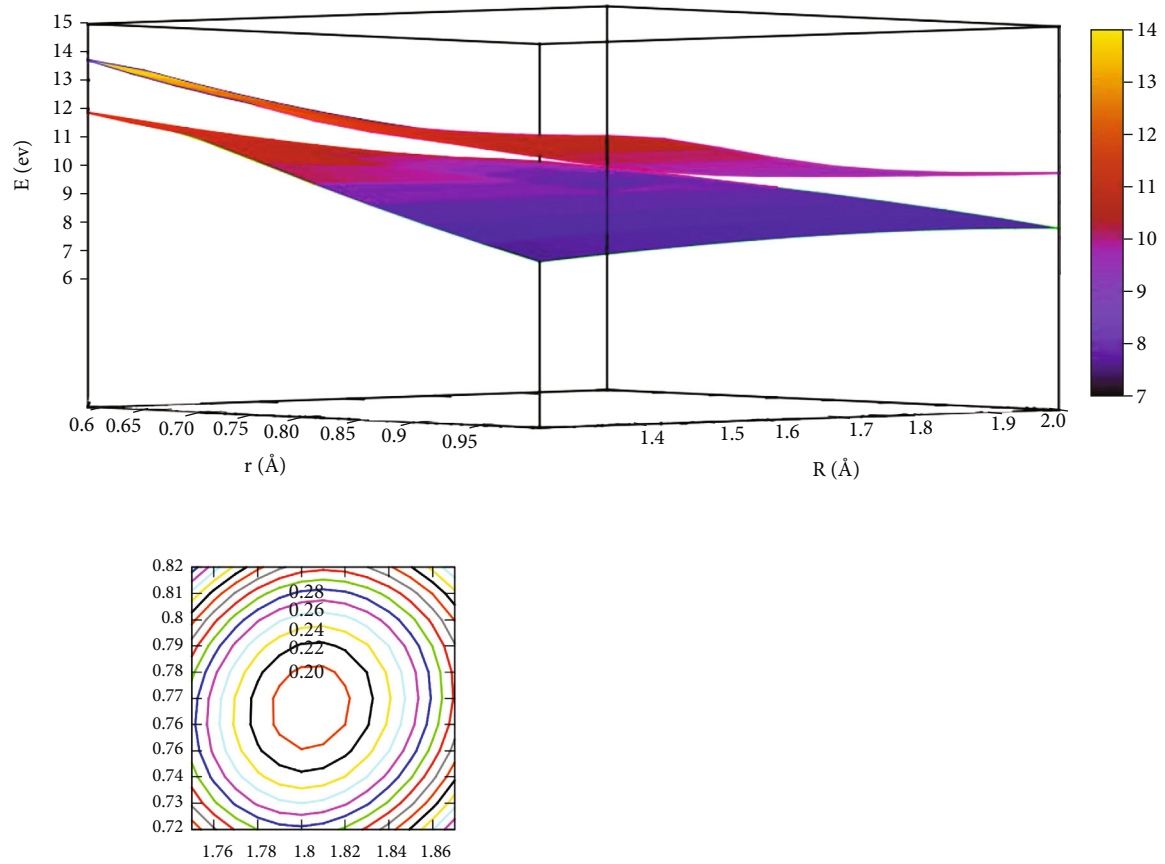


Figure 13: Avoid crossing point for the lowest state and the first excited state of He + H_2 reaction.

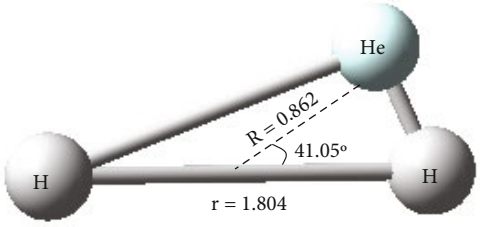


Figure 14: The conical intersection structure for He + H_2.

There was a complex (M_5) formed, which had an energy value of 10.006 eV when r =1.77 Å and R =0.83 Å. It was different from the energy value of θ =50°. It can be seen that M_5 was not the global minimum of the whole reaction system.

Figure 11 presents the lowest state potential energy surface of the He + H_2 reaction system when θ =90.0°. It was seen that for r >2.0 Å and R >2.5 Å, there was no minimum. At the 0.5 Å < r <1.0 Å and at R >4.5 Å, which was entrance region for the reaction. Figure 12 represents the first excited state potential energy. When the helium atom attacks the hydrogen atom along the angle of 90.0°, the entrance region was R >1.5 Å, 1.0 Å < r <1.5 Å. As the helium atoms close H_2, they passed a transition state TS_4, at R =0.65 Å and r =1.97 Å. Crossing this transition state, the minimum (M_6) was formed, and it was at r =2.10 Å, R =0.46 Å. The minimum energy was 10.837 eV.

3.3. Two-Dimensional Diabatic Potential Energy Surfaces

3.3.1. Avoid Crossing Point. As seen in Figure 13, it described the difference energy between the lowest state and the first excited state. At r =1.80 Å and R =0.77 Å, the energy interval between the two lowest states was the smallest, and the energy difference was 0.186 eV. It indicated that the diabatic reaction happens easily in this area.

3.3.2. Conical Intersection Structure. The configuration of the conical intersection point was optimized and obtained with MOLPRO 2012 software, and the distance between two hydrogen atoms was 1.804 Å. The distance from the helium atom to the center of mass between the two hydrogen molecules was 0.862 Å, as shown in Figure 14.

The mixing angle formula was used to calculate the diabatic potential energies of the He + H_2 reaction system. The kinetic characteristics of the entire reaction system were modified by the interaction zone. The adiabatic and diabatic potential energy surfaces with fixed θ =50.0° are shown in Figure 15. These panels (a), (b), (c), and (d) correspond to the potential energy curves, when the distance between two hydrogen atoms was equal to 1.1 Å, 1.2 Å, 1.3 Å, and 1.4 Å, respectively. When mix angle is equal to 0.0°, $E_1 = H_{22}$ and $E_2 = H_{11}$, and when the mixing angle is equal to 90.0°, adiabatic energy and diabatic energy were $E_1 = H_{11}$ and $E_2 = H_{22}$. The cross point appeared when the mixing angle was

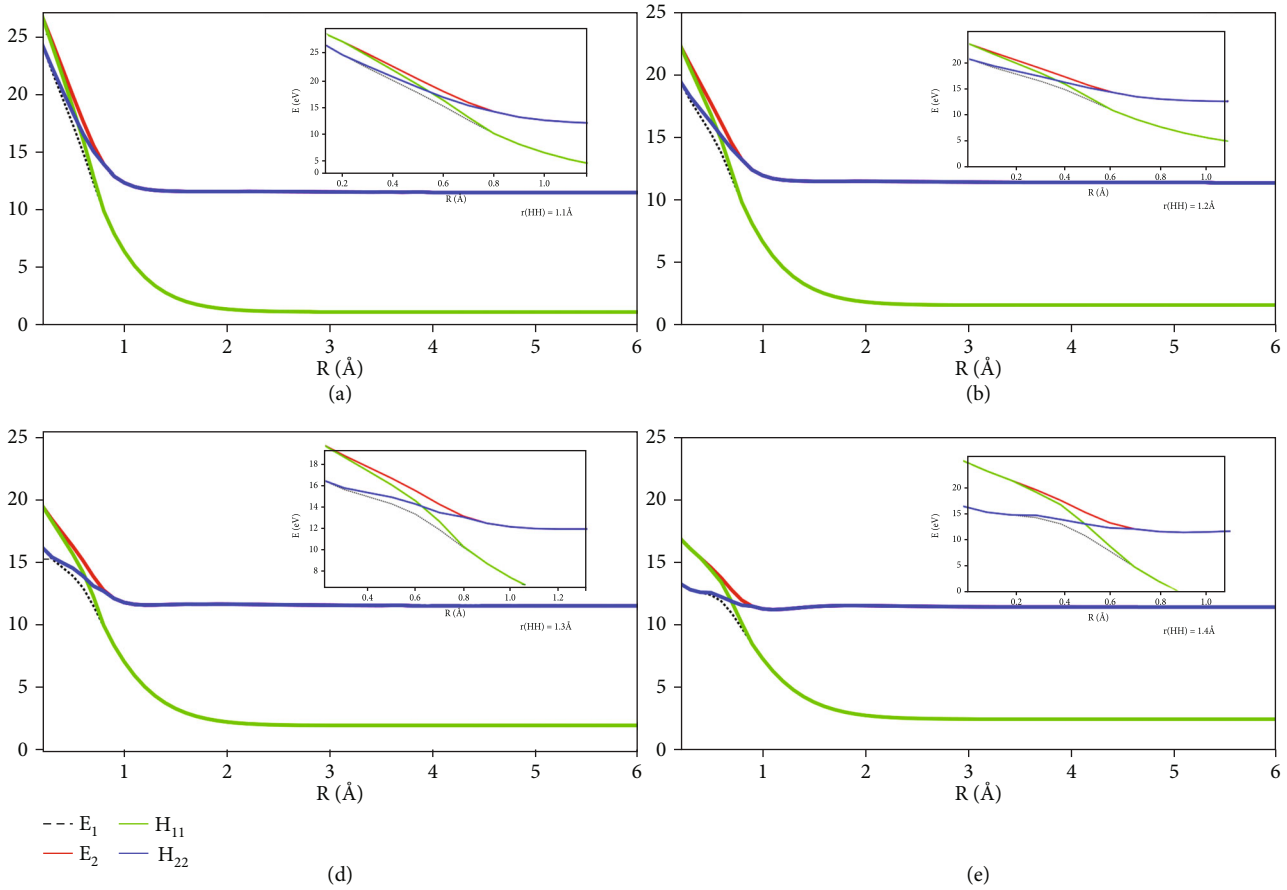


Figure 15: Adiabatic and diabatic potential energy surface at (a) $r_{H\text{-}H}$=1.1 Å, (b) $r_{H\text{-}H}$= 1.2 Å, (c) $r_{H\text{-}H}$=1.3 Å, (d) $r_{H\text{-}H}$=1.4 Å at θ= 50.0°.

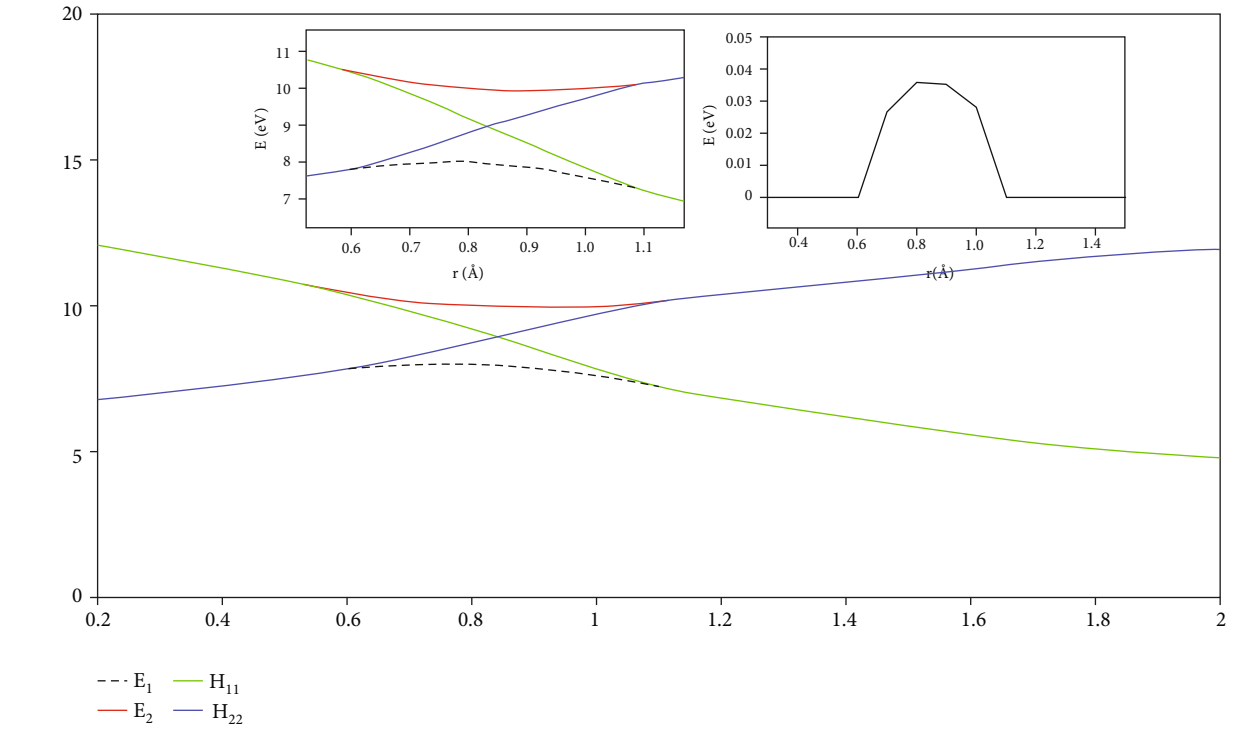


Figure 16: Cut-out plot of the diabatic potential energy surface (in eV) as a function of distance r (in Å)

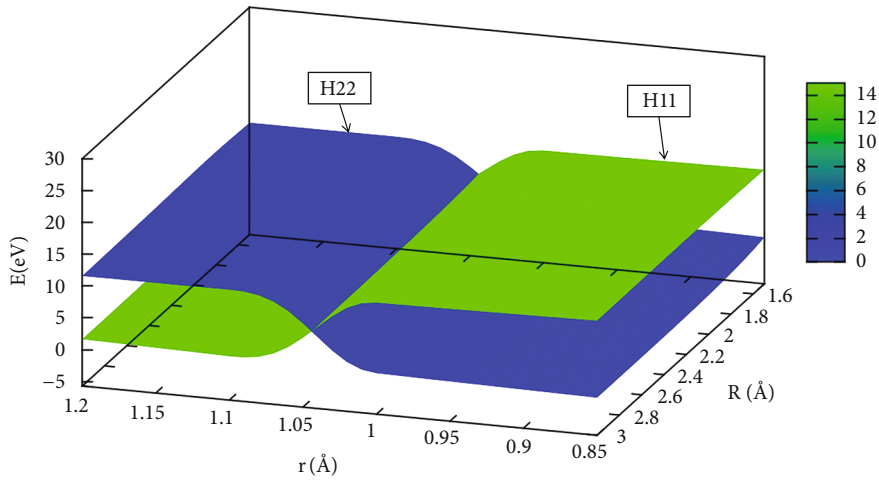


Figure 17: When θ=0.0°, diabatic potential energy surfaces (in eV) as a function of distances r (in Å) and R (in A).

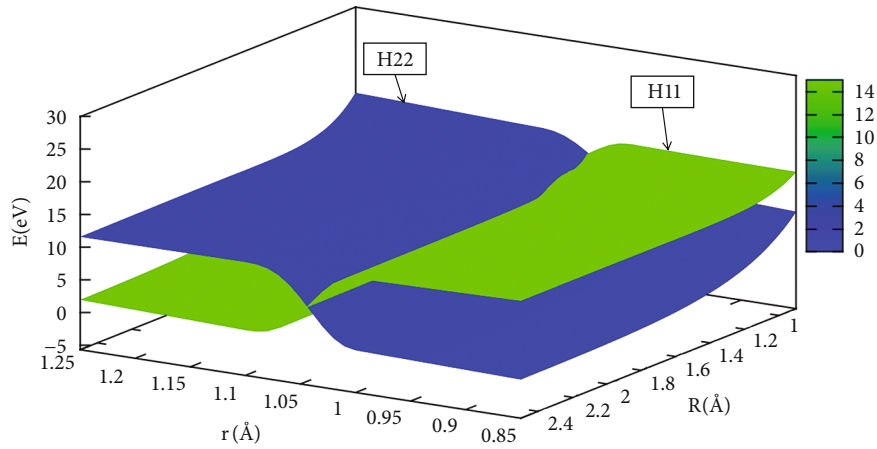


Figure 18: When θ=30.0°, Diabatic petential energy surfaces (in eV) as the function of r and R (in Å) for theta=30.0° r (in Å) and R (in A).

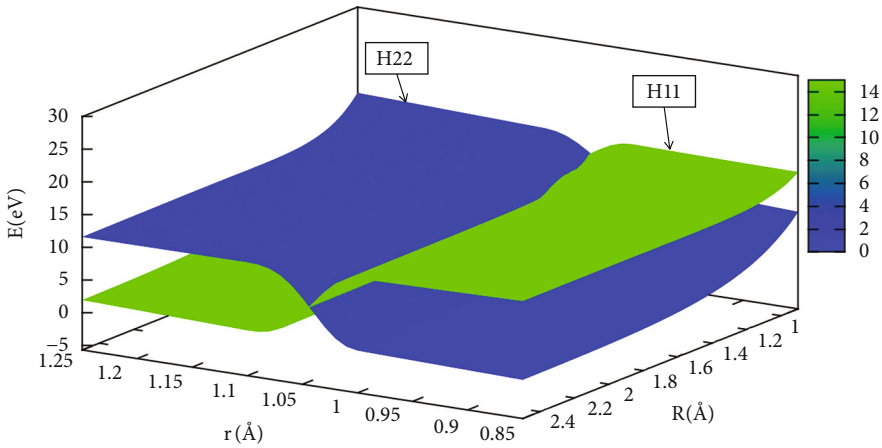


Figure 19: When θ=60.0°, diabatic potential energy surfaces (in eV) as a function of distances r (in Å) and R (in A).

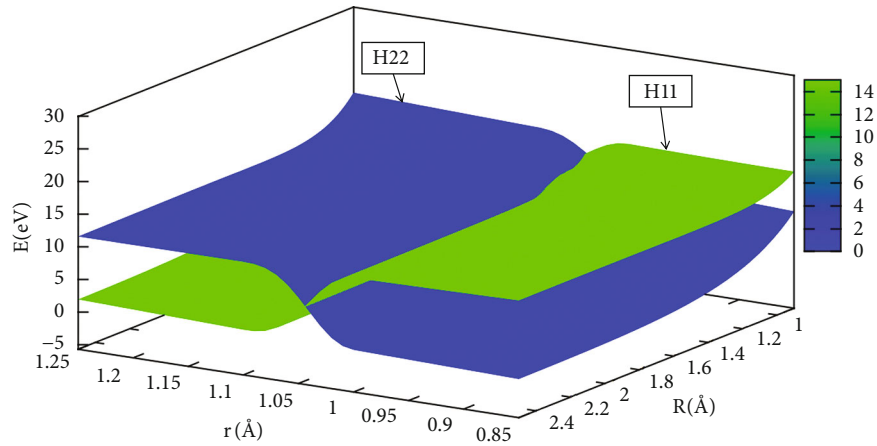


Figure 20: When θ=90.0°, diabatic potential energy surfaces (in eV) as a function of distances r (in Å) and R (in A).

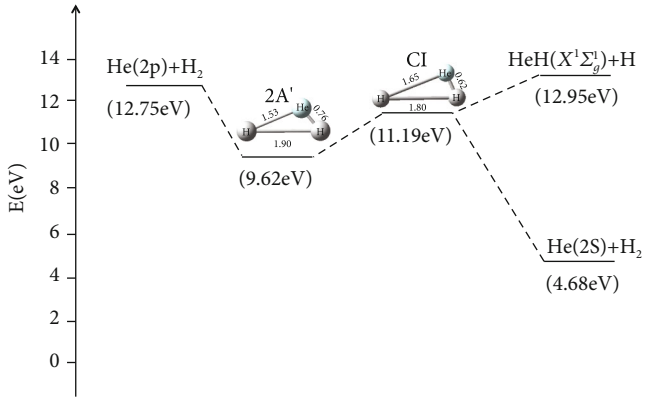


Figure 21: The possible reaction path in the He + H_2 reaction system.

equal to 45.0°. It was seen that the R of the crossing point increased with r.

While r was equal to 2.20 and the angle at 50.0°, Figure 16 is the enlarged view of the conical intersection area. The path of the transition from energy H_{11} to energy H_{22} in this range was clearly visible.

3.3.3. Two-Dimensional Diabatic Potential Energy Surfaces. In order to better understand the reaction process, the diabatic potential energy surfaces are expressed by Figures 17–20 for angles set as 0.0°,30.0°, 60.0°, 90.0°, respectively. The H_{11} and H_{22} represented the diabatic potential energy surfaces for the He + H_2 reaction system at angles (including 0.0°, 30.0°, 60.0°, and 90.0°), respectively.

The diabatic potential energy surfaces in the He + H_2 system were displayed at 0.0°, 30.0°, 60.0°, and 90.0°, where the blue part represented the H_{22} and the green part represented the H_{11}. As the distance between two hydrogens (r) atoms increased, the cross point appeared, and the H_{11} of the reaction was transformed into H_{22}. All the points on these crossing regions were almost located on the same straight line, which also meant that the diabatic potential energy surfaces constructed by the mixing angle formula were accurate and reliable.

The most likely reaction path for the He + H_2 reaction system was predicted, which is illustrated in Figure 21. Firstly, the first excited helium atom attacked the ground state H_2 molecule, which formed the intermediate with bond lengths of 1.53 Å, 1.90 Å, and 0.76 Å, respectively, and the energy was 9.62 eV. Secondly, the intermediate crossed the conical intersection, and the conical intersection bond lengths of 1.65 Å, 1.80 Å, and 0.62 Å. The configuration of conical intersection was very similar to the intermediate configuration in the first excited state for the reaction. Finally, the product is gradually produced. For this reaction, it was seen that the product of HeH $(X^1\Sigma_g^1)$+ H was the most probable reaction path to reach the product.

4. Conclusion

In this paper, the potential energy points of the He + H_2 reaction system were calculated using the MOLPRO 2012 software package, which selected the aug-cc-pV5Z basis sets, and applied the *ab initio* calculation method at the MCSCF and MRCI levels. The study calculated the region of 0°-90° for the reactants and 0°-180° for the products. A total of three lowest electronic states with 34848 potential energy points were calculated, and these energy points were fitted precisely using the B-Spline method.

First of all, based on the accurate potential energies, the diatomic potential energy curves of the H_2 molecule were constructed. We found a total of 13 vibrational states and studied the equilibrium bond distance of H_2 was 0.742 Å. The potential energy surface features were analyzed, which deduced that the global minimum energy of the first excited state was 9.616 eV. Meanwhile, the energy difference of two lowest states was 0.186 eV in interaction area, which was consistent with the conical intersection. Finally, the diabatic potential energy surfaces were constructed in the present work.

It was worth performing the full dynamical study with this global potential energy surface. We would continue this work in the following study [54].

Data Availability

The authors can supply the study data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jing Cao and Nan Gao contributed equally to this work.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2019YFC1804800). This work was completed under cooperation between the Institute of Theoretical Chemistry and High Performance Computing Center of Jilin University, China.

References

[1] A. I. Boothroyd, P. G. Martin, and M. R. Peterson, "Accurate analytic He-H_2 potential energy surface from a greatly expanded set of ab initio energies," *The Journal of Chemical Physics*, vol. 119, no. 6, pp. 3187–3207, 2003.

[2] U. Manthe, G. Capecchi, and H. J. Werner, "The effect of spin-orbit coupling on the thermal rate constant of the H_{2+} Cl → H + HCl reaction," *Physical Chemistry Chemical Physics*, vol. 6, no. 21, pp. 5026–5030, 2004.

[3] B. Jiang and D. Q. Xie, "New ab initio potential energy surfaces for Cl(2P3/2,2P1/2)+H2 reaction," *Chinese Journal of Chemical Physics*, vol. 22, no. 6, pp. 601–604, 2009.

[4] S. L. Mielke, D. G. Truhlar, and D. W. Schwenke, "Quantum photochemistry. The competition between electronically nonadiabatic reaction and electronic-to-vibrational, rotational, translational energy transfer in Br collisions with H," *The Journal of Physical Chemistry*, vol. 99, no. 44, pp. 16210–16216, 1995.

[5] A. J. Dobbyn and P. J. Knowles, "A comparative study of methods for describing non-adiabatic coupling: diabatic representation of the 1Sigma +/1Pi HOH and HHO conical intersections," *Molecular Physics*, vol. 91, no. 6, pp. 1107–1124, 1997.

[6] F. J. Aoiz, L. Bañares, and V. J. Herrero, "Dynamics of insertion reactions of H2 molecules with excited atoms," *The Journal of Physical Chemistry A*, vol. 110, no. 46, pp. 12546–12565, 2006.

[7] G. J. Halász, Á. Vibók, R. Baer, and M. Baer, "D matrix analysis of the Renner-Teller effect: an accurate three-state diabatization for NH2," *The Journal of Chemical Physics*, vol. 125, no. 9, article 094102, 2006.

[8] G. J. Halász, Á. Vibók, R. Baer, and M. Baer, "Renner-Teller nonadiabatic coupling terms: an ab-initio study of the HNH molecule," *The Journal of Chemical Physics*, vol. 124, no. 8, article 081106, 2006.

[9] G. L. Li, H. J. Werner, F. Lique, and M. H. Alexander, "New ab initio potential energy surfaces for the F+H_2 reaction," *The Journal of Chemical Physics*, vol. 127, no. 17, p. 174302, 2007.

[10] P. Defazio, B. Bussery-Honvault, P. Honvault, and C. Petrongolo, "Nonadiabatic quantum dynamics of C(1D)+H2→CH+H: coupled-channel calculations including Renner-Teller and Coriolis terms," *Journal of Chemical Physics*, vol. 135, no. 11, p. 114308, 2011.

[11] G. Hirsch, R. J. Buenker, and C. Petrongolo, "Ab initio study of NO2," *Molecular Physics*, vol. 70, no. 5, pp. 835–848, 1990.

[12] C. Petrongolo, G. Hirsch, and R. J. Buenker, "Diabatic representation of the Ã2A1[Btilde]2B2 conical intersection in NH2," *Molecular Physics*, vol. 70, no. 5, pp. 825–834, 1990.

[13] G. Hirsch, R. J. Buenker, and C. Petrongolo, "Ab initio study of NO2," *Molecular Physics*, vol. 73, no. 5, pp. 1085–1099, 1991.

[14] D. Simah, B. Hartke, and H. J. Werner, "Photodissociation dynamics of H_2S on new coupled ab initio potential energy surfaces," *The Journal of Chemical Physics*, vol. 111, no. 10, pp. 4523–4534, 1999.

[15] M. S. Gordon, V. A. Glezakou, and D. R. Yarkony, "Systematic location of intersecting seams of conical intersection in triatomic molecules: the $1\,2A'$–$2\,2A'$ conical intersections in BH2," *The Journal of Chemical Physics*, vol. 108, no. 14, pp. 5657–5659, 1998.

[16] F. Rebentrost and W. A. Lester, "Nonadiabatic effects in the collision of F(^{2}P) with $H_2(^1\Sigma_g^+)$. III. Scattering theory and coupled-channel computations," *The Journal of Chemical Physics*, vol. 67, no. 7, pp. 3367–3375, 1977.

[17] R. Gengenbach and C. Hahn, "Measurements of absolute integral total cross sections of He– H_2 and the interaction potential," *Chemical Physics Letters*, vol. 15, no. 4, pp. 604–608, 1972.

[18] R. Gengenbach, J. Strunck, and J. P. Toennies, "He–H2 potential parameters from molecular beam scattering experiments," *The Journal of Chemical Physics*, vol. 54, no. 4, pp. 1830–1832, 1971.

[19] P. Barletta, J. Tennyson, and P. F. Barker, "Creating ultracold molecules by collisions with ultracold rare-gas atoms in an optical trap," *Physical Review A*, vol. 78, no. 5, article 052707, 2008.

[20] S. Grebenev, B. Sartakov, J. P. Toennies, and A. F. Vilesov, "Evidence for superfluidity in para-hydrogen clusters inside helium-4 droplets at 0.15 kelvin," *Science*, vol. 289, no. 5484, pp. 1532–1535, 2000.

[21] F. Mezzacapo and M. Boninsegni, "Local superfluidity of para-hydrogen clusters," *Physical Review Letters*, vol. 100, no. 14, p. 145301, 2008.

[22] M. Karplus and H. J. Kolker, "Van der waals forces in atoms and molecules," *The Journal of Chemical Physics*, vol. 41, no. 12, pp. 3955–3961, 1964.

[23] M. Krauss and F. H. Mies, "Interaction potential between He and H2," *The Journal of Chemical Physics*, vol. 42, no. 8, pp. 2703–2708, 1965.

[24] M. D. Gordon and D. Secrest, "Helium-atom-hydrogen-molecule potential surface employing the LCAO-MO-SCF and CI methods," *The Journal of Chemical Physics*, vol. 52, no. 1, pp. 120–131, 1970.

[25] B. Tsapline and W. Kutzelnigg, "Interaction potential for He/H_2 including the region of the van der waals minimum," *Chemical Physics Letters*, vol. 23, no. 2, pp. 173–177, 1973.

[26] C. W. Wilson, R. Kapral, and G. Burns, "Potential energy surface for the hydrogen molecule-helium system," *Chemical Physics Letters*, vol. 24, no. 4, pp. 488–491, 1974.

[27] P. J. M. Geurts, P. E. S. Wormer, and A. V. D. Avoird, "Interaction potential for He-H_2 in the region of the van der Waals minimum," *Chemical Physice Letters*, vol. 35, no. 4, pp. 444–449, 1975.

[28] A. W. Raczkowski and W. A. Lester, "Extension of a He-H_2 potential energy surface," *Chemical Physics Letters*, vol. 47, no. 1, pp. 45–49, 1977.

[29] K. T. Tang and J. P. Toennies, "A simple theoretical model for the van der waals potential at intermediate distances. II. Anisotropic potentials of H_e–H_2 and Ne–H2," *The Journal of Chemical Physics*, vol. 68, no. 12, pp. 5501–5517, 1978.

[30] W. Meyer, P. C. Hariharan, and W. Kutzelnigg, "Ab initio potential surface for H_2-He," *The Journal of Chemical Physics*, vol. 73, p. 1880, 1980.

[31] A. Russek and R. Garcia, "Analytic fit of the He-H2 energy surface in the repulsive region," *Physical Review A*, vol. 26, no. 4, pp. 1924–1930, 1982.

[32] U. E. Senff and P. G. Burton, "An ab-initio study of the isotropic and anisotropic potential energy surfaces of the helium-molecular hydrogen interaction," *The Journal of Physical Chemistry*, vol. 89, no. 5, pp. 797–806, 1985.

[33] F. M. Tao, "An accurate ab initio potential energy surface of the He–H2 interaction," *The Journal of Chemical Physics*, vol. 100, no. 7, pp. 4947–4954, 1994.

[34] P. Muchnick and A. Russek, "The H_eH_2 energy surface," *The Journal of Chemical Physics*, vol. 100, no. 6, pp. 4336–4346, 1994.

[35] C. S. Roberts, "Inelastic scattering from a diatomic molecule: rotational excitation upon collision between He and H2 and H2 and H2," *Physical Review*, vol. 131, no. 1, pp. 209–217, 1963.

[36] S. C. Farantos, J. N. Murrell, and S. Carter, "Analytical ab initio potential-energy surfaces for the ground and the first singlet excited states of HeH_2," *Chemical Physics Letters*, vol. 108, no. 4, pp. 367–372, 1984.

[37] H. F. Schaefer, D. Wallach, and C. F. Bender, "Interaction potential between ground state helium atom and the B1Σu+ state of the hydrogen molecule," *The Journal of Chemical Physics*, vol. 56, no. 3, pp. 1219–1223, 1972.

[38] J. E. Dove and S. Raynor, "A quasiclassical trajectory study of the collisional dissociation of H_2 by He," *Chemical Physics*, vol. 28, no. 1-2, pp. 113–124, 1978.

[39] M. Słowiński, F. Thibault, Y. Tan et al., "H_2-He collisions: ab initio theory meets cavity-enhanced spectra," *Physical Review A*, vol. 101, no. 5, article 052705, 2020.

[40] H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: a general-purpose quantum chemistry program package," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 2, pp. 242–253, 2012.

[41] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning Jr., and A. K. Wilson, "Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg," *Theoretical Chemistry Accounts*, vol. 128, no. 1, pp. 69–82, 2011.

[42] D. Q. Wang, L. W. Fu, Z. X. Qu, Y. K. Chen, and X. R. Huang, "Accurate potential surfaces for the ground state of H+C_2 reaction," *The European Physical Journal D*, vol. 71, no. 10, pp. 1–7, 2017.

[43] L. W. Fu, D. Q. Wang, and X. R. Huang, "Accurate potential energy surfaces for the first two lowest electronic states of the Li (2p) + H2 reaction," *RSC Advances*, vol. 8, no. 28, pp. 15595–15602, 2018.

[44] D. Q. Wang, D. G. Wang, L. W. Fu, and X. Huang, "Accurate potential surfaces for the first three lowest states of reaction $O(^3P)$ + C_2(a3Πu) -> CO(X1Σ) + $C(^1D)$," *Chemical Physics*, vol. 517, pp. 228–236, 2019.

[45] D. Q. Wang, D. G. Wang, L. W. Fu et al., "An accurate ground state potential surface for the scattering reaction F−+ F2(v,j) → F2(v',j') + F−," *RSC Advances*, vol. 9, no. 4, pp. 1929–1932, 2019.

[46] D. Q. Wang, G. Shi, L. W. Fu, R. Yin, and Y. Ji, "Accurate potential energy surfaces for the three lowest electronic states of N(^{2}D) + H2(X1∑g+) scattering reaction," *ACS Omega*, vol. 4, no. 7, pp. 12167–12174, 2019.

[47] R. L. Yin, N. Gao, J. Cao, Y. Li, D. Wang, and X. Huang, "Global accurate diabatic potential surfaces for the reaction H + Li2," *RSC Advances*, vol. 10, no. 64, pp. 39226–39240, 2020.

[48] R. L. Yin, N. Gao, R. M. Zhang, D. Wang, and X. Huang, "Accurate potential energy surfaces for the excited state of CF_2 molecule," *Chemical Physics*, vol. 538, p. 110906, 2020.

[49] H. S. Lee, Y. S. Lee, and G. H. Jeung, "Potential energy surfaces for LiH2 and photochemical reactions Li∗+ H2↔ LiH + H," *The Journal of Physical Chemistry A*, vol. 103, no. 50, pp. 11080–11088, 1999.

[50] J. C. Yuan, D. He, and M. D. Chen, "A new potential energy surface for the ground electronic state of the LiH2 system, and dynamics studies on the H(2S) + LiH(X1Σ+) → Li(2S) + H2(X1Σg+) reaction," *Physical Chemistry Chemical Physics*, vol. 17, no. 17, pp. 11732–11739, 2015.

[51] D. He, J. C. Yuan, H. X. Li, and M. Chen, "Global diabatic potential energy surfaces and quantum dynamical studies for the Li(2p) + $H_2(X^1\Sigma^+_g)$ -> LiH($X^1\Sigma^+$) + H reaction," *Scientific Reports*, vol. 6, no. 1, pp. 1–10, 2016.

[52] W. C. Stwalley and W. T. Zemke, "Spectroscopy and structure of the Lithium Hydride diatomic molecules and ions," *Journal of Physical and Chemical Reference Data*, vol. 22, no. 1, pp. 87–112, 1993.

[53] E. E. Eyler and N. Melikechi, "Near-threshold continuum structure and the dissociation energies of H2, HD, and D2," *Physical Review A*, vol. 48, no. 1, pp. R18–R21, 1993.

[54] M. M. Rahaman, L. Chen, Y. D. Yao et al., "Identification of COVID-19 samples from chest X-Ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

Hindawi
BioMed Research International
Volume 2022, Article ID 5164970, 19 pages
https://doi.org/10.1155/2022/5164970

Review Article

Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review

Shah Hussain,[1] Iqra Mubeen,[1] Niamat Ullah,[2] Syed Shahab Ud Din Shah,[1] Bakhtawar Abduljalil Khan,[3] Muhammad Zahoor,[4] Riaz Ullah,[5] Farhat Ali Khan,[6] and Mujeeb A. Sultan[7]

[1]*Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University Islamabad, Pakistan*
[2]*Department of Pharmacy, Faculty of Biological Sciences, University of Malakand, Pakistan*
[3]*Women's Wellness and Research Center, Hamad Medical Corporation, Doha, Qatar*
[4]*Department of Biochemistry, University of Malakand, Chakdara, Dir Lower, KPK, Pakistan*
[5]*Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia*
[6]*Department of Pharmacy, Shaheed Benazir Bhutto University, Sheringal, Dir Upper, KPK, Pakistan*
[7]*Department of Pharmacy, Faculty of Medical Sciences, Aljanad University for Science and Technology, Taiz, Yemen*

Correspondence should be addressed to Mujeeb A. Sultan; mujeeb.aa@just.edu.ye

Received 19 April 2022; Accepted 25 May 2022; Published 6 June 2022

Academic Editor: Chen Li

Copyright © 2022 Shah Hussain et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medical imaging is the process of visual representation of different tissues and organs of the human body to monitor the normal and abnormal anatomy and physiology of the body. There are many medical imaging techniques used for this purpose such as X-ray, computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), digital mammography, and diagnostic sonography. These advanced medical imaging techniques have many applications in the diagnosis of myocardial diseases, cancer of different tissues, neurological disorders, congenital heart disease, abdominal illnesses, complex bone fractures, and other serious medical conditions. There are benefits as well as some risks to every imaging technique. There are some steps for minimizing the radiation exposure risks from imaging techniques. Advance medical imaging modalities such as PET/CT hybrid, three-dimensional ultrasound computed tomography (3D USCT), and simultaneous PET/MRI give high resolution, better reliability, and safety to diagnose, treat, and manage complex patient abnormalities. These techniques ensure the production of new accurate imaging tools with improving resolution, sensitivity, and specificity. In the future, with mounting innovations and advancements in technology systems, the medical diagnostic field will become a field of regular measurement of various complex diseases and will provide healthcare solutions.

1. Introduction

Medical imaging is the process of visual representation of the structure and function of different tissues and organs of the human body for clinical purposes and medical science for detailed study of normal and abnormal anatomy and physiology of the body. Medical imaging techniques are used to show internal structures under the skin and bones, as well as to diagnose abnormalities and treat diseases [1]. Medical imaging has changed into healthcare science. It is an important part of biological imaging and includes radiology which uses the imaging technologies like X-ray radiography, X-ray computed tomography (CT), endoscopy, magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS), positron emission tomography (PET), thermography, medical photography, electrical source imaging (ESI), digital mammography, tactile imaging, magnetic source imaging (MSI), medical optical imaging, single-photon emission computed tomography

(SPECT), and ultrasonic and electrical impedance tomography (EIT) [2].

Imaging technologies play a vital role in the diagnosis of abnormalities and therapy, the refined process of visual representation which contributes to medical personnel access to awareness about their patient's situation [3, 4]. Electroencephalography (EEG), magnetoencephalography (MEG), and electrocardiography (ECG) are recording and measurement techniques that are not responsible to produce images, but these represent the data as a parameter graph *vs.* time or maps which shows the susceptible information with less accuracy. Therefore, these technologies can be said to form medical imaging on a limited scale. Worldwide, up until 2010, approximately 5 billion medical imaging techniques studies have been shown [5].

In the United States, approximately 50% of total ionizing radiation exposure is composed of radiation exposure from medical imaging [6]. Medical imaging technologies are used to measure illnesses, manage, treat, and prevent. Nowadays, imaging techniques have become a necessary tool to diagnose almost all major types of medical abnormalities and illnesses, such as trauma disease, many types of cancer diseases, cardiovascular diseases, neurological disorders, and many other medical conditions. Medical imaging techniques are used by highly trained technicians like medical specialists, from oncologists to internists [1].

Medical imaging technologies are mostly used for medical diagnoses. Medical diagnosis is the process of identification of patient disease and its symptoms. The medical diagnosis gives the information about the disease or condition needed for treatment that is collected from patient history and physical checkups or surveys. Due to no specificity of the many signs and symptoms of a disorder, its diagnosis becomes a challenging phase in medical science. For example, the case of erythema (redness of the skin) gives a sign of many diseases. Thus, there is a need for different diagnostic procedures the determination the causes of different diseases and their cure or prevention [7].

Historically, the first medical diagnosis composed by humans was dependent upon the observation of ancient doctors with their eyes, ears, and sometimes examination of human specimens. For example, the oldest methods were used to test on body fluids like urine and saliva (before 400 B.C). In ancient Egypt and Mesopotamia, doctors were able to measure the problem of the digestive tract, blood circulation, heartbeat, spleen, liver menstrual problems, etc. But unfortunately, medicine for curing diseases was only for wealthy and royal people.

At around 300 B.C., the use of the mind and senses as diagnostic tools was promoted by Hippocrates. He got a reputation as the "Father of Medicine." Hippocrates supported a diagnostic protocol by testing the patient urine, observing the skin color, and listening to the lungs and other outward appearances. The link between disease and heredity also had been recorded by them [8]. In the Islamic world, Abu al-Qasim al-Zahrawi (Arabic physician) provided the first report on a hereditary genetic disease referred as hemophilia. In this report, he wrote about a family of Andalusia, whose males died due to hemophilia [9]. In the Middle Ages, many different techniques were used by physicians to detect the causes of imbalance function of the body. Uroscopy was the most common method of diagnosis. The patient's urine was collected in a special type of flask known as "Matula." Urine was checked on the basis of color, smell, density, and presence of precipitate [10]. The viscosity and color of the blood were also examined by physicians to detect chronic or acute diseases [11]. The pulse rate, power, and tempo of a patient's artery were observed by physicians through a technique known as palpation [12]. In Middle Ages, physicians were also used to combining the study of medicine and zodiac signs [13].

In the 19^{th} century, X-rays and microscopes were the diagnostic tools that helped to diagnose and treat illnesses. At the beginning of the 19^{th} century, medical doctors diagnosed diseases by the examination of symptoms and signs. By the 1850s, many diagnostic tools such as ophthalmoscopes, stethoscopes, and laryngoscopes lead to evoke the medical doctors with the sensory power to develop other novel methods and techniques for diagnosing different illnesses. And in this way, a series of diagnostic tools including chemical tests, bacteriological tests, microscopic tests, X-ray tests, and many other medical tests were generated [8].

Medical imaging techniques are developed after the discovery of X-rays. In November 1895, Wilhelm Conrad Roentgen discovered X-ray. He got the Nobel Prize in 1901 for his discovery. Radiologists gave names to X-ray basis as "X-rays" or "plane film" used for diagnosing bone fractures and chest abnormalities. Fluoroscopy was developed due to a more powerful beam of X-ray for diagnosing the patient abnormalities. In 1920s, radiologists started giving information about various diseases like cancer of the esophagus, ulcers, and stomach. Fluoroscopy is now converted into computed tomography (CT).

Today CT scan is commonly used to diagnose many diseases. The mammography technique also uses an X-ray beam, to generate high-resolution breast images, monitoring breast cancer. In the 1940s, the X-ray tomography technique was developed, looking for a desired part of the tissue. In this technique, the whole process was accomplished by rotating the tube of X-ray focus on part of the tissue. Today, tomography is replaced with advanced imaging techniques such as CT scanning or computerized axial tomography (CAT) scanning. X-ray is also a source of a technique known as "angiography," which is used to obtain images of blood vessels. In 1950s, diagnostic imaging tests along with nuclear medicine were started. Radioactive compounds are used as X-ray sources rather than X-ray tubes. Radioactive compounds produce gamma rays. They are joined with other complexes that are an essential part of the disease analysis to study a certain illness. For an instant, technetium 99m is combined with methylene diphosphonate, which is absorbed by bone tumors. In this way, breast or lung cancer spread to other body parts such as bone can be detected from this type of nuclear bone scan technique [14].

2. Advance Modalities in Medical Imaging

Many advanced techniques are developed and can be explained with their principle of work, application in

medical labs, and development in imaging techniques. Computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), digital mammography, and sonography are included in advanced medical imaging techniques. These are all mentioned below to understand their advantages and applications in the diagnosis, management, and treatment of different diseases such as cardiovascular disease, cancer, neurological illnesses, and trauma. These techniques are readily used by clinicians because through images, they can easily choose how to manage diseases.

3. Computed Tomography (CT)

In the 1969s, Hounsfield invented the first CT-scanner prototype [15]. Computed tomography is also known as X-ray CT. A CT scan is used by radiologists, biologists, archaeologists, and many other scientists to generate cross-sectional images of different scanned objects. A modern CT scanner system is shown in Figure 1. In the medical field, technicians use CT scanners, machines to produce the images that lead to diagnosing the abnormalities and other therapeutic measurements. In this technology, X-rays are produced from different angles that are eventually processed by computers to create tomographic images. This computer-based technology has been greatly improved, developing reconstructed images with high revolution [16]. In the pharmaceutical industry, it has been used to study and improve the medicine manufacturing process to generate good quality products [17].

Positron emission tomography (PET) and single-photon emission computed tomography (SPECT) are the types of CT scan. An X-ray generator is used to generate the X-rays that rotate nearby the object to be scanned. X-rays are detected by an X-ray detector located on the opposite side of the source of X-rays. A sonogram is obtained, which is a visual representation of raw form data. This scanned data is processed in the form of tomographic reconstruction that leads to generating a series of cross-sectional photos. CT scan is performed by special individuals called radiology technologists or radiographers. Over the last two decades, CT is used largely in many clinical labs in different countries [18]. According to an estimated study, almost a 72 million scans were achieved in the US in 2007 and 80 million scans in 2015 [19].

CT is an effective technique for monitoring various types of cancers such as cancer of the bladder, kidneys, skeleton, neck, and head and for diagnosing infection [20–22]. CT also identifies distant metastases to the lungs, skeleton, liver, and brain. CT has made a high impact on the brain and lungs [23, 24]. CT scan is the best method than other techniques in detection as well as recording modifications in tumor mass during treatment [25]. It may show a bloated belly with enlarged lymph nodes in patients with bronchus carcinoma. In this way, CT scan help in performing before surgery [26]. Another major application of CT scans is the detection of heart diseases like myocardial disease, congenital heart disease, and coronary artery bypass grafts [27]. Gastroenterologists mostly use computed tomography for the analysis of the liver or pancreas of patients. Tumors of size 1.5-2.0 cm in diameter can be detected by CT scan. Furthermore, biliary obstruction caused by lesions can also be monitored by this technology [28]. One of the rewarding roles of technology is to study suspected intra-abdominal abnormalities with 95% accuracy, and treatment decisions can be easily made [29].

A big drawback of computed tomography is that large masses within the gastrointestinal tract may not be visible during the abdominal investigation. There is also no finding of some of the mucosal abnormalities by it. CT scan is highly useful to manage abdominal disorders such as carcinoma of the stomach, esophagus, and rectum more accurately as compared to other modalities [30, 31] The middle column of the spine can be visualized by using the computed tomography technique during dislocation type of fractures in many thoracolumbar fractures. CT also detects lesions and provides nonsurgical management of some disorders, for example, unstable burst damages [31].

All spinal injuries are unstable and known as translational injuries. Before surgery of such patients, complete information about the site of ligament discontinuity of respective vertebrae is provided by computed tomography. It also gives the prediction of whether Harrington-rod stabilization is possible or not. CT scan can provide detailed evidence of distraction injuries and fractures. For example, flexion-distraction injury between the 11^{th} and 12^{th} thoracic vertebrae and spinal injury between the 2^{nd} and 3^{rd} lumbar vertebrae have been scanned by computed tomography scan as shown in Figures 2 and 3, respectively [32].

Advanced CT bone imaging techniques include volumetric quantitative CT (QCT), high-resolution CT (CT), and micro-CT. High-resolution CT and high-resolution MR are generally used *in vivo*; micro-CT and micro-MR are usually used *in vitro* systems. These advanced modalities are used for bone imaging to investigate bone diseases especially osteoporosis and bone cancer. In osteoporosis, disorder advanced CT bone imaging provides information about bone mineral density (BMD), bone strength, a risk factor for osteoporosis, and recovery factors after medication or bone therapy.

Dual-energy X-ray absorptiometry (DXA) and volumetric QCT are the quantitative methods used for weighing the macrostructure of suspected bone. High-resolution CT and micro-CT methods are applicable for measuring the microstructure of trabecular bone without any invasiveness or destructiveness. CT and MRI have been used to obtain bone structure. However, the CT field has been more developed as compared to other techniques, because there are more advantages of CT-based modalities; for example, QCT generates three-dimensional (3D) images in such a way that trabecular and cortical bone can be distinctly measured. vQCT technology is quicker than MRI [33].

4. Volumetric Quantitative CT (vQCT)

Initially, QCT has been used to measure trabecular BMD of the forearm and lumber midvertebrae through a particular transverse CT slice. The measurement of BMD is a static

Figure 1: CT scanner.

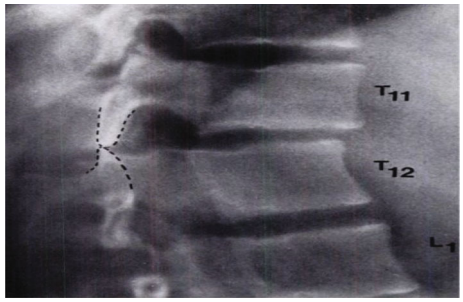


Figure 2: Distraction injury scanned by CT scan (showing damage occurrence at the 11th and 12th thoracic vertebrae).

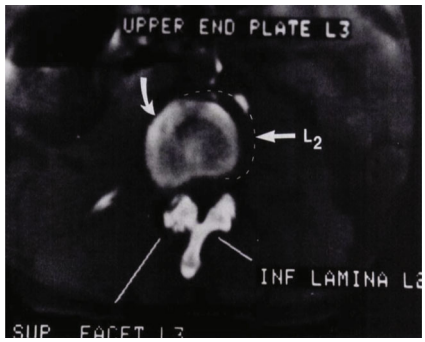


Figure 3: An axial scan of a spinal injury by computed tomography (CT) at the 2nd and 3rd lumbar vertebrae.

property of the advanced spiral QCT [34]. Trabecular bone in the spine and cortical bone in the hip may be indicated by this technology to estimate the fracture risk [35].

For the improvement of the 3D structure of the cortex, almost 0.5 mm isotropic spatial resolution is required, but still, almost 1.5 to 2 mm resolution is provided by QCT which is not adequate to make accurate images. This is a drawback of QCT. In general, the measurement of accurate cortical thickness for the femur is easier than the thickness in the spine, especially in aged people. Researchers have shown that women grow faster not only with small vertebrae as well as reduced bone mass but also with a slow rate of increase in cross-sectional area as compared to men [35]. QCT is a CT imaging technique that can provide information about bone density. For example, a QCT scan of the femur for the measurement of macrostructure and bone mineral density (BMD) has been shown in Figure 4(a) [33].

5. High-Resolution CT (hrCT)

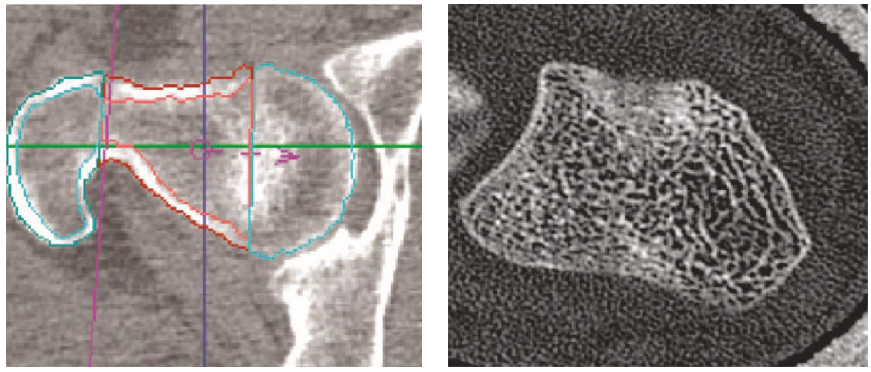
High-resolution CT is a modern CT scanner that usually requires a high radiation dose but produces high-resolution images of bone such as forearm bone submitted to CT to determine the trabecular and cortical network and texture as shown in Figure 4(b) [33]. According to many different cross-sectional studies, CT gave better imaging results in distinguishing fractured vertebral trabecular structures from nonfractured structures as compared to DXA measurements of BMD [36].

6. Micro-CT (μCT)

Micro-CT with 1-100 μm spatial resolution is typically known as microscopy. Micro-CT has abilities to replace the standing techniques used in in vivo measurements in rats and mice like animals. Initially, the micro-CT technique used synchrotron radiations to obtain ultra-high-resolution applications [37].

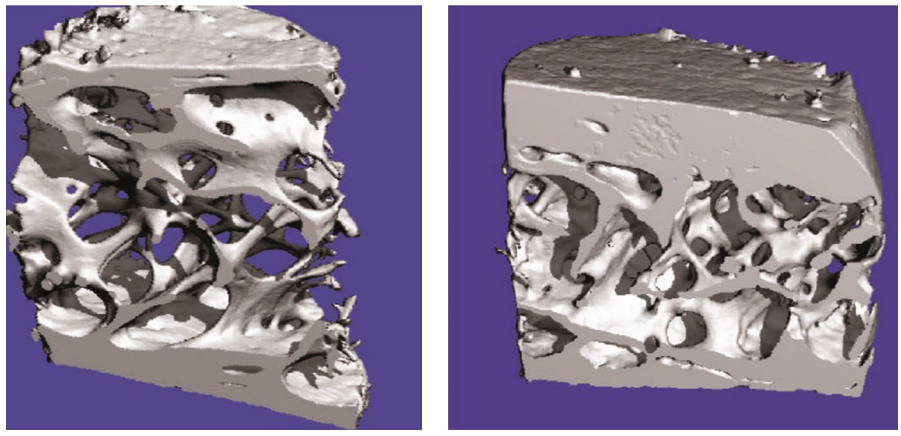
Mostly, now, the convenient method of X-ray tube-based micro-CT is used in university-based research laboratories and special clinical centers. To make 3D structures of bone, some special software (for example, FEM) is attached with a micro-CT scanner. Finite element modeling (FEM) is a software mostly used in engineering. Its goal is to help information of 3D structures of bone for analysis of fractured bone part structures from nonfractured bone structures. Currently, structural models are generated by volumetric QCT, and computer-based programs give the element elastic properties from the bone density at the site of elements [38].

Arlot and coworkers determined the 3D microstructure of bone of postmenopausal women with osteoporotic disease; they have accomplished treatment with proper strontium ranelate therapy for 36 months [39]. Researchers investigate these 3D micro-bone structures as shown in Figure 5 [33]. Over the past two decades, considerable development occurred in imaging technologies for osteoporosis bone disease analysis. Despite the development in these technologies, there are many challenges for bone imaging,



(a) (b)

Figure 4: (a) Femora undergo vQCT to determine BMD and macrostructure. (b) Ultradistal forearm undergoes CT to measure the structure of the trabecular complex network and its texture.



(a) (b)

Figure 5: Microstructure of transiliac bone biopsies is determined to undergo 3D micro-CT of two postmenopausal women who have accomplished strontium ranelate therapy for 36 months: (a) strontium ranelate therapy; (b) placebo.

such as the sample size, spatial resolution, complexity, radiation exposure, time, and cost. Finally, there is still a requirement of high accuracy, availability, reproducibility, and proper monitoring procedure for better bone imaging [33].

CT scan for lung disease is highly used in present days. For example, a low-radiation helical chest CT scan is used to investigate lung cancer (bronchopulmonary cancer) [40, 41]. Another lung disease is the most common type of progressive idiopathic interstitial pneumonia mostly in adults known as idiopathic pulmonary fibrosis (IPF). In IPF patients, CT-based methods include density histogram analysis, CT scan of whole lungs, and density mask technique, and other structural or texture classification methods are greatly used to examine the pulmonary function, lung disease progression, and mortality. For example, lung images of a 73-year-old male IPF patient have been taken by CT as shown in Figure 6. These methods have the property of time efficiency, availability, and reproducibility. Still, there are many issues interrelated to computer-based CT in IPF disease analysis. But it is promising by scientists to develop advanced CT imaging techniques that must play a vital role in the future to manage lung diseases as well as other abdominal diseases [42, 43].

There are many possible reasons for the usage of CT scans, for example, to determine or investigate the acute stroke in the patient's head. CT is applicable to establish the diagnosis, investigate the type of stroke, respond to surgery, and finally manage the disease [44]. CT scan of the head is also responsible for the investigation of dementia disease. Accurate diagnosis is directly related to proper management of symptoms and signs of dementia. Patients with treatable lesions can also be identified by computed tomography [45]. Abdominal computed tomography is a new technology for identifying fungal infection known as disseminated fungal infection (DFI) in pediatric cancer patients. Currently, abdominal CT is greatly applicable for the diagnosis and management of DFI in cancer patients [46]. During the past few years, the usage of CT scan has become a national trend in emergency departments, especially in the US. Computed tomography plays an expanding role in diagnosing acute and chronic diseases as well as life-threatening diseases such as stroke, head injury, major trauma, heart disease, abdominal pain, pulmonary embolism, severe chest pain, and renal abnormalities [47].

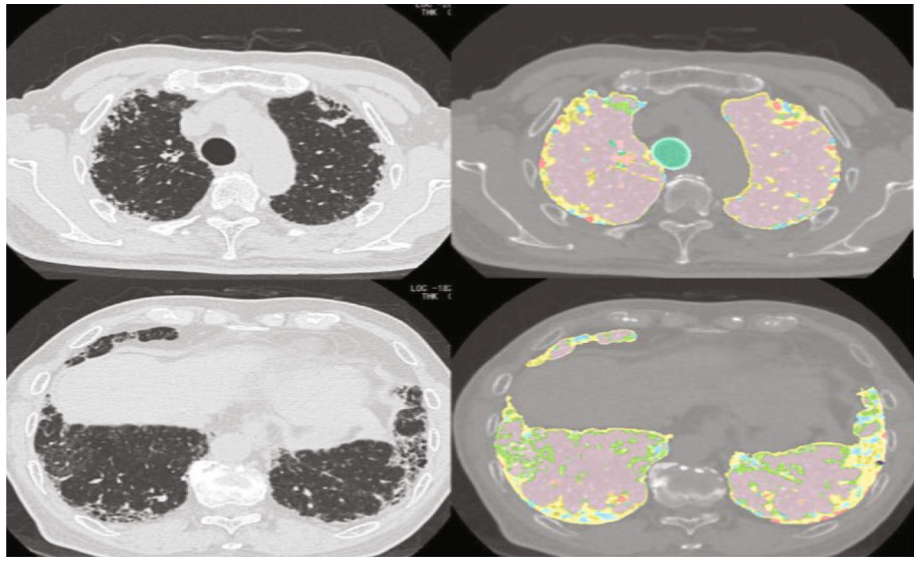


Figure 6: Images derived using a system known as GHNC (Gaussian Histogram Normalized Correlation). Lung images of a 73-year-old male IPF patient have been taken by computed tomography (CT). Light blue and yellow color, fibrosis; dark blue color, emphysema; pink color, normal; and light green color is indicated as ground-glass opacity.

7. 3D Ultrasound Computed Tomography (3D USCT)

3D USCT is a promising technology for imaging breast cancer. Simultaneous recording of reproducible reflection, speed of sound volume, fast data collection, attenuation, and high image quality production are all the main advantages of the USCT system. 3D USCT system is a full potential device used for clinical purposes. Only in 4 minutes, the full volume of breast can be picked up [48].

8. Risks of Computed Tomography

Computed tomography risks are small but if these small risks are produced by million numbers of scans, they may drive into serious public health concerns in the future, especially from pediatric CT. The risk of cancer due to computed tomography scanning is increasing [49]. Children are of specific concern due to the sensitivity of radiation-induced cancer as compared to adults. According to a study, the risks of leukemia and brain tumors are mostly revealed after exposure to radiation from CT scans [50].

According to the authors of a recent report "long-term risks from CT scans directly would require very large-scale studies with lifelong follow-up." The author gave this statement after the study date on CT scan exposure leading to the risk of future cancer [51]. Radiologists should be a source of discussion earlier to perform imaging technologies that contain high doses of radiation. They should also know about the risk factors of imaging technologies that can cause more adverse effects than recovery. Both families and patients should also raise the question/answers about the benefits and risks of CT scans [52].

9. Positron Emission Tomography (PET)

Positron emission tomography is a nuclear medicine functional technique that is used to display the total concentration of radioactive labeled elements in the body with clear images. It has the potential to diagnose biological processes within living bodies and is highly applicable for clinical purposes [53]. 3D images of positron-emitting radionuclides within the body are made by a computer system. In PET-CT scanners, 3D imaging is created with the help of a CT X-ray scan implemented on the patient body in the same machine and session. Positron emission tomography (PET) and nuclear magnetic resonance (NMR) both are quantitative radiological techniques that display information about biochemistry and physiology, normality, or abnormality. Nuclear magnetic resonance is not more sensitive to give high-resolution images by the distribution of substances except hydrogen. It can measure the total concentration of ATP and creatine phosphate (CP) in particular areas of the brain. Thus, both NMR and PET performed their specific function in the diagnosis. In 1953, the first PET system was established at Massachusetts General Hospital. It was followed by many other devices in a series manner such as tomographic positron camera, PET scanner, and other PET instruments [54].

10. Working Principle of PET

The PET technique detects radioactivity emission when a small concentration of radioactive tracer is intravenously injected. These tracers are frequently labeled with carbon-11, nitrogen-13, oxygen-15, and fluorine-18, as shown in Figure 7. There is no positron emitter of hydrogen. The radioactive dose amount is the same as used in CT. 10-40

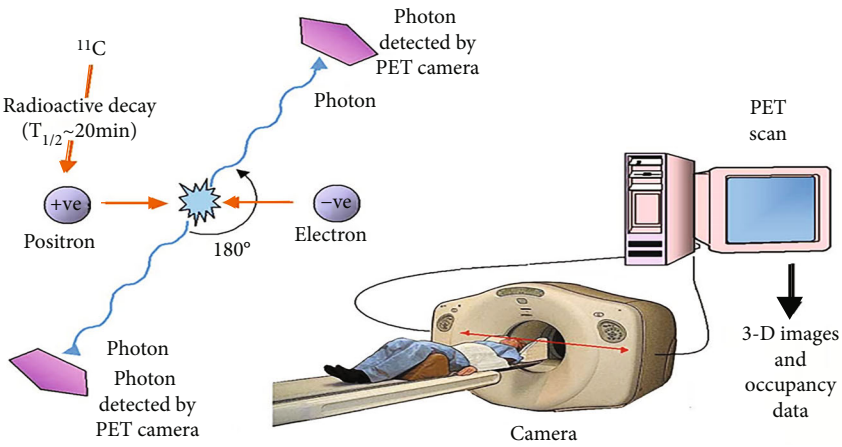


FIGURE 7: Basic principle of PET scan: (1) a positive electron (positron) is emitted by radioactive decay of radioisotope, for example, carbon-11; (2) this positron hits an electron present in the tissue to be analyzed and emits two photons having low energy; (3) scintillation crystals are present in the PET camera to absorb this emitted photon with low energy; (4) the light is produced that is converted into another signal such as electrical signals used by the computer system to produce 3D images.

minutes is required to complete the process to perform a PET scan. The patient is fully clothed during scanning. There are specific steps in PET scan processing, explained in Figure 7 [55].

Two molecular probes are mostly used to explain PET assay: 2-[F-18] fluoro-2-deoxy-D-glucose (FDG) and 3-deoxy-3-[F-18] fluorothymidine (FLT). FDG is the analog of F-18-labeled glucose, and it is used to identify diseases by changing the metabolism of glucose in heart diseases, Alzheimer's disease, and cancer. FLT is the analog of F-18 labeled thymidine and is highly used to estimate processes like cell proliferation and DNA replication by analysis of the phosphorylation process and thymidine transport. Thus, FLT and FDG are considered as best candidate probes/tracers for molecular imaging.

An early diagnosis of Alzheimer's disease can be scanned by PET technology with 93% accuracy. Huntington's disease, a hereditary disease was also detectable by PET scan. The development of PET technology provides accurate whole-body images for examining early primary and metastatic diseases. Imaging of transgenes provides information on the regulation of gene expression during cell proliferation, growth, response to environmental stimuli, the aging process, and gene therapy. Such endogenous gene expression can be monitored through the developed PET approach, which uses F-18-labeled oligodeoxynucleotides having a short single strand of almost fifteen nucleotides. In monkeys with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine- (MPTP-) induced lesions on one side of the brain, the study of restoring dopamine production by gene therapy can be assessed by PET imaging, as shown in Figure 8 [56].

The combination of PET and computed tomography (PET-CT) forms a hybrid imaging approach that is highly used to gain functional and metabolic information to measure inflammatory and infectious diseases to assess their proper treatment. PET-CT hybrid imaging technique can provide quick information during the diagnosis of a disease and its treatment response [57].

Haberkorn et al. measured FDG uptake that relates to the proliferation rate of tumor cells in the head and neck with different patterns, in two groups of patients [58]. The FDG uptake was also measured in the malignant neck and head tumors and metastases process by the use of FDG-PET. Minn and coworkers found that the uptake of FDG is related to the proliferation rate of tumor cells [59]. Jabour *et al.* measured the normal anatomy of the neck and head [60]. In this way, change in uptake FDG as investigated by PET provides necessary information for clinical and anticancer therapeutics [61].

In the human cerebellum, changes in local neuronal function by voluntary movement and tactile stimulation were also mapped with the help of the PET approach detection of brain blood flow. According to research, finger movement leads to the production of parasagittal and bilateral blood flow enhancement in the superior and anterior hemispheric cortex of the brain human brain cerebellum. The enhancement in midline blood flow in the posterior vermis of the human brain cerebellum is produced by saccadic eye movement. PET also allows the measurement of structural and functional relations in the cerebellum of the human brain [62]. The development of the PET brain imaging technology makes it possible to advance understanding of the anatomy of brain parts and map of neuroanatomical basis of cognitive processes and memory [63]. A pathogen SIV (simian immunodeficiency virus) causes infection in rhesus macaques (a type of monkeys) with acute viremia, and progression leads to infection in the solid tissues of lymphoid, and then, cellular degradation becomes a terminal disease, and death occurs in most cases. So, the FDG-PET imaging technique is used to take images from SIV-infected animals. In this way, infected groups can be distinguished from the uninfected control groups [64].

11. Future of PET Technology

The PET scan can be used to measure the concentration of amino acids, sugar, fatty acids, and receptor in the living

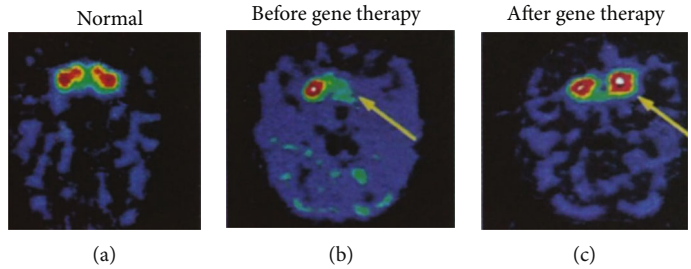


Figure 8: PET images of gene therapy in unilateral MPTP monkey as Parkinson's model: (a) image of normal dopamine production; (b) the image is representing autonomous dopamine MPTP-induced shortage (before gene therapy); (c) image of the restoration of dopamine production in the caudate and putamen (after gene therapy) [56].

body. It is a new diagnostic tool used to detect diseases such as atherosclerosis, aging, cancer, and schizophrenia, although improvement in instrumentation and modeling is still required for future purposes. Emission tomography has also been associated with a small risk of ionizing radiations [54].

12. Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging (MRI) is primarily an imaging technique that is applicable for noninvasive visualization of the anatomy and physiology of the body in both disease and health conditions. An MRI-related technique known as echo-planar imaging (EPI) was developed by physicists Peter Mansfield and Paul Lauterbur in the late 1970s [65]. Magnetic fields, electric fields, and radio waves are used in an MRI scanner to produce images of organs and the structure of the body. The SI unit of magnetic flux density (magnetic field strength) is measured in tesla (T).

The most common detections by MRI are multiple sclerosis, CNS tumors, brain and spine infections, stroke, injuries in ligaments and tendons, muscle degradation, bone tumor, and occlusion of blood vessels. MRI uses nonionizing radiation, frequently preferable compared to CT. MRI also provides excellent contrast of soft tissues; for example, the white and gray matter structure of the brain can easily be distinguished through this approach. MRI employs other different techniques such as functional MRI, magnetic resonance angiography (MRA), susceptibility-weighted, diffusion-weighted (PWI), diffusion-weighted (DWI), gradient echo, and spin-echo. It provides an image of good quality without requiring repositioning of the patient [66]. There are several benefits to MRI such as it is a painless, noninvasive technique with high spatial resolution and nonionizing radiations. MRI is mostly used independently for soft tissue analysis.

13. Working Principle of MRI

An MRI machine consists of multiple components, including a slab for patients to lie on, a superconducting magnet, a protective cage, the operator's console, and computers to analyze the data and product images. During the MRI scanning process, the machine's magnet produces a strong magnetic field. Hydrogen ions align in the target body part of the patient due to a stable magnetic field. Then, bombardment of radiofrequency waves causes the alignment of lined-up hydrogen ions to move out, and then, ions return to their equilibrium state [67]. An attached computer system converts the spin echoes (signal) of hydrogen ions into the images, after several "shifting" and "working on." A microphone is also present inside the MRI unit for communication between the patient and technologist during the imaging process. Images of only the target part of the body are created through MRI radiological analysis. The physician chooses which part of the patient's body must be analyzed by imaging, to diagnose the illness of the patient [68].

14. Applications of MRI

A major application of whole-body MRI is to investigate skeletal metastases. The MRI approach allows for visualization of the tumor because the tumor matrix contains an abundance of the proton. It is a more sensitive imaging technique than skeletal scintigraphy (bone scan) in the measurement of skeletal metastases. The whole-body MRI technique is more effective for detecting lesions in the pelvis, spine, and femur. This technique is also highly used as a primary diagnostic tool for the measurement of soft tissue diseases, whole-body fat, and polymyositis disease [69].

MRI is different from other diagnostic techniques because MRI has no risk of ionizing radiation. MRI has no side effects unlike CT and PET scans. There is no loss of image quality due to the scanning of body target parts from several angles and viewpoints [70]. Dynamic contrast-enhanced magnetic resonance (DCE-MRI) has been developed for the detection of the tumor microenvironment and its treatment. It has been supported as a useful method and improved clinical interest [71].

An advantage of using MRI to diagnose cardiovascular diseases is that examination reveals function structure perfusion, metabolism, and blood flow in the heart. A cardiovascular MRI is a source for the detection of congenital cardiac diseases, abnormalities in the thoracic aorta, and pericardium in heart patients. During the detection of myocardial tumor or right ventricular dysplasia, tissues are differentiated due to varying imaging parameters of the MRI approach. Another application of MRI, cardiovascular MRI, is applicable for determining cardiac prognosis, ischemia in a patient with heart disease, artery arteriosclerosis, and screening the myocardial viability [72].

Schizophrenia patients show mental abnormality which leads to language processing deficits and abnormal social behavior. Functional MRI has been applicable for remarks

of such types of illnesses. The region of hypoactivity can be determined in the frontotemporal cortex of the patient brain. Soft neurological signs and symptoms have also been promoted. Functional MRI can detect abnormalities in the cerebrum; cerebral asymmetry images reveal changes in patients with schizophrenia compared with control as shown in Figure 9 [73].

Microfluidic LOC (lab-on-a-chip) is a device used as an emerging technology in medical laboratories. Sample (consist of suspensions of cells) and reagents react on these devices. For monitoring reaction on LOC, MRI is considered an ideal tool. MRI records the signals from the expended fluid leavings in the device. MRI combined with MRS (magnetic resonance spectroscopy) monitors fluid flow processes, chemical reaction separations, and diffusion processes in LOC. But MRI and MRS both show low sensitivity. In the future, there is a hope that MRI will be applicable for the advancement of microfluidic LOCs with powerful usage in medical diagnostic libraries [74].

Mutation in BRCA1 and BRCA2 genes leads to losing their ability to repair the damaged DNA, causing cancer, especially breast cancer. MRI diagnoses breast cancer which is due to a genetic mutation. Mostly, these hidden breast cancers are not detected by mammography. For a decade, doctors use MRI imaging tools to detect breast cancer [75]. According to previous research, 27-37% of patients have shown lesions on MRI, which are not seen through mammography. The researchers noted that mammography had a low value of positive prediction of 52.8%, as compared to MRI which is high at 72.4% [74].

Molecular MRI employed for specific and early detection of pulmonary metastatic cancer cells can improve its treatment. In research, pulmonary cancer cells are besieged by iron oxide nanoparticles having the ability to bind with ligand expressed on the cells. Then, images were taken by high-resolution hyperpolarized ^{3}He MRI (HP ^{3}He MRI). The study confirmed that HP ^{3}He MRI pooled with targeted superparamagnetic iron oxide nanoparticle (SPION) contrast agent detects specific and early metastatic pulmonary cancer in mice. A researcher used the LHRH-SPION agent to explore new drug procedures. For this purpose, they injected breast adenocarcinoma cells into mice and then detect pulmonary defects as cancer formation in mouse lungs with the help of LHRH-SPIONs and HP ^{3}He MRI results are shown in Figure 10 [76].

Multinuclear 3D solid-state MRI allows images of tooth bone and calcium phosphate components of bone substances. It also gives information on the bone composition and texture of the bone [77]. Recent neuroimaging techniques including high-resolution MRI can investigate myeloarchitectural patterns in the cortex of the human brain. The bands of myelination have been revealed by the staining technique. Now, the same band in good quality image form can also be obtained by high-resolution MRI imaging technique. Although the advanced technology has been largely applied in the visual system, further improved methodologies are required for the investigation of another brain region. To overcome high ratio of "signal" to "noise" is a challenge for MRI machines that is produced due to the

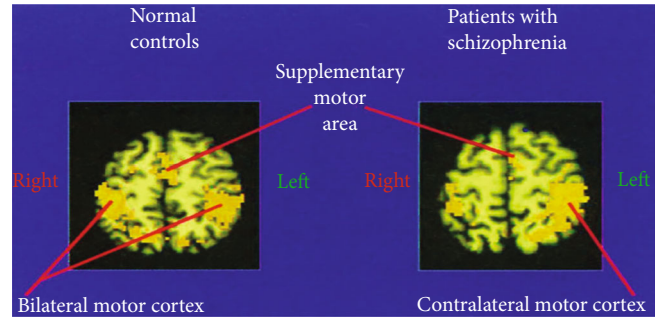


Figure 9: Abnormal cerebral asymmetry in schizophrenia patients compared with control is shown by functional-MRI imaging technique.

increased resolution of the image [78]. In this way, fMRI and other types of MRI have been powerfully used to change our understanding of diseases, their causes, and how to manage the conditions.

15. Simultaneous Imaging with MRI and PET

In *in vivo* study, imaging of small animals, for example, mouse imaging by combined MRI and PET modalities, produces constant information of different parts of the body. An experimental study reveals that the combined PET/MRI technique improves the understanding of malignant tissues and heterogeneous tumors, edema, and necrosis that are not done by MRI alone. Particular ionic ^{18}F is also used for PET/MRI combined imaging of small animals as models of bone metastasis, osteoporosis, and arthritis to study the complete skeletal system [79].

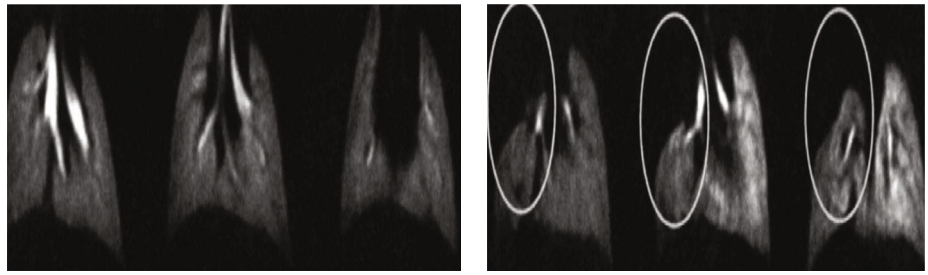
16. Risks of Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is highly expensive, low in sensitivity, and time-consuming for scanning and processing compared to other imaging modalities. A probe with a bulk quantity may be needed for MRI. It cannot detect abnormalities of intraluminal body parts. It gives no real-time information. And it can create a suffocating environment for some people [2].

17. Single-Photon Emission Computed Tomography (SPECT)

Single-photon emission computed tomography (SPECT) is an advanced imaging technique using gamma rays and provides three-dimensional (3D) representations of objects with high accuracy. In 1963, Kuhl and Edwards [80] gave the first report about single positron emission computed tomography. Gradually, modification with new instruments such as computer-attached systems and rotating gamma cameras leads to the development of a novel modality of single-photon emission tomography.

SPECT has become a great medical imaging technique used in research and clinical area. A dual-headed single-photon emission tomography (SPECT) system has been



(a) (b)

Figure 10: Breast adenocarcinoma mouse model formed for the detection of lung metastases. (a) High-resolution hyperpolarized ^{3}He MRI (HP ^{3}He MRI) images were taken from the control mouse. Screening as normal ventilation forms of lungs. (b) After injection of LHRH-SPIONs, images were produced from human breast adenocarcinoma abnormal mouse (model), showing defects in the right lobe (under circles).

Figure 11: Dual-headed single-photon emission computed tomography (SPECT) system.

shown in Figure 11 [81]. It monitors the 3D information of an object by producing series of thin slices from tomographic images. These essential tomographic images can improve the ability detection of deep and very small fractures in patients [81].

SPECT assesses the multiple two-dimensional (2D) images from different angles by using high-energy gamma rays. Data is reconstructed and recorded, and 3D images of the target portion of the body are produced by a computer program. SPECT is greatly used in clinics and research laboratories like other tomographic modalities such as PET, MRI, and CT. SPECT and PET both use the radioactive tracer and then measure the emitted gamma rays. In the case of SPECT, emitted gamma rays by radioactive tracers are directly detected by the detector. The computer system analyzes the data from the detector and produces the true image of the area where the radioactive tracers are injected. SPECT imaging technique is less expensive than exclusively used for imaging small animals. It is sensitive to monitoring target bone metabolism, myocardial disorders, and blood flow in the cerebrum [82].

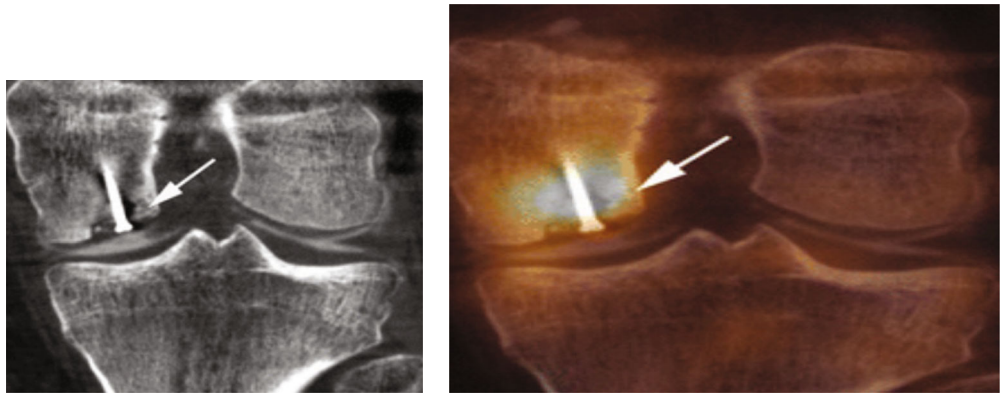
SPECT has also been designed for imaging of the brain known as neurochemical brain imaging. It has a powerful imaging technique to elaborate the neuropsychiatric diseases. It is an essential developmental technique that has great potential to monitor the pathophysiology and many other complex disorders of the brain [83].

Today, hybrid SPECT and CT are progressively employed and available in the nuclear medicine field. SPECT/CT provides exact abnormal bone turnover during inflammation, bone tumor, bone regeneration, bone infection, and trauma in complex bone joints such as the knee, hip, foot, shoulder, and hand/wrist. In most cases, CT with specificity and SPECT with high sensitivity are performed together for a complete diagnosis. SPECT/CT is also responsible for giving the proper information about therapy planning. For example, it tells us about the decision of using SPECT/CT alone or joint arthroplasty. It was observed that joint arthrography of the knee gives better results as compared to SPECT/CT alone as shown in Figure 12 [84].

SPECT/CT technique has great potential for the measurement of articular cartilage, soft bodies, and synovial structures. It also gives promising results in the examination of osteochondral abnormalities [84].

18. Risks of Single-Photon Emission Computed Tomography (SPECT)

Single-photon emission computed tomography (SPECT) is expensive and requires additional care for performance with



(a) (b)

Figure 12: (a) The SPECT/CT scan of the knee joint shows clear articular cartilage. (b) The SPECT/CT arthrography image of the knee joint showing enhance the value of SPECT/CT screening.

radioactive materials. Like PET, it uses ionizing radiation and creates radiation side effect for the patients [2].

19. Digital Mammography

Mammography is a technique used to screen and diagnose human breasts. Low-energy X-rays (30 kVp) are used in the mammography in performing early diagnosis and screening of human breast cancer. Ionizing radiations are used in mammograms to create images for analysis of abnormal conditions. Ultrasound is usually employed to give additional information about masses, detected by mammography. MRI, discography, and positron emission mammography (PEM) are also supporters of mammography. Mammography is more accurate for women 50 or above 50 years old as compared to younger women because old women have high breast density [85]. Today, conventional mammography has been replaced by digital mammography.

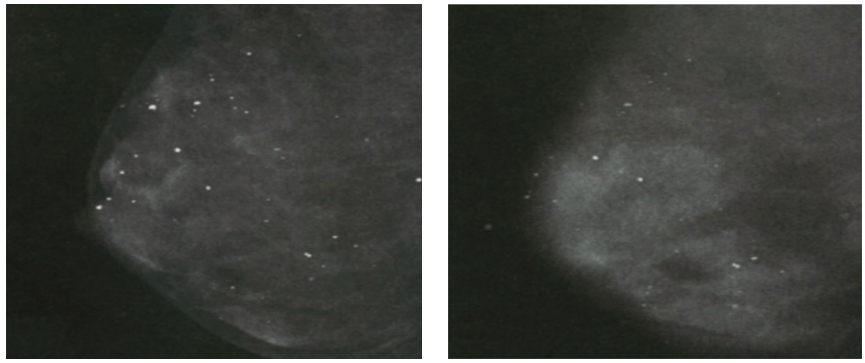
An advanced technique is used for creating 3D images of breast tissues for detailed analysis of breast cancer, known as 3D mammography. When 3D mammography is used along with usual mammography, it gives more positive results [86]. Cost-effectiveness and high radiation exposure are of high concern to 3D mammography [87].

Digital mammography is a special form of mammography employed to investigate breast tissues for breast tumor study. Digital mammography contrasts with film (conventional) mammography by using a special detector that detects the transmitted X-rays energy and converts it into an image signal by a computer system instead of a film X-ray. Digital mammography is a rapid and advanced modality that has the potential for diagnosis and proper screening of breast cancer. Such new diffusion technologies can alter the health care pattern through many mechanisms. There are different results related to breast health care for digital-screen and film mammography. Digital mammography may also change the application of diagnostic services ensuring mammograms with positive screening [88].

Digital mammography has been considered a better technique as compared to film mammography in the detection of breast cancer in premenopausal, premenopausal, and young women. A digital system has more cost (approximately 1.5 to 4 times) than a film system. Digital mammography has the advantage of diagnosis in computer-based system that generates images with easy access and better-quality transmission, recovery, and image storage. Advanced digital mammography uses an average low dose of radiation without cooperation with diagnostic accuracy [89].

A healthy breast tissue mammogram recorded by digital mammography is clear as compared to a mammogram on film as shown in Figure 13 [90]. Scientists used both digital mammography and film mammography for 42,760 women's breast X-rays. Cancer is almost equally well detected through these techniques. But digital mammography detected 28% more breast cancer in younger women or those under 50, who have dense breast tissues. Digital mammography uses a specific detector that captures the transmitted X-rays and sends the information in the form of energy that converts into an image through a computer system [90]. Digital mammography screening for breast cancer is not cost-effective, relative to conventional mammography [91].

Now, screening MRI is highly used as an adjunct to mammography, recommended for women to ensure 20-25% or more threat of breast cancer [92]. Currently, breast cancer diagnostic programs have been recognized widely at least in 22 countries. Collective struggles lead to the formation of the Breast Cancer Screening Network (IBSN), an international program for policymaking, administration, funding, and handling of results that come from breast cancer screening of huge populations [93]. The ultrasound technology combined with mammography is used to detect elevated risk factors for breast cancer. According to research on elevated risk factors, by using ultrasound with mammography, more than 90% of cancer risk was seen in over 50% of women having dense breast tissues, and 25% of cancer risk factors were determined in women having just 26%-40% dense tissues of the breast. It is suggested that screening with ultrasound may be beneficial for those women having other risk factors and less dense breast tissues [94].



(a) (b)

Figure 13: (a) A healthy breast tissue mammogram recorded by digital mammogram; (b) the same breast tissue mammogram recorded by film mammography. White spots shown in the above images are deposits of calcium which can consider the mark of cancer when they form clusters.

20. Medical Ultrasound

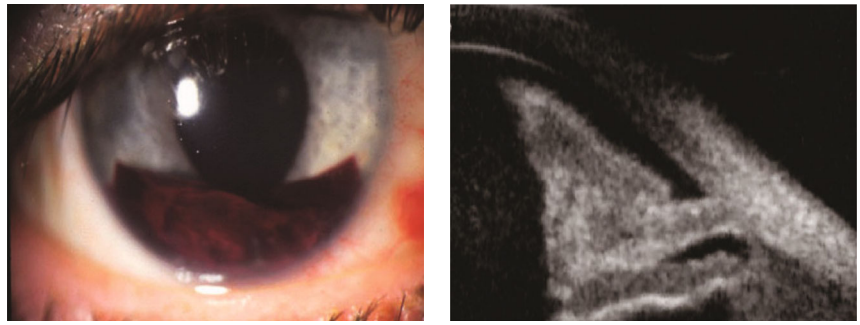
The ultrasound imaging technology was used earlier as a diagnostic tool for brain images. Today, ultrasound is a widespread imaging technology used in diagnostic laboratories and clinics. It is free from radiation exposure risk, comparatively less expensive, and highly portable as compared to other imaging techniques like MRI and CT [95]. This system is used in different fields. In the medical field, ultrasound uses sound waves of high frequency, to diagnose the organs and structure of the body. The ultrasound system is performed with high frequency. Special technicians or doctors use it to observe the kidney, heart, liver, blood vessels, and other organs of the body. The most critical component of ultrasound is a transducer. An ultrasound transducer can convert an electrical signal into sound waves and sound waves into an electrical signal. An ultrasonic image or sonogram is formed by transmitting pulses of sound waves into tissues using a special probe. Different tissues reflect these sounds to a different degree. These reflected sound waves (echoes) are detected and presented as an image with the help of the operator. In the medical field, there are many applications of ultrasound. Ultrasound scanning is a very effective and reliable technique that is greatly used for monitoring normal pregnancy, placenta previa, multiple pregnancies, and different abnormalities during pregnancy and rest [96].

An ultrasound imaging technique also known as transvaginal ultrasound (TVS) alters our understanding of the management and diagnosis of pregnancy. The TVS gives clear knowledge of early pregnancy problems. It investigates the pregnancy location as well as viability. *In utero*, the TVs have determined the fetal heart activity that is initial proof of pregnancy viability. Abnormal development in fetal heart rate pattern indicates the subsequent miscarriage. Less fetal heart rate especially at 6 to 8 weeks demonstrates subsequent fetal disease. Fetal heart pulsation can be seen on TVS. The routine use of TVS also leads to development in managing early pregnancy failure. Awareness of pregnant women and improvement in early pregnancy units can directly manage miscarriage. TVS is a very sensitive approach for diagnosing early miscarriage. It detects trophoblastic tissue and blood flow in the intervillous space, and the use of color Doppler images leads to estimate the level of expectant management [97].

Functional ultrasound (fUS) is highly applicable for imaging the brain and detects transient alternation of blood volume in the brain at high resolution than other brain imaging modalities. The blood volume in small vessels can be measured by functional ultrasound (fUS), which uses plane-wave illumination with a high frame rate. Functional ultrasound can detect the brain's active portion [98]. Ultrasound has major advantages as noninvasive and out-patient scanning in children to investigate neuromuscular disorders. Ultrasonography has been used for finding the normal function of muscle, muscle contraction, muscle thickness, and muscle fiber length. Real-time ultrasound is used for muscle imaging. When ultrasound applies to neuromuscular patients, different muscle disorders can be detected by a pattern of muscle echo, such as a bright spotted pattern of increase in muscle echo obtained in muscular dystrophy patients and a moderate increase in echo showed in spinal muscular atrophy [99].

Currently, early care physicians can easily understand complex patient conditions with the help of advanced 3D ultrasound algorithms. And high-speed networks are used in special health centers to enhance patient care facilities [100]. 3D ultrasound is widely used due to the reason of 2D ultrasound limitations. Clinical 3D ultrasound experience has an advantage in the diagnosis of disorders and produces a 3D image that guides invasive therapy. Further improvement in 3D imaging software, as well as hardware, will lead to routine usage of this tool [101].

High-resolution ultrasound which is also known as ultrasound biomicroscopy (UBM) has clinical applications in imaging the human eye. UBM uses 35 MHz or above frequencies to provide images of high resolution as compared to conventional ophthalmic ultrasound techniques. UBM can be used to diagnose ocular trauma and complex hypotony. It can determine eye lens displacement, iridodialysis, zonular flaw, cataract, lens subluxation, and hyphemia. Hyphemia is a condition in which blood diffuses the anterior chamber of the eye due to injury. Hyphemia scan image by Sonomed UBM is shown in Figure 14 [102].



(a) (b)

Figure 14: (a) Hyphemia is due to injury; blood diffuses the anterior chamber of the eye. (b) Hyphemia scan image by Sonomed UBM.

UBM can detect several eye abnormalities, as it allows for the diagnosis of trauma, glaucoma, and foreign bodies. It can evaluate eyelids, neoplasms, normal eye anatomy, and extraocular muscles during strabismus surgery. Currently, for high-resolution diagnostic eye imaging, new advanced technologies such as pulse encoding, transducer array, and ultrasound combination with light are developed [102].

A high-frequency (40 MHz) ultrasound imaging technique has been developed for checking the programmed cell death process known as apoptosis. It is a noninvasive procedure used to monitor the apoptosis process that happens because of agents especially anticancer agents in cells of body tissue in vitro or in vivo. The procedure of detection monitored alternations in subcellular nuclear such as the condensation process after proper destruction of the cell during the programmed cell death process. The high-frequency ultrasound technique shows a high scattering rate due to these intense alternations (approximately 25-50-fold) in apoptotic cells as compared to normal tissue cells. As a result, the apoptotic tissues show greatly brighter areas as compared to normal tissue. In the future, this noninvasive imaging technique will use to check the effect of anticancer treatment and chemotherapeutic agents in laboratory model systems and then in patients [103].

Another application of ultrasound technology is the successful detection of renal masses. According to research results, 86% carcinomas and 98% renal cysts were accurately determined among 111 patients by the ultrasound imaging technique. Ultrasound is a safe, simple, and cheap diagnostic tool to diagnose complex renal masses [104]. Ultrasound screening allows imaging the cartilage for checking instability, abnormal location of the femoral head inside the acetabulum, and developmental dysplasia in newborns [105]. A powerful low-frequency ultrasound system can also be used as a noninvasively drug delivery system. Many drugs and proteins having high molecular weight can be delivered easily with excessive permeability into human skin with the help of a low-frequency ultrasound modality. For example, insulin, erythropoietin, and interferon-gamma molecules are easily and safely delivered across human skin [106].

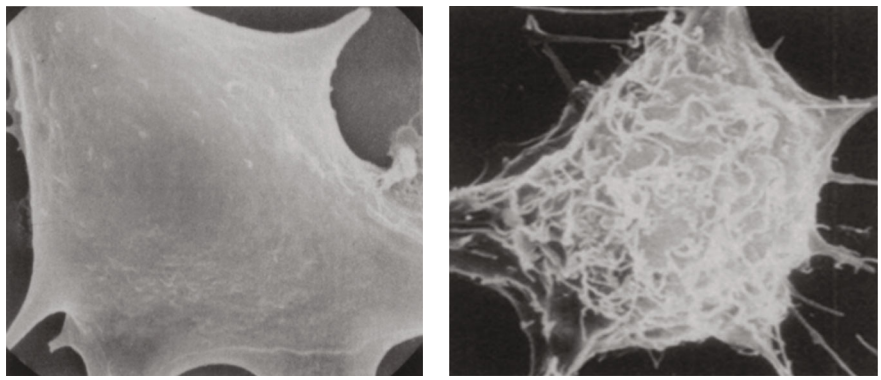
A novel molecular imaging technique known as molecular ultrasound has been used by researchers in the molecular biology field to monitor the alternation in the expression rate of molecular markers located on intravascular targets. Contrast agents used in the advanced molecular ultrasound imaging technology are mostly micro- or nanosized particles having ligands on the surface also known as microbubbles. Specific molecular markers are targeted by these microbubbles such as selectin, vascular cell adhesion molecule 1, and integrin. In this way, these agents lead to detect specific molecular markers on intravenous targets by gathering at that specific tissue site.

Molecular ultrasound has many advantages such as low-effective cost, high resolution, portable, noninvasiveness, absence of ionizing radiation, real-time imaging potential, and high availability. It has the potential for regular investigations of different abnormalities at the molecular level, such as inflammation, tumor angiogenesis, and thrombus. In addition, improvement is still required in the field of molecular ultrasound to design the novel targeting ligand to form a more effective contrast agent. In the future, advancements in molecular ultrasound imaging technology will play a clinical role with high sensitivity and accuracy for imaging complex abnormalities at the molecular level [107, 108].

21. Disadvantages of Medical Ultrasound

There are many useful applications of ultrasound in the medical field. But medical ultrasonography also had some side effects such as hormone change effect, breakage of chromosomes with very low frequency, chemical effects, and other health problems. It was examined that for 10-week gestation, chorionic gonadotropin hormone in humans increased after routine usage of ultrasound modality [109].

It was observed in an experiment that no chromosome damage occurred after diagnostic ultrasound exposure to human lymphocyte culture. But experimental results suggest that ultrasound can cause chromosome breakage with very low frequency [110]. Similarly, lots of experiments have been done to check fetal ultrasound safety. *In vivo* study demonstrates no major neurological defects, fetal growth problems, birth flaws, or childhood cancer caused by ultrasound imaging. But *in vitro* study demonstrates the possibility of some health problems that can occur due to diagnostic ultrasonography. For example, the effect of diagnostic ultrasound on the normal architecture of the mouse fibroblast cell with



(a) (b)

Figure 15: Images of mouse fibroblast cell (a) normal smooth form and (b) abnormal cell in its rough shape. Due to the diagnostic ultrasound effect, fingerlike projections are produced on the smooth surface of the fibroblast cell.

the production of fingerlike projections on the fine surface of the cell is demonstrated in Figure 15 [111].

Another disadvantage of an ultrasound system in the medical field is generating significant heat. When strong ultrasound waves migrate through the liquid environment, small cavities are produced, and these cavities expand, then collapse to each other, and generate heat. This heat-generating condition leads to create an unnecessary chemical environment [112]. Dyslexia, growth limitation, non-right handedness, and late speech-like effects are also examined after diagnostic ultrasound exposure. Continuous research is required to find the side effects of medical ultrasonography on human health [113].

22. Radiation Exposure Risk from Medical Imaging and Its Management

Radiological imaging by X-ray radiology, positron emission tomography (PET), single-photon emission computed tomography (SPECT), mammography, and computed tomography (CT) uses high-energy X-rays that leads to high radiation dose in some patients. Pregnant women and children are mostly affected by radiation exposure from radiological imaging. Possibility of cancer, genetic mutations, growth and developmental retardation in the fetus, and cardiovascular abnormalities can occur by the exposure to radiation after radiotherapy. Direct radiation exposure can cause hair loss, cataracts, skin redness, and skin damage. Radiation exposure risks can be reduced by making "National Guidelines" that aid the physician to manage their effects on patients.

Many online tools have been developed to enable physicians or technicians to record the calculation of radiation exposure from each radiological imaging technique. For this purpose, a Thermoluminescent dosimeter (TLD) is used to calculate radiation dose through different software depending on the modality being used such as Monte Carlo PENRADIO which is used for CT [114]. Magnetic resonance imaging (MRI) and ultrasound techniques are free from ionizing radiation. To minimize the risk, MRI and ultrasound can be used instead of radiological imaging techniques. Reduction in unnecessary computed tomography screening leads to the direct reduction of radiation exposure risks. Today, advanced and safe technologies are used that allow measuring signals with a low dose of radiation; e.g., low-dose computed tomography scanners permit less radiation exposure [115].

23. Advanced Machine Medical Image Analysis

4D medical imaging analysis is an advanced technology that is used in combination with different modalities such as 4D CT, 4D US, and 4D MRI. 4D CT is an excellent choice for radiation oncology, which is prone to motion artifacts. Similarly, 4D ultrasound is particularly useful in prenatal research. 4D flow MRI can help doctors diagnose and treat heart issues more precisely. For big data integration in medical imaging, researchers need to develop algorithms to store images by converting them into numerical format which will be helpful for physicians in diagnosis.

24. Artificial Intelligence in Medical Imaging

Machine learning and deep learning are branches of artificial intelligence (AI) that solves problems in medical imaging applications such computer-aided diagnosis, lesion segmentation, medical image analysis, image-guided treatment, annotation, and retrieval. AI assesses image quality, interprets image, and analyzes biomarkers, and finally, reporting is done. AI has impact on oncologic imaging. Lung cancer is one of the most prevalent and severe tumors in thoracic imaging. AI can assist in recognizing and classifying these nodules as benign or cancerous [116]. Machine learning focuses on pattern recognition. The traditional AI systems relied on preset engineering feature algorithms with specified parameters based on expert knowledge. Such features were intended to assess certain radiographic characteristics such as a tumor's 3D shape and intratumoral texture. Following that, a selection process ensures that only the most important features are used. The data is then fed into statistical machine learning models to find potential imaging-based biomarkers [117]. Deep learning algorithms learn by navigating the data space and by providing them with

greater problem-solving capabilities. Convolutional neural networks (CNNs) are the most used deep learning architectural typologies in medical imaging today, even though many deep learning designs have been researched to handle diverse objectives [118].

25. Critical Analysis

Medical imaging often contains several techniques that are noninvasive to make images of different parts of the body. New imaging techniques such as computed tomography (CT), positron emission computed tomography (PET), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), ultrasound (US), and digital mammography reveal the internal anatomy and physiology of the body [2].

Advanced imaging technologies are used to diagnose various external as well as internal human illnesses that can also minimize diagnostic errors and produce novel and better information about the target object. There are benefits and risks to every imaging technique. Ultrasound is also employed in the medical field to look at the kidneys, heart, liver, blood vessels, and other organs of the body [95]. Computed tomography (CT) measures cancer and various abnormalities in the heart, abdomen, bone, and spinal cord with high resolution [33]. 3D ultrasound computed tomography (3D USCT) is a promising technology for the investigation of breast cancer [48]. PET is a powerful technique to visualize, characterize, and quantify the biological processes and pathological changes at the cellular and subcellular levels within a living body. The development of MRI is now employed to examine several musculoskeletal, neurologic problems, and cancer. It can be used for both soft and hard tissues [1]. Digital mammography is a rapid and computer-based modality. It is used for the diagnosis and screening of breast cancer [89].

Some imaging techniques, such as CT, PET, SPECT, and digital mammography using X-rays, lead to high ionizing radiation exposure risk in some patients. There are some management steps for minimizing the radiation exposure risks from imaging techniques [114]. The development in medical imaging techniques such as the use of PET/CT hybrid, SPECT/CT hybrid, 3D USCT, and simultaneous PET/MRI leads to an increase in our understanding of diseases and their treatment [84]. With advanced medical imaging techniques, detection of early-stage diseases is possible and then eventually aids patients to live longer and better lives. In the future, with mounting innovations and advancements in technology systems, the medical diagnostic field would become a field of regular measurement of various complex diseases and will provide healthcare solutions [1].

Abbreviations

AI: Artificial intelligence
BMD: Bone mineral density
CAT: Computed axial tomography
CT: Computed tomography
DCE-MRI: Dynamic contrast-enhanced magnetic resonance imaging
DFI: Disseminated fungal infection
DXA: Dual-energy X-ray absorptiometry
ECG: Electrocardiography
EEG: Electroencephalography
EIT: Electrical impedance tomography
EPI: Echo-planar imaging
ESI: Electrical source imaging
FDG: 2-[F-18] Fluoro-2-deoxy-D-glucose
FLT: 3-Deoxy-3-[F-18] fluorothymidine
fMRI: Functional MRI
fUS: Functional ultrasound
IPF: Idiopathic pulmonary fibrosis
LOC: Lab-on-a-chip
MEG: Magnetoencephalography
MPTP: 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MRA: Magnetic resonance angiography
MRI: Magnetic resonance imaging
MRS: Magnetic resonance spectroscopy
NMR: Nuclear magnetic resonance
PEM: Positron emission mammography
PET: Positron emission tomography
PET-CT: PET and computed tomography
QCT: Quantitative CT
SIV: Simian immunodeficiency virus
SPECT: Single-photon emission computed tomography
SPION: Superparamagnetic iron oxide nanoparticles
TVS: Transvaginal ultrasound
UBM: Ultrasound biomicroscopy
3D USCT: Three-dimensional ultrasound computed tomography.

Data Availability

This is a review article. All data are taken from published research papers and available online.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

All authors contributed equally.

Acknowledgments

The authors wish to thank Research Center College of Pharmacy at King Saud University, Riyadh, Saudi Arabia, for their financial support and for providing free access to digital library and laboratory.

References

[1] M. Laal, "Innovation process in medical imaging," *Procedia-Social and Behavioral Sciences*, vol. 81, pp. 60–64, 2013.

[2] H. Kasban, M. A. M. El-Bendary, and D. H. Salama, "A comparative study of medical imaging techniques," *International Journal of Information Science and Intelligent System*, vol. 4, no. 2, pp. 37–58, 2015.

[3] National Research Council, *Mathematics and Physics of Emerging Biomedical Imaging*, National Academies Press, 1996.

[4] M. A. Flower, *Webb's Physics of Medical Imaging*, CRC Press, 2012.

[5] C. A. Roobottom, G. Mitchell, and G. Morgan-Hughes, "Radiation-reduction strategies in cardiac computed tomographic angiography," *Clinical radiology*, vol. 65, no. 11, pp. 859–867, 2010.

[6] A. Kaur and M. Goyal, "ROI based image compression of medical images," *International Journal of Computer Science Trends and Technology.*, vol. 2, no. 5, pp. 2347–8578, 2014.

[7] S. J. McPhee, M. A. Papadakis, and M. W. Rabow, *Current Medical Diagnosis & Treatment 2010*, McGraw-Hill Medical, New York, 2010.

[8] D. Berger, "A brief history of medical diagnosis and the birth of the clinical laboratory. Part 1–ancient times through the 19th century," *MLO: Medical Laboratory Observer*, vol. 31, no. 7, pp. 28–30, 1999.

[9] P. Skinner, "The New Cambridge Medieval History. Vol. III: c.900-c.1024. Timothy Reuter," *The English Historical Review*, vol. 116, no. 465, pp. 137-138, 2001.

[10] R. Wittern-Sterzel, "Diagnosis: the doctor and the urine glass," *The Lancet*, vol. 354, p. SIV13, 1999.

[11] A. A. Diamandopoulos and P. C. Goudas, "The late Greco-Roman and Byzantine contribution towards the evolution of laboratory examinations of bodily excrement. Part 2: sputum, vomit, blood, sweat, autopsies," *Clinical Chemistry and Laboratory Medicine*, vol. 43, no. 1, pp. 90–96, 2005.

[12] F. Wallis, "Signs and senses: diagnosis and prognosis in early medieval pulse and urine texts," *Social History of Medicine*, vol. 13, no. 2, pp. 265–278, 2000.

[13] L. R. Mooney, "A middle English verse compendium of astrological medicine," *Medical History*, vol. 28, no. 4, pp. 406–419, 1984.

[14] W. G. Bradley, "History of medical imaging," *Proceedings of the American Philosophical Society*, vol. 152, no. 3, pp. 349–361, 2008.

[15] E. Ambrose, T. Gould, and D. Uttley, *Jamie Ambrse*, BMJ Publishing Group, 2006.

[16] S. R. Stock, "Trends in the micro-and nano computed tomography 2010-2012," in *Developments in X-Ray Tomography VIII*, vol. 8506, article 850602, International Society for Optics and Photonics, 2012.

[17] B. C. Hancock and M. P. Mullarney, "X-ray microtomography of solid dosage forms," *Pharmaceutical Technology*, vol. 29, no. 44, pp. 92–100, 2005.

[18] R. Smith-Bindman, J. Lipson, R. Marcus et al., "Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer," *Archives of internal medicine*, vol. 169, no. 22, pp. 2078–2086, 2009.

[19] A. B. de González, M. Mahesh, K. P. Kim et al., "Projected cancer risks from computed tomographic scans performed in the United States in 2007," *Archives of internal medicine*, vol. 169, no. 22, pp. 2071–2077, 2009.

[20] N. J. Hodson, J. E. Husband, and J. S. Mac Donald, "The role of computed tomography in the staging of bladder cancer," *Clinical Radiology*, vol. 30, no. 4, pp. 389–395, 1979.

[21] E. Levine, K. R. Lee, J. R. Neff, N. F. Maklad, R. G. Robinson, and D. F. Preston, "Comparison of computed tomography and other imaging modalities in the evaluation of musculoskeletal tumors," *Radiology*, vol. 131, no. 2, pp. 431–437, 1979.

[22] S. Naz, M. Zahoor, M. U. K. Sahibzada, R. Ullah, and A. S. Alqahtani, "COVID-19 and SARS-CoV-2: everything we know so far–a comprehensive review," *Open Chemistry*, vol. 19, no. 1, pp. 548–575, 2021.

[23] J. R. Muhm, L. R. Brown, J. K. Crowe, P. F. Sheedy, R. R. Hattery, and D. H. Stephens, "Comparison of whole lung tomography and computed tomography for detecting pulmonary nodules," *American Journal of Roentgenology*, vol. 131, no. 6, pp. 981–984, 1978.

[24] R. Paxton and J. Ambrose, "The EMI scanner. A brief review of the first 650 patients," *The British journal of radiology*, vol. 47, no. 561, pp. 530–565, 1974.

[25] J. E. Husband, A. Barrett, and M. J. Peckham, "Evaluation of computed tomography in the management of testicular teratoma," *The British Journal of Radiology*, vol. 53, no. 2, pp. 179–183, 1981.

[26] G. H. Underwood Jr., R. G. Hooper, S. P. Axelbaum, and D. W. Goodwin, "Computed tomographic scanning of the thorax in the staging of bronchogenic carcinoma," *New England Journal of Medicine*, vol. 300, no. 14, pp. 777-778, 1979.

[27] D. F. Guthaner, W. R. Brody, M. Ricci, P. E. Oyer, and L. Wexler, "The use of computed tomography in the diagnosis of coronary artery bypass graft patency," *Cardiovascular and interventional radiology*, vol. 3, no. 1, pp. 3–8, 1980.

[28] H. Shimizu, M. Ida, S. Takayama et al., "The diagnostic accuracy of computed tomography in obstructive biliary disease: a comparative evaluation with direct cholangiography," *Radiology*, vol. 138, no. 2, pp. 411–416, 1981.

[29] J. Q. Knochel, P. R. Koehler, T. G. Lee, and D. M. Welch, "Diagnosis of abdominal abscesses with computed tomography, ultrasound and 111In leukocyte scans," *Radiology*, vol. 137, no. 2, pp. 425–432, 1980.

[30] R. H. Daffner, M. D. Halber, R. W. Postlethwait, M. Korobkin, and W. M. Thompson, "CT of the esophagus. II. Carcinoma," *American Journal of Roentgenology*, vol. 133, no. 6, pp. 1051–1055, 1979.

[31] A. K. Dixon, I. K. Fry, B. C. Morson, R. J. Nicholls, and A. Y. Mason, "Pre-operative computed tomography of carcinoma of the rectum," *The British journal of radiology*, vol. 54, no. 644, pp. 655–659, 1981.

[32] P. C. McAfee, H. A. Yuan, B. E. Fredrickson, and J. P. Lubicky, "The value of computed tomography in thoracolumbar fractures. An analysis of one hundred consecutive cases and a new classification," *The Journal of Bone and Joint Surgery. American*, vol. 65, no. 4, pp. 461–473, 1983.

[33] H. K. Genant, K. Engelke, and S. Prevrhal, "Advanced CT bone imaging in osteoporosis," *Rheumatology*, vol. 47, Supplement 4, p. iv9, 2008.

[34] A. Mastmeyer, K. Engelke, C. Fuchs, and W. A. Kalender, "A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine," *Medical image analysis*, vol. 10, no. 4, pp. 560–577, 2006.

[35] B. L. Riggs, L. J. Melton, R. A. Robb et al., "Population-based study of age and sex differences in bone volumetric density, size, geometry, and structure at different skeletal sites," *Journal of Bone and Mineral Research*, vol. 19, no. 12, pp. 1945–1954, 2004.

[36] M. Ito, K. Ikeda, M. Nishiguchi et al., "Multi-detector row CT imaging of vertebral microstructure for evaluation of fracture risk," *Journal of Bone and Mineral Research*, vol. 20, no. 10, pp. 1828–1836, 2005.

[37] W. Graeff and K. Engelke, "Microradiography and microtomography," in *Handbook on Synchrotron Radiation*, vol. 11, no. 4, 1991North-Holland, 1991.

[38] T. F. Lang, J. H. Keyak, M. W. Heitz et al., "Volumetric quantitative computed tomography of the proximal femur: precision and relation to bone strength," *Bone*, vol. 21, no. 1, pp. 101–108, 1997.

[39] M. E. Arlot, Y. Jiang, H. K. Genant et al., "Histomorphometric and μCT analysis of bone biopsies from postmenopausal osteoporotic women treated with strontium ranelate," *Journal of Bone and Mineral Research*, vol. 23, no. 2, pp. 215–222, 2008.

[40] C. I. Henschke, D. I. McCauley, D. F. Yankelevitz et al., "Early Lung Cancer Action Project: overall design and findings from baseline screening," *The Lancet*, vol. 354, no. 9173, pp. 99–105, 1999.

[41] U. Pastorino, M. Bellomi, C. Landoni et al., "Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results," *The Lancet*, vol. 362, no. 9384, pp. 593–597, 2003.

[42] H. Ohkubo, H. Nakagawa, and A. Niimi, "Computer-based quantitative computed tomography image analysis in idiopathic pulmonary fibrosis: a mini review," *Respiratory investigation*, vol. 56, no. 1, pp. 5–13, 2018.

[43] A. Asakura, T. Gotoh, T. Iwasawa et al., *Classification system of the CT images with nonspecific interstitial pneumonia*, Korean Society of Electronic Engineering Workshop, 2004.

[44] P. Sandercock, A. Molyneux, and C. Warlow, "Value of computed tomography in patients with stroke: Oxfordshire community stroke project," *British Medical Journal (Clinical Research Ed.)*, vol. 290, no. 6463, pp. 193–197, 1985.

[45] J. R. Bradshaw, L. Thomson, and M. J. Campbell, "Computed tomography in the investigation of dementia," *British Medical Journal (Clinical Research Ed.)*, vol. 286, no. 6361, pp. 277–280, 1983.

[46] P. M. Flynn, J. L. Shenep, R. Crawford, and W. T. Hughes, "Use of abdominal computed tomography for identifying disseminated fungal infection in pediatric cancer patients," *Clinical infectious diseases*, vol. 20, no. 4, pp. 964–970, 1995.

[47] K. E. Kocher, W. J. Meurer, R. Fazel, P. A. Scott, H. M. Krumholz, and B. K. Nallamothu, "National trends in use of computed tomography in the emergency department," *Annals of emergency medicine*, vol. 58, no. 5, pp. 452–462.e3, 2011.

[48] N. V. Ruiter, M. Zapf, T. Hopp et al., "3D ultrasound computer tomography of the breast: a new era?," *European Journal of Radiology*, vol. 81, pp. S133–S134, 2012.

[49] D. J. Brenner and E. J. Hall, "Computed tomography—an increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.

[50] J. M. Meulepas, C. M. Ronckers, A. M. Smets et al., "Leukemia and brain tumors among children after radiation exposure from CT scans: design and methodological opportunities of the Dutch Pediatric CT Study," *European journal of epidemiology*, vol. 29, no. 4, pp. 293–301, 2014.

[51] T. Worth, "Cancer threat from CT scans," *AJN The American Journal of Nursing*, vol. 110, no. 3, p. 18, 2010.

[52] A. S. Brody, D. P. Frush, W. Huda, R. L. Brent, and and the Section on Radiology, "Radiation risk to children from computed tomography," *Pediatrics*, vol. 120, no. 3, pp. 677–682, 2007.

[53] D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, *Positron Emission Tomography*, Springer, London, 2005.

[54] G. L. Brownell, T. F. Budinger, P. C. Lauterbur, and P. L. McGeer, "Positron tomography and nuclear magnetic resonance imaging," *Science*, vol. 215, no. 4533, pp. 619–626, 1982.

[55] A. Berger, "Positron emission tomography," *British Medical Journal*, vol. 326, no. 7404, p. 1449, 2003.

[56] M. E. Phelps, "Positron emission tomography provides molecular imaging of biological processes," *Proceedings of the National Academy of Sciences*, vol. 97, no. 16, pp. 9226–9233, 2000.

[57] A. Haroon, A. Zumla, and J. Bomanji, "Role of fluorine 18 fluorodeoxyglucose positron emission tomography–computed tomography in focal and generalized infectious and inflammatory disorders," *Clinical infectious diseases*, vol. 54, no. 9, pp. 1333–1341, 2012.

[58] U. Haberkorn, L. G. Strauss, C. H. Reisser et al., "Glucose uptake, perfusion, and cell proliferation in head and neck tumors: relation of positron emission tomography to flow cytometry," *Journal of Nuclear Medicine*, vol. 32, no. 8, pp. 1548–1555, 1991.

[59] H. Minn, H. Joensuu, A. Ahonen, and P. Klemi, "Florodeoxyglucose imaging: a method to assess the proliferative activity of human cancer in vivo. Comparison with DNA flow cytometry in head and neck tumors," *Cancer*, vol. 61, no. 9, pp. 1776–1781, 1988.

[60] B. A. Jabour, Y. Choi, C. K. Hoh et al., "Extracranial head and neck: PET imaging with 2-[F-18] fluoro-2-deoxy-D-glucose and MR imaging correlation," *Radiology*, vol. 186, no. 1, pp. 27–35, 1993.

[61] P. Price, T. Jones, and EC PET Oncology Concerted Action, and EORTC PET Study Group, "Can positron emission tomography (PET) be used to detect subclinical response to cancer therapy?," *European Journal of Cancer*, vol. 31, no. 12, pp. 1924–1927, 1995.

[62] P. T. Fox, M. E. Raichle, and W. T. Thach, "Functional mapping of the human cerebellum with positron emission tomography," *Proceedings of the National Academy of Sciences*, vol. 82, no. 21, pp. 7462–7466, 1985.

[63] A. L. Roskies, "Mapping memory with positron emission tomography," *Proceedings of the National Academy of Sciences*, vol. 91, no. 6, pp. 1989–1991, 1994.

[64] A. M. Scharko, S. B. Perlman, P. Hinds, J. M. Hanson, H. Uno, and D. Pauza, "Whole body positron emission tomography imaging of simian immunodeficiency virus-infected rhesus macaques," *Proceedings of the National Academy of Sciences*, vol. 93, no. 13, pp. 6425–6430, 1996.

[65] P. Mansfield and P. K. Grannell, "" Diffraction" and microscopy in solids and liquids by NMR," *Physical Review B*, vol. 12, no. 9, pp. 3618–3634, 1975.

[66] M. W. Grey and M. L. Carr, "Magnetic resonance imaging," *The American Journal of Neursing*, vol. 102, no. 12, pp. 26–33, 2002.

[67] J. Hinkle, "Magnetic resonance imaging," *The American Journal of Nursing*, vol. 99, no. 11, pp. 24CC–24DD, 1999.

[68] A. Prasad, "Making images/making bodies: visibilizing and disciplining through magnetic resonance imaging (MRI)," *Science, Technology, & Human Values*, vol. 30, no. 2, pp. 291–316, 2005.

[69] S. J. Eustace and E. Nelson, "Whole body magnetic resonance imaging," *British Medical Journal*, vol. 328, no. 7453, pp. 1387-1388, 2004.

[70] E. D. Plankey and J. Knauf, "What patients need to know about magnetic resonance imaging," *AJN The American Journal of Nursing*, vol. 90, no. 1, pp. 27-28, 1990.

[71] T. A. Egeland, T. G. Simonsen, J. V. Gaustad, K. Gulliksrud, C. Ellingsen, and E. K. Rofstad, "Dynamic contrast-enhanced magnetic resonance imaging of tumors: preclinical validation of parametric images," *Radiation research*, vol. 172, no. 3, pp. 339–347, 2009.

[72] S. N. Darty, M. S. Thomas, C. M. Neagle, K. M. Link, F. D. Wesley, and W. G. Hundley, "Cardiovascular magnetic resonance imaging," *The American Journal of Nursing*, vol. 102, no. 12, pp. 34–38, 2002.

[73] G. D. Honey, W. Soni, E. T. Bullmore et al., "Evidence of abnormal lateralisation of motor systems in schizophrenia using functional MRI," *Schizophrenia Research*, vol. 29, no. 1-2, pp. 70-71, 1998.

[74] M. Utz and J. Landers, "Magnetic resonance and microfluidics," *Science*, vol. 330, no. 6007, pp. 1056–1058, 2010.

[75] N. Seppa, *MRI detects missed breast cancers*, vol. 163, no. 25, 2003Science News, 2003.

[76] R. T. Branca, Z. I. Cleveland, B. Fubara et al., "Molecular MRI for sensitive and specific detection of lung metastases," *Proceedings of the National Academy of Sciences*, vol. 107, no. 8, pp. 3693–3697, 2010.

[77] Y. Wu, D. A. Chesler, M. J. Glimcher et al., "Multinuclear solid-state three-dimensional MRI of bone and synthetic calcium phosphates," *Proceedings of the National Academy of Sciences*, vol. 96, no. 4, pp. 1574–1578, 1999.

[78] H. Bridge and S. Clare, "High-resolution MRI: in vivo histology?," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1465, pp. 137–146, 2006.

[79] C. Catana, D. Procissi, Y. Wu et al., "Simultaneous in vivo positron emission tomography and magnetic resonance imaging," *Proceedings of the National Academy of Sciences*, vol. 105, no. 10, pp. 3705–3710, 2008.

[80] D. E. Kuhl and R. Q. Edwards, "Image separation radioisotope scanning," *Radiology*, vol. 80, pp. 653–661, 1963.

[81] F. H. Fahey and B. A. Harkness, "Single positron emission computed tomography," in *Pediatric Nuclear Medicine/PET*, pp. 457–483, Springer, New York, 2007.

[82] M. V. Green, J. Seidel, J. J. Vaquero, E. Jagoda, I. Lee, and W. C. Eckelman, "High resolution PET, SPECT and projection imaging in small animals," *Computerized medical imaging and graphics*, vol. 25, no. 2, pp. 79–86, 2001.

[83] A. Kugaya, M. Fujita, and R. B. Innis, "Applications of SPECT imaging of dopaminergic neurotransmission in neuropsychiatric disorders," *Annals of nuclear medicine*, vol. 14, no. 1, pp. 1–9, 2000.

[84] M. W. Huellner and K. Strobel, "Clinical applications of SPECT/CT in imaging the extremities," *European journal of nuclear medicine and molecular imaging*, vol. 41, no. S1, pp. 50–58, 2014.

[85] T. R. Stewart and J. L. Mumpower, "Detection and selection decisions in the practice of screening mammography," *Journal of Policy Analysis and Management*, vol. 23, no. 4, pp. 908–920, 2004.

[86] R. Hodgson, S. H. Heywang-Köbrunner, S. C. Harvey et al., "Systematic review of 3D mammography for breast cancer screening," *The Breast*, vol. 27, pp. 52–61, 2016.

[87] F. J. Gilbert, L. Tucker, and K. C. Young, "Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool," *Clinical radiology*, vol. 71, no. 2, pp. 141–150, 2016.

[88] R. A. Hubbard, W. Zhu, T. L. Onega et al., "Effects of digital mammography uptake on downstream breast-related care among older women," *Medical care*, vol. 50, no. 12, pp. 1053–1059, 2012.

[89] D. Spurgeon, "Digital mammography is more accurate only for certain groups of women," *BMJ: British Medical Journal*, vol. 331, no. 7518, p. 653.1, 2005.

[90] N. Seppa, *Sharpening the focus of mammograms*, vol. 168, no. 13, 2005Science News, 2005.

[91] A. N. Tosteson, N. Z. Stout, D. G. Fryback et al., "Cost-effectiveness of digital mammography breast cancer screening," *Annals of internal medicine*, vol. 148, no. 1, pp. 1–10, 2008.

[92] D. Saslow, C. Boetes, W. Burke et al., "American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography," *CA: a cancer journal for clinicians*, vol. 57, no. 2, pp. 75–89, 2007.

[93] S. Shapiro, E. A. Coleman, M. Broeders et al., "Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. International Breast Cancer Screening Network (IBSN) and the European Network of Pilot Projects for Breast Cancer Screening," *International journal of epidemiology*, vol. 27, no. 5, pp. 735–742, 1998.

[94] W. A. Berg, J. D. Blume, J. B. Cormack et al., "Combined screening with ultrasound and mammography vs. mammography alone in women at elevated risk of breast cancer," *Journal of the American Medical Association*, vol. 299, no. 18, pp. 2151–2163, 2008.

[95] V. Chan and A. Perlas, "Basics of ultrasound imaging," in *Atlas of Ultrasound-Guided Procedures in Interventional Pain Management*, pp. 13–19, Springer, New York, 2011.

[96] S. Barik, M. H. Jones, J. E. Spring et al., "Routine ultrasound scanning in pregnancy. The benefits are clinical," *British Medical Journal*, vol. 307, no. 6903, p. 559, 1993.

[97] E. Jauniaux, J. Johns, and G. J. Burton, "The role of ultrasound imaging in diagnosing and investigating early pregnancy failure," *Ultrasound in obstetrics & gynecology*, vol. 25, no. 6, pp. 613–624, 2005.

[98] E. Macé, G. Montaldo, I. Cohen, M. Baulac, M. Fink, and M. Tanter, "Functional ultrasound imaging of the brain," *Nature Methods*, vol. 8, no. 8, pp. 662–664, 2011.

[99] J. Z. Heckmatt, N. Pier, and V. Dubowitz, "Real-time ultrasound imaging of muscles," *Muscle & nerve*, vol. 11, no. 1, pp. 56–65, 1988.

[100] T. R. Nelson and D. H. Pretorius, "Three-dimensional ultrasound imaging," *Ultrasound in Medicine and Biology*, vol. 24, no. 9, pp. 1243–1270, 1998.

[101] A. Fenster and D. B. Downey, "Three-dimensional ultrasound imaging," *Annual review of biomedical engineering*, vol. 2, no. 1, pp. 457–475, 2000.

[102] R. H. Silverman, "High-resolution ultrasound imaging of the eye – a review," *Clinical & experimental ophthalmology*, vol. 37, no. 1, pp. 54–67, 2009.

[103] G. J. Czarnota, M. C. Kolios, J. Abraham et al., "Ultrasound imaging of apoptosis: high-resolution non-invasive monitoring of programmed cell death in vitro, in situ and in vivo," *British journal of cancer*, vol. 81, no. 3, pp. 520–527, 1999.

[104] T. Sherwood, "Renal masses and ultrasound," *British Medical Journal*, vol. 4, no. 5998, pp. 682-683, 1975.

[105] N. F. Woolacott, M. A. Puhan, J. Steurer, and J. Kleijnen, "Ultrasonography in screening for developmental dysplasia of the hip in newborns: systematic review," *BMJ*, vol. 330, no. 7505, p. 1413, 2005.

[106] S. Mitragotri, D. Blankschtein, and R. Langer, "Ultrasound-mediated transdermal protein delivery," *Science*, vol. 269, no. 5225, pp. 850–853, 1995.

[107] R. Gessner and P. A. Dayton, "Advances in molecular imaging with ultrasound," *Molecular imaging*, vol. 9, no. 3, pp. 7290.2010.00022–7290.2010.02010, 2010.

[108] N. Deshpande, A. Needles, and J. K. Willmann, "Molecular ultrasound imaging: current status and future directions," *Clinical radiology*, vol. 65, no. 7, pp. 567–581, 2010.

[109] R. M. Vojin Sulovic, M. Jr, D. S. Aleksandar, M. M. Danica, and L. J. Lana, "Hormonal effects of ultrasound," *Epidemiology*, vol. 1, no. 1, pp. 87-88, 1990.

[110] E. Boyd, U. Abdulla, I. Donald, J. E. E. Fleming, A. J. Hall, and M. A. Ferguson-Smith, "Chromosome breakage and ultrasound," *British Medical Journal*, vol. 2, no. 5760, pp. 501-502, 1971.

[111] J. Arehart-Treichel, *Fetal ultrasound: how safe?*, vol. 121, no. 24, 1982Science News, 1982.

[112] K. S. Suslick, "The chemical effects of ultrasound," *Scientific American*, vol. 260, no. 2, pp. 80–86, 1989.

[113] D. Marinac-Dabic, C. J. Krulewitch, and R. M. Moore Jr., "The safety of prenatal ultrasound exposure in human studies," *Epidemiology*, vol. 13, Supplement, pp. S19–S22, 2002.

[114] C. Adrien, M. L. Noriega, G. Bonniaud, J. M. Bordy, C. Le Loirec, and B. Poumarède, "Monte Carlo PENRADIO software for dose calculation in medical imaging," in *SNA+ MC 2013-Joint International Conference on Supercomputing in Nuclear Applications+ Monte Carlo*, p. 01601, EDP Sciences, 2014.

[115] H. E. Davies, C. G. Wathen, and F. V. Gleeson, "Risks of exposure to radiological imaging and how to minimize them," *BMJ*, vol. 342, no. 7797, pp. 589–593, 2011.

[116] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

[117] E. R. Ranschaert, S. Morozov, and P. R. Algra, *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks*, Springer, 1 edition, 2019.

[118] O. Oren, B. J. Gersh, and D. L. Bhatt, "Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints," *The Lancet Digital Health*, vol. 2, no. 9, pp. e486–e488, 2020.

Hindawi
BioMed Research International
Volume 2022, Article ID 1376659, 15 pages
https://doi.org/10.1155/2022/1376659

Research Article

Texture Analysis of Enhanced MRI and Pathological Slides Predicts EGFR Mutation Status in Breast Cancer

Tianming Du and Haidong Zhao

Department of Breast Surgery, Second Affiliated Hospital of Dalian Medical University, 467, Zhongshan Road, Shahekou District, Dalian, Liaoning 116023, China

Correspondence should be addressed to Haidong Zhao; z.hddl@hotmail.com

Received 5 April 2022; Revised 25 April 2022; Accepted 29 April 2022; Published 26 May 2022

Academic Editor: Chen Li

Copyright © 2022 Tianming Du and Haidong Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Image texture information was extracted from enhanced magnetic resonance imaging (MRI) and pathological hematoxylin and eosin- (HE-) stained images of female breast cancer patients. We established models individually, and then, we combine the two kinds of data to establish model. Through this method, we verified whether sufficient information could be obtained from enhanced MRI and pathological slides to assist in the determination of epidermal growth factor receptor (EGFR) mutation status in patients. *Methods.* We obtained enhanced MRI data from patients with breast cancer before treatment and selected diffusion-weighted imaging (DWI), T1 fast-spin echo (T1 FSE), and T2 fast-spin echo (T2 FSE) as the data sources for extracting texture information. Imaging physicians manually outlined the 3D regions of interest (ROIs) and extracted texture features according to the gray level cooccurrence matrix (GLCM) of the images. For the HE staining images of the patients, we adopted a specific normalization algorithm to simulate the images dyed with only hematoxylin or eosin and extracted textures. We extracted texture features to predict the expression of EGFR. After evaluating the predictive power of each model, the models from the two data sources were combined for remodeling. *Results.* For enhanced MRI data, the modeling of texture information of T1 FSE had a good predictive effect for EGFR mutation status. For pathological images, eosin-stained images can achieve a better prediction effect. We selected these two classifiers as the weak classifiers of the final model and obtained good results (training group: AUC, 0.983; 95% CI, 0.95-1.00; accuracy, 0.962; specificity, 0.936; and sensitivity, 0.979; test group: AUC, 0.983; 95% CI, 0.94-1.00; accuracy, 0.943; specificity, 1.00; and sensitivity, 0.905). *Conclusion.* The EGFR mutation status of patients with breast cancer can be well predicted based on enhanced MRI data and pathological data. This helps hospitals that do not test the EGFR mutation status of patients with breast cancer. The technology gives clinicians more information about breast cancer, which helps them make accurate diagnoses and select suitable treatments.

1. Introduction

Breast cancer is the most common cancer in women in both developed and developing countries. In 2020, there were 2.26 million new cases of breast cancer worldwide, accounting for 11.7% of all cases, and 685,000 deaths, accounting for 6.9% of all cases [1]. Many endogenous and exogenous factors have been identified as being associated with breast cancer etiology. Age is the strongest risk factor for the disease. More than two-thirds of new cases occur after the age of 55, with a higher risk of 4.0 for women older than 65 compared to those younger than 65. Several other risk factors have also been identified in relation to breast cancer. Some risk factors are constant, such as age and mutations in the BRCA1 and BRCA2 genes, which are estimated to account for 20% to 40% of inherited breast cancers with harmful mutations in BRCA1 and BRCA2 [2]. Family and reproductive history are also important risk factors. Other factors are dynamic, such as high endogenous estrogen, hormone therapy, obesity, and alcohol consumption [3, 4]. Among all kinds of diagnostic techniques, imaging techniques are the main diagnostic method used and can provide valuable data

for breast cancer patients. It has been shown that various imaging techniques can be used to diagnose and monitor patients with different stages of breast cancer [5]. In addition, there are a number of biochemical biomarkers available as new diagnostic and therapeutic tools for breast cancer patients [6]. Among them, epidermal growth factor receptor (EGFR) can be used as a new biomarker for monitoring the diagnosis and treatment of breast cancer patients [7].

EGFR is one of the four members of the HER family receptors, which are composed of EGFR, HER2, HER3, and HER4. EGFR signaling cascades are key regulators of cell proliferation, differentiation, division, survival, and cancer development. It is expressed in a variety of cancers, including breast, brain, lung, and prostate cancer. In breast cancer, EGFR is overexpressed in approximately half of triple-negative breast cancer (TNBC) and inflammatory breast cancer (IBC) patients [8]. EGFR is one of the first important targets identified for novel antitumor agents. High EGFR expression was identified as an independent predictor of poor outcome in TNBC [9]. This may be because the abnormal expression of the EGFR kinase domain may have an important impact on therapeutic resistance. Recent studies have shown that targeting EGFR enhances the chemical sensitivity of TNBC cells by reconnecting apoptotic signaling networks in TNBC. Some drugs that target EGFR also enhance the effectiveness of other therapies [10]. A study suggests that anti-EGFR-directed radioimmunotherapy combined with radiosensitizing chemotherapy and PARP inhibitor is more effective in treating triple-negative breast cancer [11]. These studies suggest that EGFR-targeted therapy may play a positive role in TNBC and IBC.

Hematoxylin and eosin (HE) staining has stood the test of time as a standard stain for the histological examination of human tissues [12]. In most hospitals, this technology is used as the primary source of pathological diagnosis. This simple dye combination can highlight the fine structure of cells and tissues. Most organelles and the extracellular matrix are eosinophilic, while the nucleus, rough endoplasmic reticulum, and ribosomes are basophilic. Much research has been done on this technology. Many different studies have improved the technique for different diseases. Different from traditional manual recognition, with the development of digital image processing technology, more research has focused on the digitalization and information extraction of pathological images. A study showed that the normalization of HE staining histological images by cycle-consistent generative adversarial networks could effectively enhance the training effect of the network [13]. Some experiments have also shown that HE staining images can be used to determine the Ki-67 score of breast cancer [14]. In this study, a normalization method was adopted to separate the staining effects of the two dyes, which allowed us to focus more on the medical information reflected by a particular dye [15].

Immunohistochemistry (IHC) is an inexpensive and effective technique that is easily available to most pathologists. It has been considerably used in several tumors. Immunohistochemical detection of EGFR is a well-established technique in non-small-cell lung cancer. IHC has good specificity and fairly good sensitivity to mutation-specific antibodies for common EGFR mutations. Moreover, IHC tests can accurately predict responses to EGFR-tyrosine kinase inhibitors (TKIs) [16]. This has been proven in both biopsies and cell blocks [17]. In breast cancer, this method is also widely used. Advances in clinical IHC have greatly improved the ability of clinicians to access information in terms of cost and efficiency. In oncology, these new techniques are highly reproducible, providing a useful alternative to and adjunct to molecular detection. IHC has become a valuable tool in modern cancer treatment.

Enhanced magnetic resonance imaging (MRI) is a highly sensitive breast imaging detection method. Compared with molybdenum target and ultrasound, enhanced MRI has higher resolution and can observe the tissue perfusion status. Different functional MRI sequences can be used to measure spatial differences in cell density, tissue structure, perfusion, and metabolism. With the improvement of medical treatment, enhanced MRI has become widely used as a routine examination for patients highly suspected of having breast cancer. Studies on MRI have shown that the texture features of MRI images can reflect a series of clinical information, such as molecular typing, prognosis, and gene mutations of the tumor itself. MRI-based radiomics analysis has been used to predict mutations in thoracic spinal metastases tumors in lung adenocarcinoma patients [18]. In gliomas, MRI imaging data and radiometric analysis based on these data were also used to analyze the EGFR mutation status of tumors [19, 20]. Another study showed that enhanced MRI data provided a better predictive model for head and neck squamous cell tumors [21]. The above research indicates that it is feasible to construct a prediction model for the texture analysis of image data by using enhanced MRI as the data source. In this experiment, diffusion-weighted imaging (DWI), T1 fast-spin echo (T1 FSE), and T2 fast-spin echo (T2 FSE) were selected to explore their prediction ability for EGFR status. DWI is the only noninvasive method to detect the diffusion of water molecules in living tissues. T1 and T2 sequences ensure that the image data can cover most of the tumor information.

Texture analysis can quantify complex medical image information. With the development of technology, many computer-aided diagnosis methods represented by texture analysis, radiomics, and deep learning play an important role in medical research [22–24]. Specifically, regional heterogeneity can be represented by texture features calculated using a variety of mathematical methods to assess the gray intensity of pixels, with 2D or higher texture features providing more complex information for tumor characterization than simple first-order histogram analysis. Such differences in features can quantify the heterogeneity of the tumor itself and can be used to predict the clinical information of the tumor. The texture analysis of the tumor can classify the tumor itself [25]. Compared with the widely used neural network in medical image field, texture analysis can directly reflect the correlation between parameters and labels. But this also means that this method is not suitable for the classification of complex parameters [26]. The technique could also be used to predict certain genetic mutations in tumors

and patient outcomes [27, 28]. This means that texture analysis can provide supplementary information and guidance throughout patient care. Among many texture information extraction methods, using gray level cooccurrence matrix is the most common method. This method has been applied in many tumor images [29]. In this study, we used texture analysis to model information images from multiple sources individually or jointly and explored methods that could predict tumor EGFR mutation status by comparing the models.

2. Materials and Methods

2.1. Study Participants. The radiology database of the Second Affiliated Hospital of Dalian Medical University was reviewed. We identified 187 patients who underwent MRI from March 2018 to June 2021. Fifty-one patients were excluded owing to incomplete image sequences. Moreover, 22 patients were excluded because of the absence of distinct EGFR results. Therefore, a total of 114 patients were included in this study. The inclusion criteria were as follows: (i) no breast diseases before imaging examination and (ii) did not receive treatment for breast cancer or may have artificially changed breast imaging. The patients underwent MRI examinations before surgery and were diagnosed as grades 3, 4, or 5 according to the Breast Imaging Reporting and Data System. Surgery or biopsy was performed within 1 week to confirm the primary breast cancer diagnosis and EGFR status. We excluded specific breast malignancies, such as IBC, Paget's disease, and breast cancer due to metastasis. Moreover, men and pregnant women were excluded. There was no specific information about the patients in the study, so the study did not involve ethical issues.

2.2. Enhanced MRI. We used the American GE 1.5 T Signa HDxt MRI scanner, and the receiving coil was a special one for the surface of the breast. In the prone position, the bilateral breasts naturally hung in the concave hole of the coil. The scan sequence and parameters were as follows: DWI, $b-\text{value} = 800\,\text{s/mm}$, repetition time $(\text{TR}) = 5{,}600$ ms, echo time $(\text{TE}) = 74.4\,\text{ms}$, matrix $= 130 \times 128$, field of view $(\text{FOV}) = 33\,\text{cm} \times 33\,\text{cm}$, and layer thickness $= 5\,\text{mm}$. All patients underwent dynamic contrast-enhanced MRI following DWI sequence scanning. Gadolinium diamine was used as the contrast agent. The injection volume was 0.2 mmol/kg, and the flow rate was 2 ml/s to 3 ml/s. Following injection, 20 ml of normal saline was used to flush the tube. We performed continuous nonstop scanning. T1-weighted image plain scanning was initially performed. Following gadolinium injection, we continuously scanned nine phases, with 47 s for each phase. A total of 10 phases were scanned. The scanning time was 7 minutes 6 s, and the turning angle was 15°. Other scanning parameters were as follows: TR = 5.1 ms, TE = 2.5 ms to 12 ms, matrix $= 320 \times 384$, FOV $= 30.2\,\text{cm} \times 30.2\,\text{cm}$, and layer thickness $= 5\,\text{mm}$. Two radiologists with >10 years of experience in breast imaging diagnosis independently interpreted the MRI results. After discussing the images, they reached a diagnostic consensus. All data were transferred to a GE workstation (Advantage Windows 4.5, General Electric, Madison, WI, USA) [30].

Table 1: Clinical characteristics of patients with breast cancer.

Characteristics	Training group ($n = 79$)	Test group ($n = 35$)
Age (years)	55.3	55.9
ER		
Positive	64	26
Negative	15	9
PR		
Positive	53	25
Negative	26	10
HER2		
Positive	24	7
Negative	55	28
Stage I	55	28
Stage II	17	5
Stage III	4	2
Stage IV	3	0
Lymphonodus		
Positive	25	11
Negative	54	24
EGFR		
Positive	48	21
Negative	31	14

ER: estrogen receptor; PR: progesterone receptor; HERK2: human epidermal growth factor receptor 2; EGFR: epidermal growth factor receptor.

2.3. HE Staining. Pathological specimens were obtained from the enrolled patients. No data were excluded owing to missing values or ambiguity. Each sample was fixed in a 10% buffered formalin solution. The fixed tissue was dehydrated and cleared in an automatic tissue processor and stained with HE. The samples were first dewaxed in xylene and alcohol. The samples were then stained. Hematoxylin was dyed for 5 minutes, and eosin was dyed for 3 minutes. Finally, the samples were soaked in alcohol and xylene for dehydration and transparency. The slides are fixed with synthetic resin.

2.4. IHC. All specimens were fixed with 4% neutral formaldehyde, embedded in paraffin, and continuously sectioned at a thickness of 4 μm. Following IHC staining, the specimens were observed and photographed under a microscope. We used the immunohistochemical SP method to detect the expression of susceptible genes. Specific steps were performed according to the standard instructions, and professional pathologists interpreted the films. We determined the comprehensive staining intensity and the percentage of positive cells. The final results were divided into the following categories: <5% visible staining: -; 5% to 25% visible staining: +; 26% to 50% visible staining: ++; and >50% visible staining: +++ (where - was defined negative and +, ++, and +++ were defined positive).

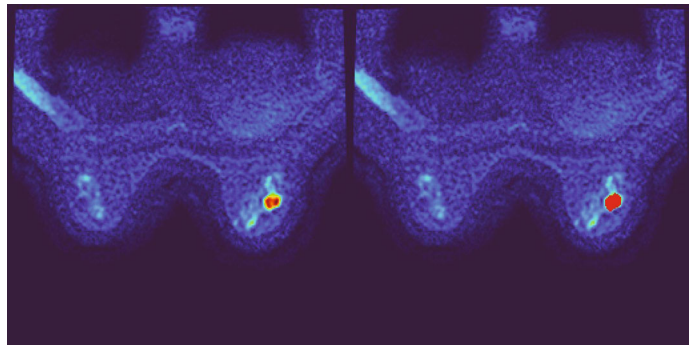


Figure 1: Example of DWI.

2.5. Texture Analysis. Texture analysis in this study was performed on two medical datasets. Therefore, the image modeling of a single information source not only explores whether it can reflect the expression of EGFR mutation status but also forms a weak classifier for the final modeling.

2.5.1. MRI Texture Analysis. The Pydicom library of Python was used to read the spatial location of the image. First, we used nearest neighbor interpolation to process the images so that they had equal pixel spacing, followed by image registration according to the spatial location. The region of interest (ROI) was manually obtained on a high-signal DWI sequence, and there was no necrosis or cystic components under ideal circumstances. In the case of no satisfactory image (usually because of low resolution), the ROIs were drawn by referencing T1 images. After standardizing the images, the Laplacian of Gaussian filter was used to process images. We used software to extract texture information, including Elongation, Flatness, LeastAxisLength, MajorAxisLength, Maximum2DDiameterColumn, Maximum2DDiameterRow, Maximum2DDiameterSlice, Maximum3DDiameter, MeshVolume, MinorAxisLength, Sphericity, SurfaceArea, SurfaceVolumeRatio, and VoxelVolume. The extraction of these parameters was based on gray level cooccurrence matrix (GLCM).

2.5.2. Pathology Image Texture Analysis. The scikit-image library of Python was used to process the pathology images. An image normalization algorithm was used to separate the dyeing effects of the two dyes and simulate two dyeing images. This method detects a color vector that meets a specific condition and converts it to a fixed value. On this basis, we extract several features of the image, including contrast, dissimilarity, homogeneity, energy, correlation, and ASM. The extraction of these parameters was also based on GLCM.

2.6. Modeling Method. For enhanced MRI images, we selected DWI, T1 FSE, and T2 FSE to extract texture features from the ROI. For pathological images, we extracted texture features from two simulated staining images. For the texture features obtained from each image, we selected features with a strong correlation with EGFR expression and adopted methods including adaboost, support vector machine (SVM), random forest, and decision tree for modeling. We selected the model with the best effect and compared it with the model from the same data source (MRI/HE). We selected the model with the best effect (with higher AUC value) as the weak classifier. In the case of similar classification effects, we chose the overfitting model as the weak classifier. The choice was made based on the advice of a software architect. Finally, we integrated the weak classifiers of the two data sources to judge EGFR mutation status.

3. Result

3.1. Clinical Features. The clinical information of the patients is shown in Table 1.

3.2. Enhanced MRI Texture Analysis

3.2.1. DWI. According to the ROI of the image (Figure 1), the gray level of the image was measured. According to this, a grayscale histogram was made (Figure 2).

A pseudocolor map of a 37-year-old patient with EGFR-positive mutation status (T1N1M0, ER: +, PR: +, HER-2: +, Ki67: -) is shown. Following imaging diagnosis, the breast tumor (considered malignancy) was confirmed by postoperative pathology. The images of T1 FSE and T2 FSE sequences and HE staining images mentioned below are also from this patient. The red part of the right picture is the ROI.

In the DWI sequence, we selected some features for modeling. As shown in Figure 3, these parameters had significant distribution differences according to the mutation status of EGFR in patients.

After comparing the modeling effects of various methods, we found that the adaboost algorithm had a relatively good classification effect on this problem (test group: AUC, 0.735; specificity, 0.9286; and sensitivity, 0.5238).

3.2.2. T1 FSE. According to the ROI of the image (Figure 4), the gray level of the image was measured. According to this, a grayscale histogram was made (Figure 5).

In the T1 FSE sequence, we selected some features for modeling. As shown in Figure 6, these parameters had significant distribution differences according to the mutation status of EGFR in patients.

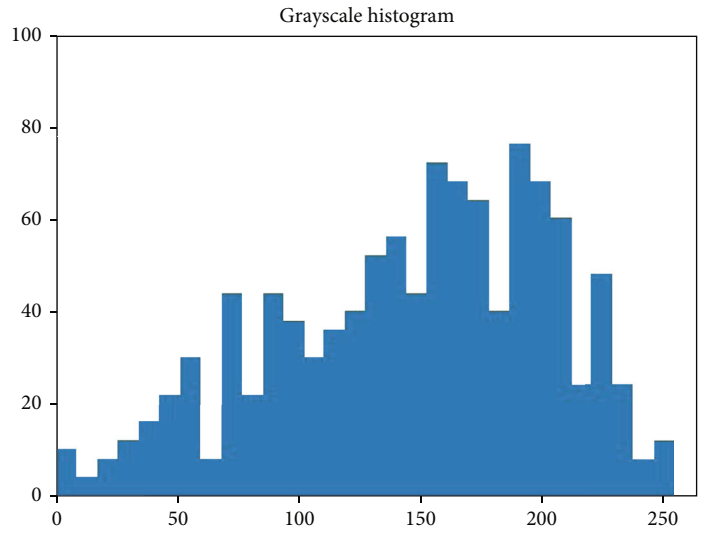


Figure 2: Grayscale histogram of pixels in the ROI of DWI.

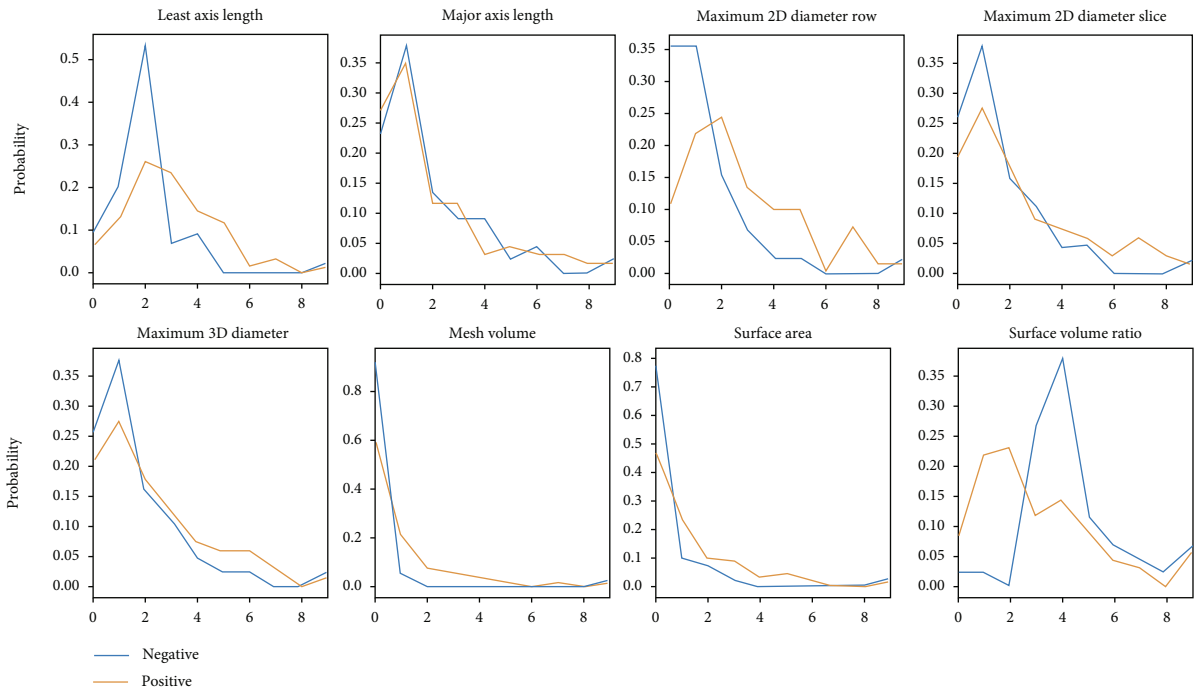


Figure 3: Feature distribution in DWI.

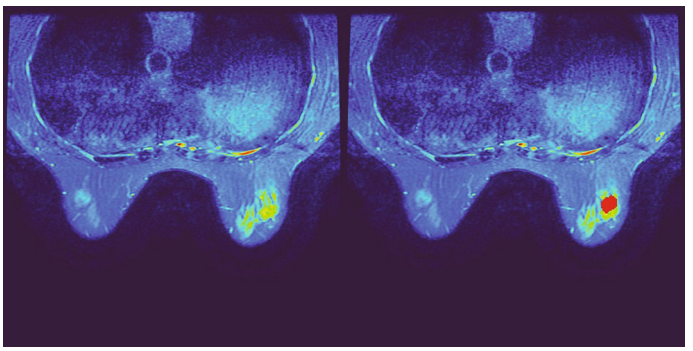


Figure 4: Pseudocolor maps of T1 FSE.

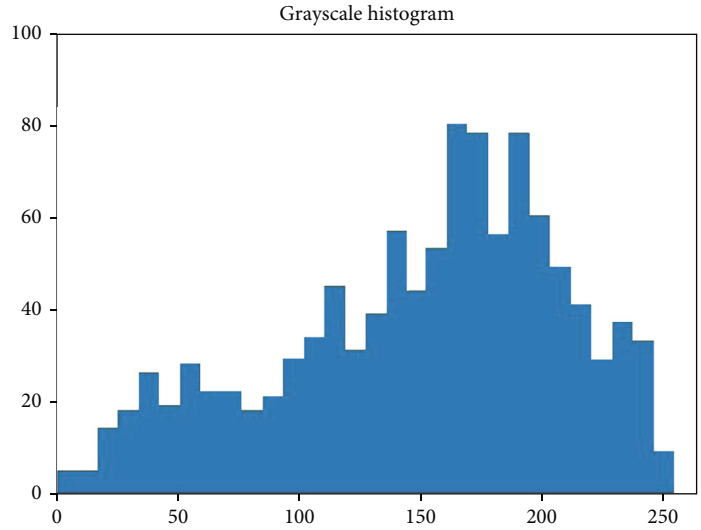


Figure 5: Grayscale histogram of pixels in the ROI of T1 FSE.

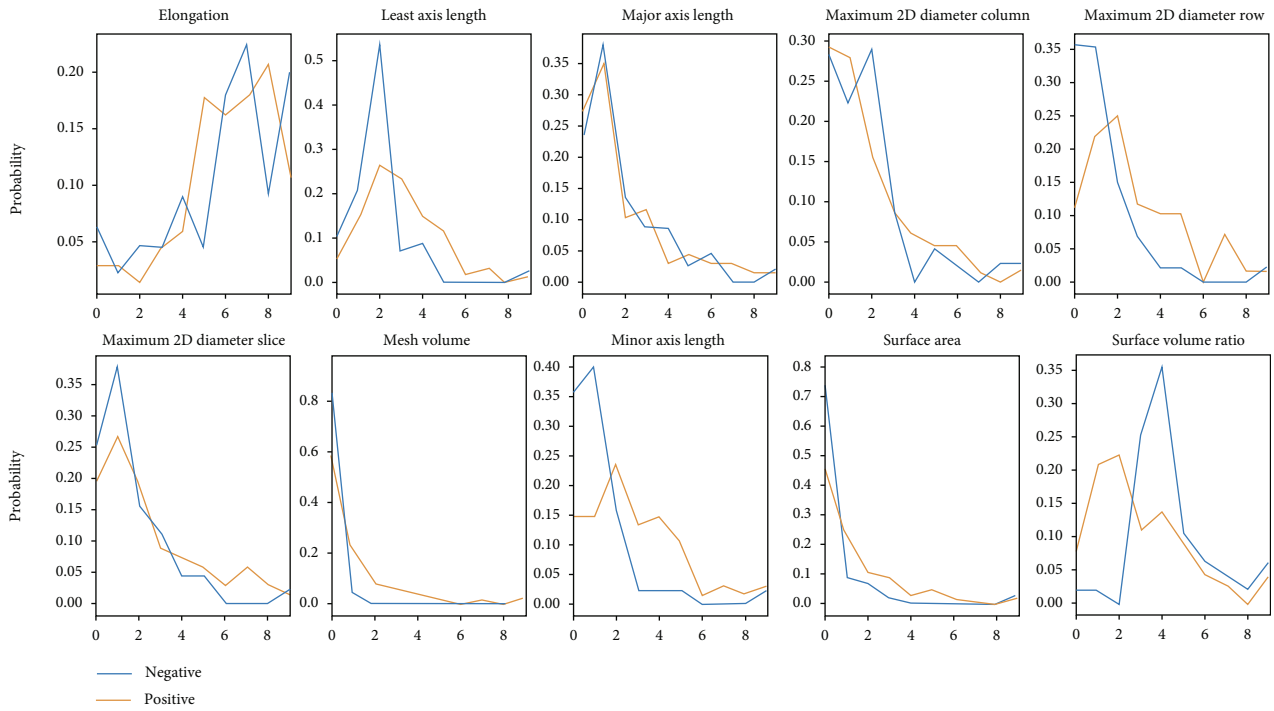


Figure 6: Feature distribution in T1 FSE.

After comparing the modeling effects of various methods, we found that the adaboost algorithm had a relatively good classification effect on this problem (test group: AUC,0.741; specificity, 0.8571; and sensitivity, 0.6667).

3.2.3. T2 FSE. According to the ROI of the image (Figure 7), the gray level of the image was measured. Based on this, a grayscale histogram was made (Figure 8).

In the T2 FSE sequence, we selected some features for modeling. As shown in the figure (Figure 9), these parameters have significant distribution differences according to the mutation status of EGFR in patients.

After comparing the modeling effects of various methods, we found that the decision tree algorithm had a relatively good classification effect on this problem (test group: AUC,0.726; specificity, 0.6429; and sensitivity, 0.8095).

By comparing the three models based on enhanced MRI, we conclude that the model based on T1 FSE is more suitable for weak classifiers (Figure 10).

3.3. Pathological Image. We used a special normalization algorithm to normalize the pathological images of patients. This method can separate the staining effect of two dyes (Figure 11).

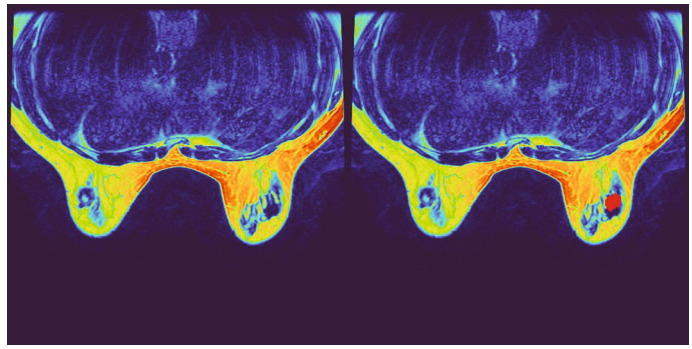


Figure 7: Pseudocolor maps of T2 FSE.

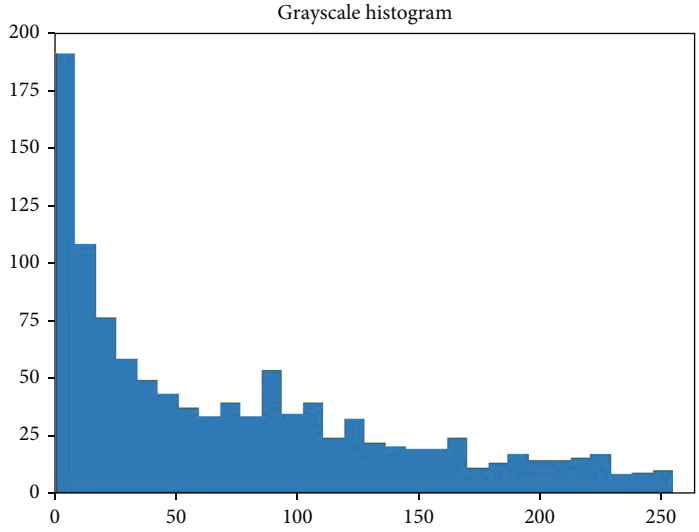


Figure 8: Grayscale histogram of pixels in the ROI of T2 FSE.

3.4. Hematoxylin. In the simulated hematoxylin staining image, we selected contrast, dissimilarity, homogeneity, energy, correlation, and ASM as the features. Finally, we chose the decision tree algorithm for modeling, which had a relatively good classification effect on this problem (test group:AUC,0.595; specificity, 0.4286; and sensitivity, 0.7619).

3.5. Eosin. In the simulated eosin staining images, we also selected contrast, dissimilarity, homogeneity, energy, correlation, and ASM as the features. We finally chose the random forest algorithm for modeling, which had a relatively good classification effect on this problem (test group; AUC,0.662; specificity, 0.7857; and sensitivity, 0.6667).

By comparing two models based on pathological images, we conclude that the model based on eosin images is more suitable for weak classifiers (Figure 12).

Finally, we selected two models from the MRI and pathological models (T1 FSE model and eosin model). As seen from the data distribution of the predicted values of the weak classifiers, EGFR mutation status can be clearly distinguished (Figure 13).

Finally, we adopted SVM to classify the data. In both the training set and the test set, the model achieved good results (Figure 14) (training group: AUC, 0.983; 95% CI, 0.95-1.00; accuracy, 0.962; specificity, 0.936; and sensitivity, 0.979; test group: AUC, 0.983; 95% CI, 0.94-1.00; accuracy, 0.943; specificity, 1.00; and sensitivity, 0.905).

4. Discussion

In this study, we developed an algorithm to predict EGFR expression in patients with untreated breast cancer based on enhanced MRI and pathologic images. This provides additional clinical information for physicians who do not perform EGFR testing. Given the growing importance of EGFR in the diagnosis and treatment of breast cancer, this approach is certainly valuable.

The EGFR gene is an oncogene-driven gene [31]. EGFR has several carcinogenic effects, including the stimulation of DNA synthesis, cell cycle, cell proliferation, cell metastasis, and invasion. Moreover, EGFR mutation was discovered to be the first molecular change in lung cancer. TKIs that target sensitizing mutations in the EGFR gene are a key pillar of the treatment of non-small-cell lung cancer [32]. To date, acquired resistance to EGFR-TKIs has been an inevitable process, usually occurring 9-12 months after treatment [33]. In breast cancer, there are also some studies describing the importance of EGFR mutation status in the diagnosis and treatment of breast cancer. Current evidence suggests

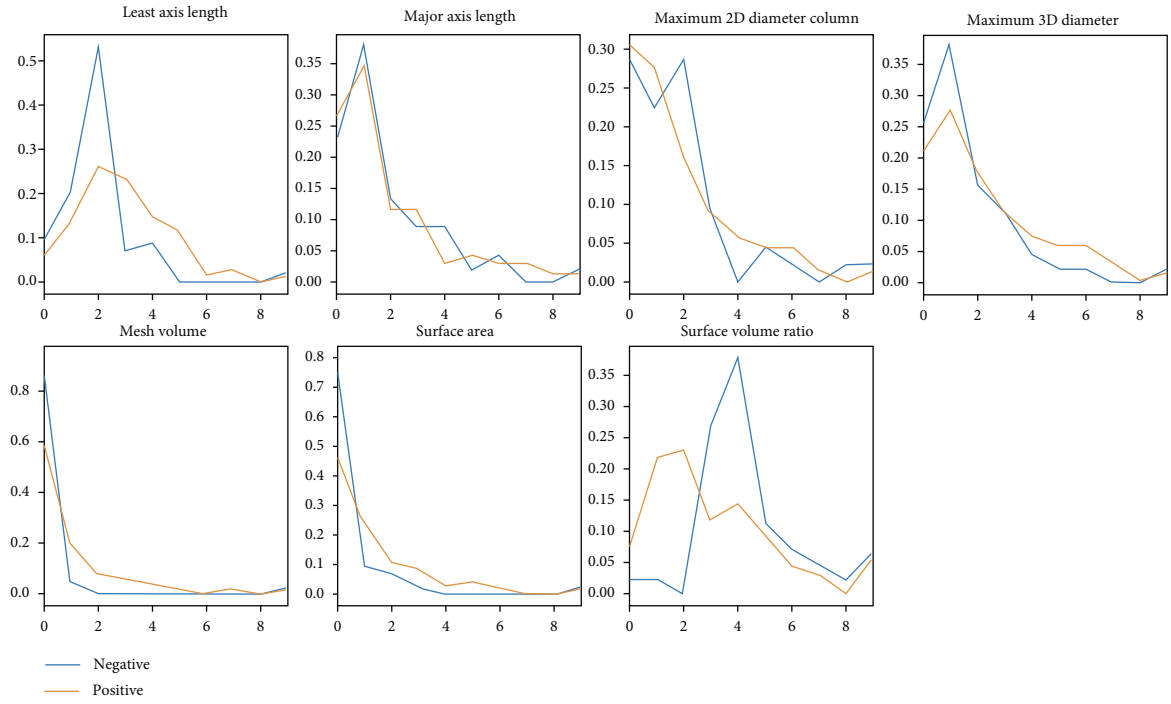


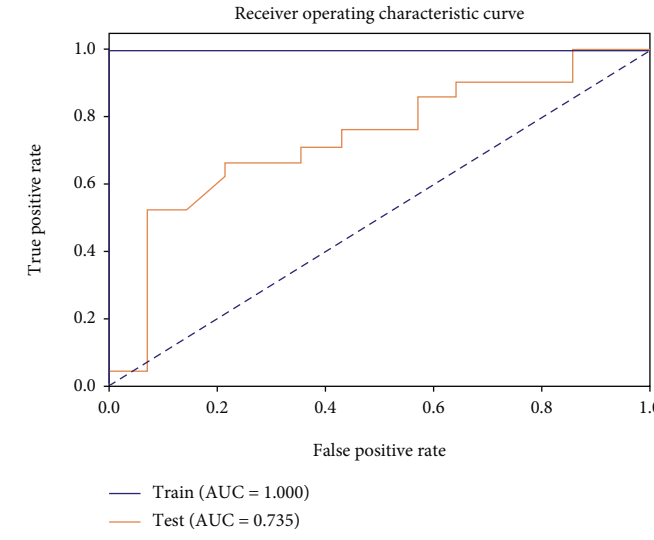
Figure 9: Feature distribution in T2 FSE.

an association between low baseline serum EGFR and shorter survival or reduced treatment responses in patients with advanced breast cancer. EGFR is usually overexpressed in metastatic breast cancer, but metastatic breast cancer is usually resistant to EGFR therapy. Anti-EGFR therapies such as cetuximab and erlotinib have had limited efficacy in clinical trials [34]. Therefore, more studies are needed to understand the underlying mechanistic link between EGFR expression and metastasis progression.

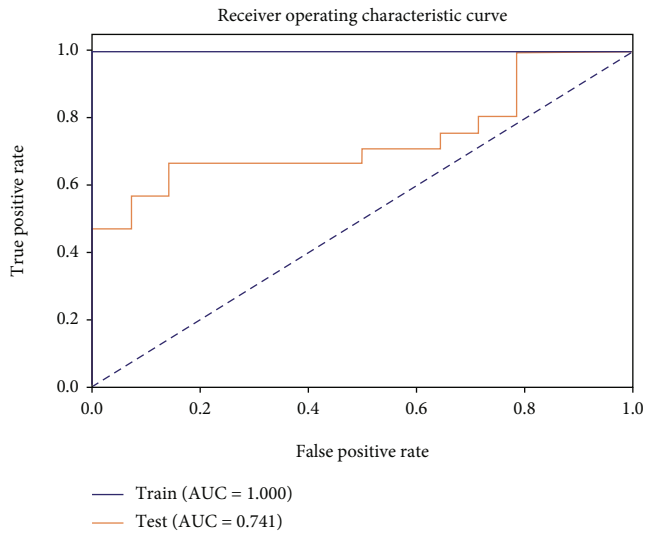
Recent studies on the texture analysis of EGFR mutation status mainly focused on lung cancer [35]. This is because EGFR-related targeted drugs have played an important role in the diagnosis and treatment of lung cancer [36]. Due to the particularity of lung tumor diagnosis, CT is often the most important imaging method for patients [37]. Therefore, the main image data sources for the texture analysis of EGFR expression are CT or PET/CT [38]. In the field of breast cancer, only a few studies have chosen to use MRI for analysis. Therefore, there is quite a gap in this area. However, in the above studies, due to the difficulty in obtaining medical records, it is rare to comprehensively consider pathological images and imaging data. In addition, many assays have been proposed to detect EGFR. A noninvasive test proved to have great potential [39]. In our experiment, we used medical information images from two different sources to predict EGFR expression. This multidimensional modeling method undoubtedly greatly improves the accuracy of the prediction results.

In this study, we adopted a specific method of pathological image normalization. This method helps us to exclude the influence of different dyeing conditions on the image coloring. At the same time, this normalization method separates the dyeing effect of the two dyes. The two dyes have different affinity for different tissues. This method gives us the ability to analyze the texture characteristics of specific tissues. Prior to this, we had conducted an imaging study on breast cancer [30]. The enhanced MRI data in both studies came from the same machine and included roughly the same patients. We can obviously see that in the aspect of imaging data analysis, compared with the simple texture analysis, the method of tumor image segmentation undoubtedly has higher accuracy.

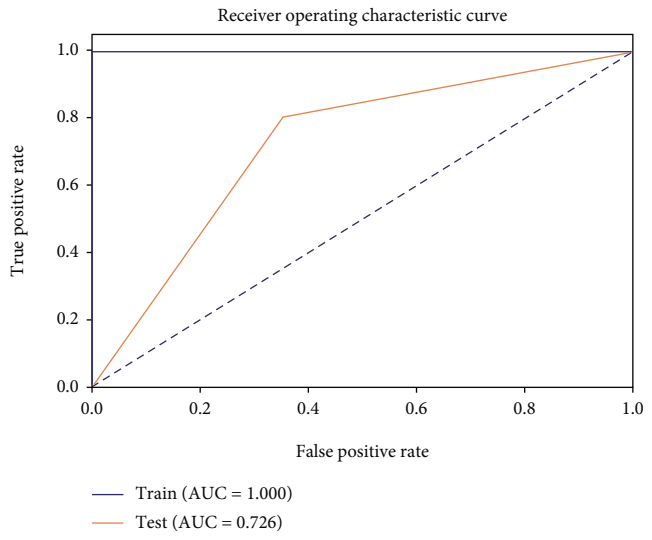
However, this experiment still has considerable limitations. First, in the process of making pathological sections, tumor tissues will shrink to different degrees after chemical process. This means that the spatial structure of the tumor has changed. Therefore, it is extremely difficult to complete the spatial correspondence between tumor images and image data. It is a very difficult task to try to analyze the tumor pathological sections and corresponding image layers together. This makes it impossible to compare our models across image sources. This difficulty forced us to separate models from different sources of clinical information, which undoubtedly reduced the efficiency of model screening. If images from different sources can correspond spatially, it means that we can give a higher information dimension to the space represented by each voxel. So, on the one hand, we can reduce the number of models. On the other hand, we were able to explore the effect of intratumor heterogeneity on EGFR status. We can even explore MRI's ability to distinguish tissue distribution by using artificial methods to distinguish tumor subareas and boundaries reflected in pathological images. On this basis, it is very potential to explore a high-precision tumor subregion segmentation method [40]. Second, it is well known that tumors are heterogeneous diseases, and the composition of each subregion



(a)

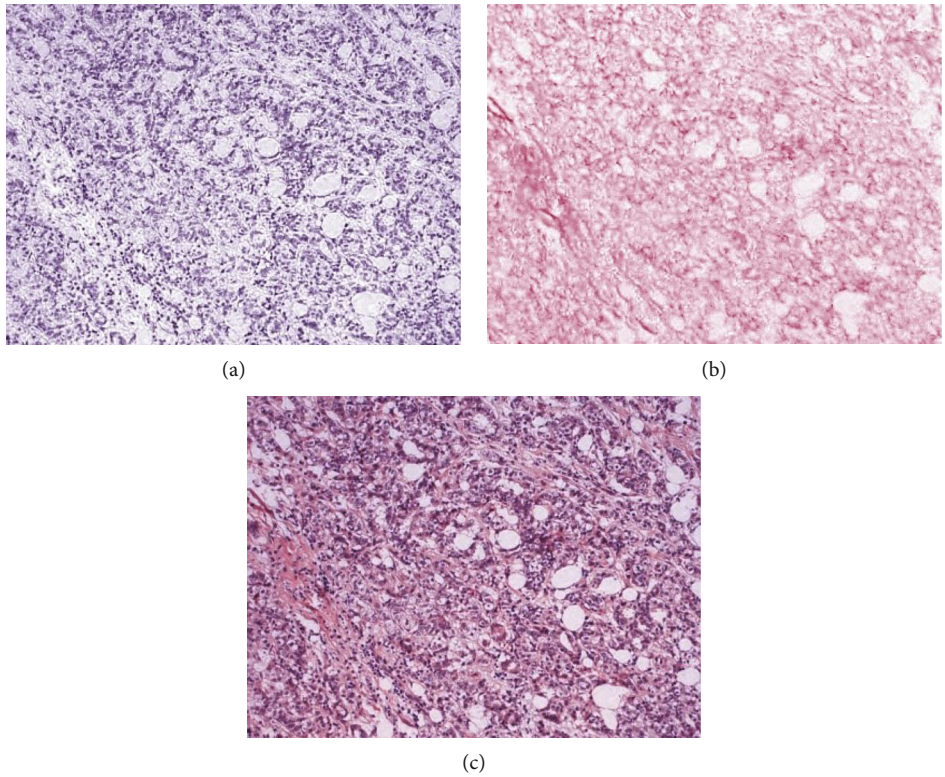


(b)



(c)

Figure 10: ROC curves of the three classification models ((a) DWI, (b)T1 FSE, and (c) T2 FSE).



(a) (b) (c)

Figure 11: Simulated staining images ((a) hematoxylin, (b) eosin, and (c) HE).

can well reflect the characteristics of tumors. This imaging method of analyzing the characteristics and relationships between tumor subareas is medically known as habitat imaging. However, due to the chemical process of pathological sections, the proportions of mesenchymal and parenchymal cells in each subregion are significantly different, and the contraction of each subregion is undoubtedly different after such treatment. This means that the spatial structure of the original subregion of the tumor has been artificially altered [41]. This change has a significant impact on habitat imaging that emphasizes relationships between subregions. At the same time, this method of research requires the acquisition of overall tumor image data. For pathological sections, it is undoubtedly very difficult to obtain such a large number of complete sections, including tumor edge conditions. Therefore, in this experiment, we only selected representative tumor sections as information sources for texture analysis. This choice also brings problems. Compared with the image data generated by artificial ROI, the overfitting of the model constructed by pathological images is greater. This means that the biopsy does not cover all the information about the tumor. With the development of computer-aided diagnosis (CAD) and image scanning technology, whole-slide image (WSI) scanners are widely used in the field of pathological diagnosis. Therefore, WSI analysis has become the key to modern digital histopathology. In this study, we used pathological images artificially generated by pathologists to ensure that the selected areas were dominated by tumor components. Due to objective conditions, this study did not use WSI as the information source for pathological images [42]. Modeling based on pathological staining images is significantly worse than modeling based on enhanced MRI data. The source of such pathological images may not adequately reflect the tumor itself. The total number of patients included in this experiment was limited due to the hospital scale and total number of patients. However, the final modeling results showed a better performance. Nevertheless, the strong overfitting characteristic of a weak classifier is undoubtedly the embodiment of insufficient data [43]. In addition, the feature extraction method adopted in this paper is relatively simple [44].

The imaging features used in this study are from GLCM formed by a single image processing method [45]. In fact, the features extracted by the texture analysis technique can be combined with a variety of filters to extract the texture information of a variety of matrices [46]. However, in our study, we found that the effect of a single model can be improved to some extent by simply adding the dimensions of texture parameters. Nevertheless, when we selected such a model as a weak classifier, the effect of the final model did not have more advantages, and its data distribution did not show more obvious differences. This finding is in accordance with the software architect's suggestion that the overfitted weak classifier will not have a significant effect on the final model.

The method adopted in this study is to model the texture features extracted from images. This usually entails converting the original image to a grayscale image for further

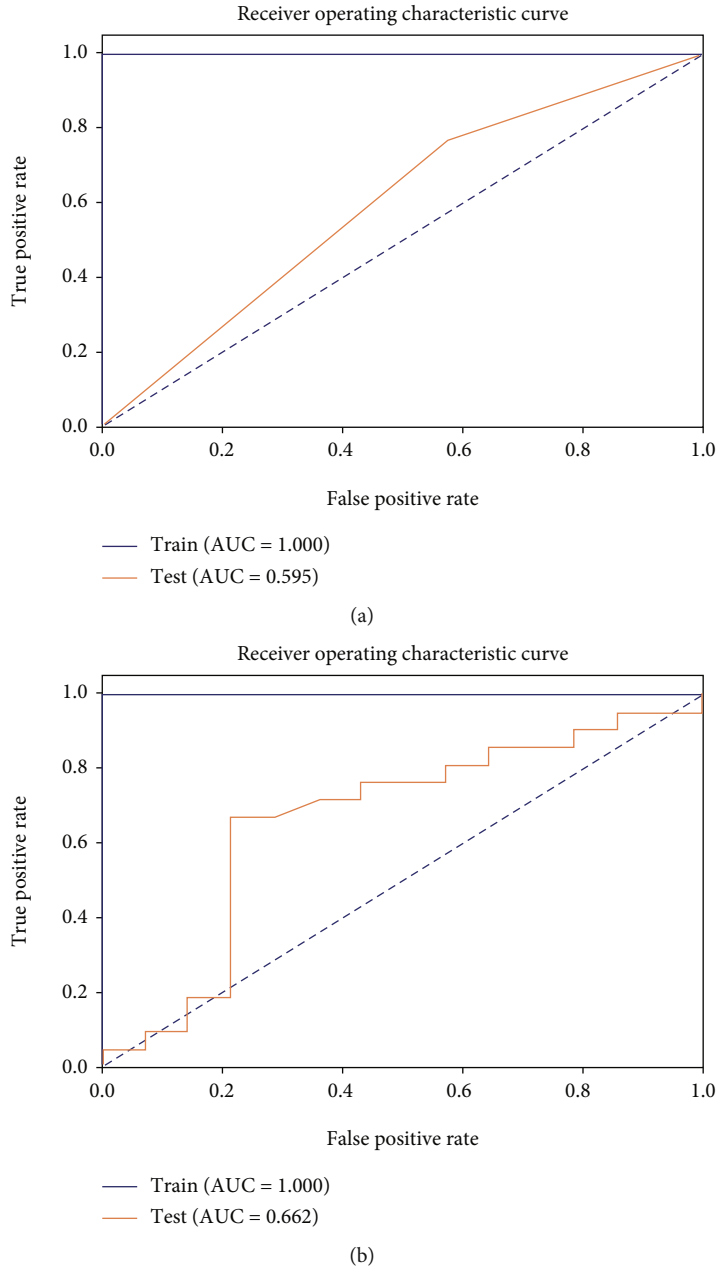


Figure 12: ROC curves of the two classification models ((a) hematoxylin and (b) eosin).

manipulation. However, some experiments show that RGB feature extraction technology has achieved good results in a variety of malignant tumors, including breast cancer and prostate cancer [47]. In recent years, artificial intelligence (AI) technology has developed rapidly. In particular, important achievements have been made in computer vision, image processing, and analysis. In pathology, there are many studies showing the use of multiple neural networks to classify and segment tumors [48, 49]. A study showed that convolutional neural networks have also made a breakthrough in the recognition of EGFR [50]. In fact, in the field of medicine, convolutional neural network has become a hot topic. From tumor image recognition to gene expression classification, convolutional neural network has made great contributions to medical research [51, 52]. In a preexperiment in this paper, we discussed the predictive power of using deep learning for EGFR mutation status. There is no doubt that deep learning can make more effective use of information in images. At present, the data sources of relevant studies mainly focus on using pathological images or radiological images alone. One feasible idea is to use Markov random field or conditional random field model to classify medical images directly [53]. Both methods have their own advantages in information processing [54, 55]. This method has made progress not only in the identification of EGFR status or breast cancer but also in gastrointestinal tumors, gliomas,

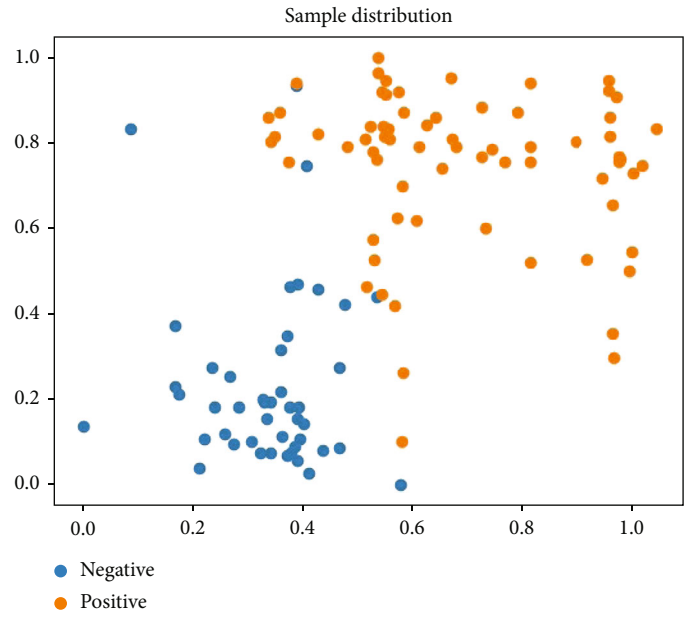


Figure 13: Scatter diagram of data distribution.

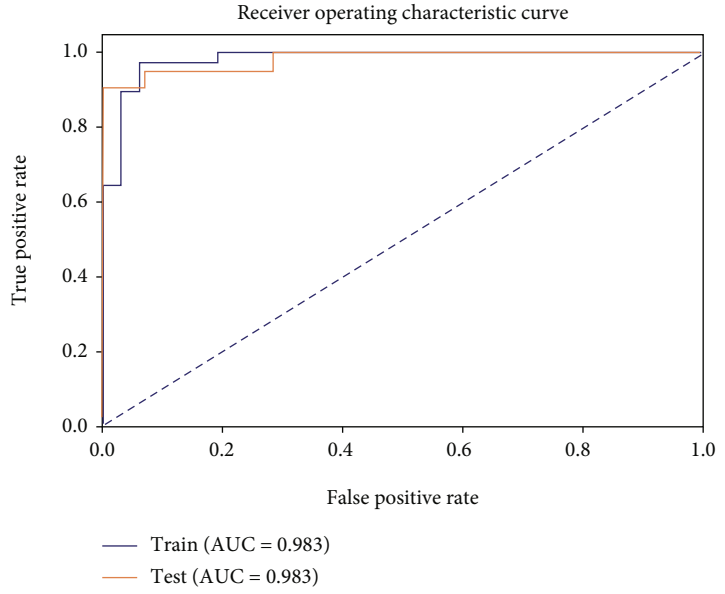


Figure 14: ROC curves of the final classification models.

reproductive system tumors, respiratory system tumors, etc. [56–58]. In both qualitative and quantitative medical problems, this approach has undisputed obvious advantages [59, 60]. Another idea is to use the image segmentation method to subsegment the tumor. These methods include *k*-means, *U*-Net, and other unsupervised or supervised algorithms [61, 62]. The final prediction model is established by describing the formed subregion. This method has made good progress in the field of imaging. All in all, we will try to develop a new comprehensive classification method in the future work. This method can make full use of WSI images and combine with imaging data to classify breast cancer. Neural networks will undoubtedly replace our current texture extraction methods.

5. Conclusions

In summary, a model was established based on features extracted from MRI and pathology images. The model can satisfactorily predict EGFR detected by immunohistochemistry. This study provides a reference for screening the high-risk population for surgery, drug therapy, and prognosis.

Data Availability

The MRI and clinic information data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors would like to thank software architect Yimin Gao for the excellent technical support and Doctor Zixin Tai for the data collection. This study was supported by the Dalian Medical University, Postgraduate Research Scholarship, 2219105127004.

References

[1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] J. D. Fackenthal and O. I. Olopade, "Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations," *Nature Reviews. Cancer*, vol. 7, no. 12, pp. 937–948, 2007.

[3] C Collaborative Group on Hormonal Factors in Breast, "Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies," *The Lancet Oncology*, vol. 13, no. 11, pp. 1141–1151, 2012.

[4] R. T. Fortner, V. Katzke, T. Kuhn, and Rudolf Kaaks, "Obesity and breast cancer," *Recent Results in Cancer Research*, vol. 208, pp. 43–65, 2016.

[5] W. C. Ou, D. Polat, and B. E. Dogan, "Deep learning in breast radiology: current progress and future directions," *European Radiology*, vol. 31, no. 7, pp. 4872–4885, 2021.

[6] M. Saleem, M. B. Ghazali, M. Wahab et al., "The BRCA1 and BRCA2 genes in early-onset breast cancer patients," *Advances in Experimental Medicine and Biology*, vol. 1292, pp. 1–12, 2020.

[7] A. Ogden, S. Bhattarai, B. Sahoo et al., "Combined HER3-EGFR score in triple-negative breast cancer provides prognostic and predictive significance superior to individual biomarkers," *Scientific Reports*, vol. 10, no. 1, p. 3009, 2020.

[8] H. Masuda, D. Zhang, C. Bartholomeusz, H. Doihara, G. N. Hortobagyi, and N. T. Ueno, "Role of epidermal growth factor receptor in breast cancer," *Breast Cancer Research and Treatment*, vol. 136, no. 2, pp. 331–345, 2012.

[9] H. S. Park, M. H. Jang, E. J. Kim et al., "High EGFR gene copy number predicts poor outcome in triple-negative breast cancer," *Modern Pathology*, vol. 27, no. 9, pp. 1212–1222, 2014.

[10] W. S. Liao, Y. Ho, Y. W. Lin et al., "Targeting EGFR of triple-negative breast cancer enhances the therapeutic efficacy of paclitaxel- and cetuximab-conjugated nanodiamond nanocomposite," *Acta Biomaterialia*, vol. 86, pp. 395–405, 2019.

[11] F. Al-Ejeh, W. Shi, M. Miranda et al., "Treatment of triple-negative breast cancer using anti-EGFR-directed radioimmunotherapy combined with radiosensitizing chemotherapy and PARP inhibitor," *Journal of Nuclear Medicine*, vol. 54, no. 6, pp. 913–921, 2013.

[12] Y. H. Fu, C. H. Hsu, J. D. Lin et al., "Using hematogram model to predict future metabolic syndrome in elderly: a 4-year longitudinal study," *The Aging Male*, vol. 18, no. 1, pp. 38–43, 2015.

[13] M. Runz, D. Rusche, S. Schmidt, M. R. Weihrauch, J. Hesser, and C. A. Weis, "Normalization of HE-stained histological images using cycle consistent generative adversarial networks," *Diagnostic Pathology*, vol. 16, pp. 1–10, 2021.

[14] M. Feng, Y. Deng, L. Yang et al., "Automated quantitative analysis of Ki-67 staining and HE images recognition and registration based on whole tissue sections in breast carcinoma," *Diagnostic Pathology*, vol. 15, no. 1, pp. 1–12, 2020.

[15] M. Macenko, M. Niethammer, J. S. Marron et al., "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE international symposium on biomedical imaging: from nano to macro*, pp. 1107–1110, IEEE, 2009.

[16] N. T. Hitij, I. Kern, A. Sadikov et al., "Immunohistochemistry for EGFR mutation detection in non-small-cell lung cancer," *Clinical Lung Cancer*, vol. 18, no. 3, pp. e187–e196, 2017.

[17] M. Ragazzi, I. Tamagnini, A. Bisagni et al., "Diamond: immunohistochemistry versus sequencing in EGFR analysis of lung adenocarcinomas," *Journal of Clinical Pathology*, vol. 69, no. 5, pp. 440–447, 2016.

[18] R. R. Sood, W. Shao, C. Kunder et al., "3D Registration of pre-surgical prostate MRI and histopathology images via super-resolution volume reconstruction," *Medical Image Analysis*, vol. 69, p. 101957, 2021.

[19] Y. W. Park, S. S. Ahn, C. J. Park et al., "Diffusion and perfusion MRI may predict EGFR amplification and the TERT promoter mutation status of IDH-wildtype lower-grade gliomas," *European Radiology*, vol. 30, no. 12, pp. 6475–6484, 2020.

[20] Y. Li, X. Liu, K. Xu et al., "MRI features can predict EGFR expression in lower grade gliomas: a voxel-based radiomic analysis," *European Radiology*, vol. 28, no. 1, pp. 356–362, 2018.

[21] H. J. Meyer, L. Leifels, G. Hamerla, A. K. Höhn, and A. Surov, "Associations between histogram analysis parameters derived from DCE-MRI and histopathological features including expression of EGFR, p16, VEGF, Hif1-alpha, and p53 in HNSCC," *Contrast Media & Molecular Imaging, vol.*, vol. 2019, pp. 1–10, 2019.

[22] C. C. Mirestean, O. Pague, C. Buzea, R. I. Iancu, and D. T. Iancu, "Radiomic machine learning and texture analysis-new horizons for head and neck oncology," *Maedica (Bucur)*, vol. 14, no. 2, pp. 126–130, 2019.

[23] M. E. Mayerhoefer, A. Materka, G. Langs et al., "Introduction to radiomics," *Journal of Nuclear Medicine*, vol. 61, no. 4, pp. 488–495, 2020.

[24] C. Li, H. Chen, L. Zhang et al., "Cervical histopathology image classification using multilayer hidden conditional random fields and weakly supervised learning," *IEEE Access*, vol. 7, pp. 90378–90397, 2019.

[25] K. Holli-Helenius, A. Salminen, I. Rinta-Kiikka et al., "MRI texture analysis in differentiating luminal A and luminal B breast cancer molecular subtypes - a feasibility study," *BMC Medical Imaging*, vol. 17, no. 1, p. 69, 2017.

[26] D. Xue, X. Zhou, C. Li et al., "An application of transfer learning and ensemble learning techniques for cervical

histopathology image classification," *IEEE Access*, vol. 8, pp. 104603–104618, 2020.

[27] G. L. Gierach, H. Li, J. T. Loud et al., "Relationships between computer-extracted mammographic texture pattern features and BRCA1/2mutation status: a cross-sectional study," *Breast Cancer Research*, vol. 16, no. 4, pp. 1–16, 2014.

[28] F. Giganti, S. Antunes, A. Salerno et al., "Gastric cancer: texture analysis from multidetector computed tomography as a potential preoperative prognostic biomarker," *European Radiology*, vol. 27, no. 5, pp. 1831–1839, 2017.

[29] S. Ai, C. Li, X. Li et al., "A state-of-the-art review for gastric histopathology image analysis approaches and future development," *BioMed Research International, vol.*, vol. 2021, pp. 1–19, 2021.

[30] T. Du and H. Zhao, "Habitat analysis of breast cancer-enhanced MRI reflects BRCA1 mutation determined by immunohistochemistry," *BioMed Research International*, vol. 2022, 2022.

[31] S. Sigismund, D. Avanzato, and L. Lanzetti, "Emerging functions of the EGFR in cancer," *Molecular Oncology*, vol. 12, no. 1, pp. 3–20, 2018.

[32] D. A. Sabbah, R. Hajjo, and K. Sweidan, "Review on epidermal growth factor receptor (EGFR) structure, signaling pathways, interactions, and recent updates of EGFR inhibitors," *Current Topics in Medicinal Chemistry*, vol. 20, no. 10, pp. 815–834, 2020.

[33] X. Liu, P. Wang, C. Zhang, and Z. Ma, "Epidermal growth factor receptor (EGFR): a rising star in the era of precision medicine of lung cancer," *Oncotarget*, vol. 8, no. 30, pp. 50209–50220, 2017.

[34] H. J. Lee, A. N. Seo, E. J. Kim et al., "Prognostic and predictive values of EGFR overexpression and EGFR copy number alteration in HER2-positive breast cancer," *British Journal of Cancer*, vol. 112, no. 1, pp. 103–111, 2015.

[35] G. M. Agazzi, M. Ravanelli, E. Roca et al., "CT texture analysis for prediction of EGFR mutational status and ALK rearrangement in patients with non-small cell lung cancer," *La Radiologia Medica*, vol. 126, no. 6, pp. 786–794, 2021.

[36] A. Leonetti, S. Sharma, R. Minari, P. Perego, E. Giovannetti, and M. Tiseo, "Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer," *British Journal of Cancer*, vol. 121, no. 9, pp. 725–737, 2019.

[37] M. Oudkerk, S. Liu, M. A. Heuvelmans, J. E. Walter, and J. K. Field, "Lung cancer LDCT screening and mortality reduction - evidence, pitfalls and future perspectives," *Nature Reviews. Clinical Oncology*, vol. 18, no. 3, pp. 135–151, 2021.

[38] X. Li, G. Yin, Y. Zhang et al., "Predictive power of a radiomic signature based on (18)F-FDG PET/CT images for EGFR mutational status in NSCLC," *Frontiers in Oncology*, vol. 9, p. 1062, 2019.

[39] W. Sun, X. Yuan, Y. Tian et al., "Non-invasive approaches to monitor EGFR-TKI treatment in non-small-cell lung cancer," *Journal of Hematology & Oncology*, vol. 8, no. 1, pp. 1–9, 2015.

[40] C. Sun, C. Li, H. Xu et al., "A comparison of segmentation methods in gastric histopathology images," *The Fourth International Symposium on Image Computing and Digital Medicine*, pp. 75–79, 2020.

[41] F. Alam and S. U. Rahman, "Challenges and solutions in multimodal medical image subregion detection and registration," *J Med Imaging Radiat Sci*, vol. 50, no. 1, pp. 24–30, 2019.

[42] X. Li, C. Li, M. M. Rahaman et al., "A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches," *Artificial Intelligence Review*, pp. 1–70, 2022.

[43] J. Demšar and B. Zupan, "Hands-on training about overfitting," *PLoS Computational Biology*, vol. 17, no. 3, p. e1008671, 2021.

[44] S. Ai, C. Li, X. Li et al., "A state-of-the-art review for gastric histopathology feature extraction methods," *The Fourth International Symposium on Image Computing and Digital Medicine*, pp. 64–68, 2020.

[45] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," *International Journal of Scientific and Research Publications*, vol. 3, no. 5, pp. 1–5, 2013.

[46] E. Trivizakis, G. S. Ioannidis, I. Souglakos, A. H. Karantanas, M. Tzardi, and K. Marias, "A neural pathomics framework for classifying colorectal cancer histopathology images based on wavelet multi-scale texture analysis," *Scientific Reports*, vol. 11, no. 1, p. 15546, 2021.

[47] G. Litjens, B. E. Bejnordi, N. Timofeeva et al., "Automated detection of prostate cancer in digitized whole-slide images of H and E-stained biopsy specimens," in *Medical Imaging 2015: Digital Pathology*, pp. 64–69, SPIE, 2015.

[48] X. Zhou, C. Li, M. M. Rahaman et al., "A comprehensive review for breast histopathology image analysis using classical and deep neural networks," *IEEE Access*, vol. 8, pp. 90931–90956, 2020.

[49] W. Li, J. Li, K. V. Sarma et al., "Path R-CNN for prostate cancer diagnosis and Gleason grading of histological images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 945–954, 2019.

[50] J. Xiong, X. Li, L. Lu et al., "Implementation strategy of a CNN model affects the performance of CT assessment of EGFR mutation status in lung cancer patients," *IEEE Access*, vol. 7, pp. 64583–64591, 2019.

[51] H. Chen, C. Li, X. Li et al., "IL-MCAM: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach," *Computers in Biology and Medicine*, vol. 143, p. 105265, 2022.

[52] R. Pratama, J. J. Hwang, J. H. Lee, G. Song, and H. R. Park, "Authentication of differential gene expression in oral squamous cell carcinoma using machine learning applications," *BMC Oral Health*, vol. 21, no. 1, pp. 1–8, 2021.

[53] Y. Li, C. Li, X. Li et al., "A comprehensive review of Markov random field and conditional random field approaches in pathology image analysis," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 609–639, 2022.

[54] J. Tohka, I. D. Dinov, D. W. Shattuck, and A. W. Toga, "Brain MRI tissue classification based on local Markov random fields," *Magnetic Resonance Imaging*, vol. 28, no. 4, pp. 557–573, 2010.

[55] L. Li, Y. Xie, L. Cen, and Z. Zeng, "A novel cause analysis approach of grey reasoning Petri net based on matrix operations," *Applied Intelligence*, vol. 52, no. 1, pp. 1–18, 2022.

[56] A. Sohail, A. Khan, N. Wahab, A. Zameer, and S. Khan, "A multi-phase deep CNN based mitosis detection framework for breast cancer histopathological images," *Scientific Reports*, vol. 11, no. 1, p. 6215, 2021.

[57] C. Li, H. Chen, X. Li et al., "A review for cervical histopathology image analysis using machine vision approaches," *Artificial Intelligence Review*, vol. 53, no. 7, pp. 4821–4862, 2020.

[58] A. S. Alhasan, "Clinical applications of artificial intelligence, machine learning, and deep learning in the imaging of gliomas: a systematic review," *Cureus*, vol. 13, no. 11, p. e19580, 2021.

[59] Y. Li, X. Wu, C. Li et al., "Intelligent gastric histopathology image classification using hierarchical conditional random field based attention mechanism," *2021 13th International Conference on Machine Learning and Computing*, pp. 330–335, 2021.

[60] M. Avanzo, J. Stancanello, G. Pirrone, and G. Sartor, "Radiomics and deep learning in lung cancer," *Strahlentherapie und Onkologie*, vol. 196, no. 10, pp. 879–887, 2020.

[61] S. Baek, Y. He, B. G. Allen et al., "Deep segmentation networks predict survival of non-small cell lung cancer," *Scientific Reports*, vol. 9, no. 1, p. 17286, 2019.

[62] C. Sun, C. Li, J. Zhang et al., "Gastric histopathology image segmentation using a hierarchical conditional random field," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1535–1555, 2020.

Hindawi
BioMed Research International
Volume 2022, Article ID 5303651, 15 pages
https://doi.org/10.1155/2022/5303651

Research Article

APU-Net: An Attention Mechanism Parallel U-Net for Lung Tumor Segmentation

Tao Zhou[1,2], **YaLi Dong**[1,2], **HuiLing Lu**[3], **XiaoMin Zheng**[4], **Shi Qiu**[5] and **SenBao Hou**[1,2]

[1]*School of Computer Science and Engineering, North Minzu University, Yinchuan, Ningxia 750021, China*
[2]*The Key Laboratory of Images and Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China*
[3]*School of Science, Ningxia Medical University, Yinchuan, Ningxia 750004, China*
[4]*Research Institute for Reproductive Medicine and Genetic Diseases, Wuxi Maternity and Child Health Hospital, Jiangsu Wuxi, 214002, China*
[5]*Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shanxi 710119, China*

Correspondence should be addressed to HuiLing Lu; lu_huiling@163.com and XiaoMin Zheng; 972514523@qq.com

Received 17 February 2022; Accepted 9 April 2022; Published 9 May 2022

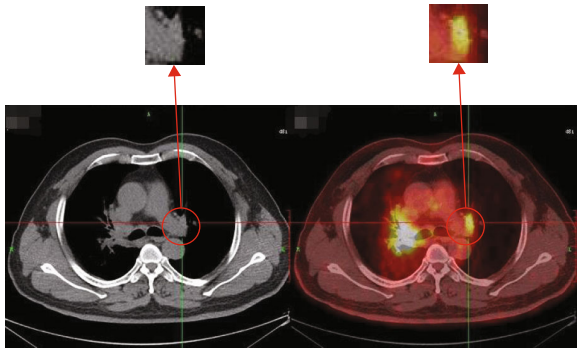
Academic Editor: Chen Li

Copyright © 2022 Tao Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lung cancer is one of the malignant tumors with high morbidity and mortality, and lung nodules are the early stages of lung cancer. The symptoms of pulmonary nodules are not obvious in the clinic, and the optimal treatment time is missed due to the missed diagnosis in the clinic. A parallel U-Net network called APU-Net is proposed. Firstly, two parallel U-Net networks are used to extract the features of different modalities. Among them, the subnetwork UNet_B extracts the CT image features, and the subnetwork UNet_A consists of two encoders to extract the PET/CT and PET image features. Secondly, multimodal feature extraction blocks are used to extract features for PET/CT and PET images in UNet_B network. Thirdly, a hybrid attention mechanism is added to the encoding paths of the UNet_A and UNet_B. Finally, a multiscale feature aggregation block is used for extracting feature maps of different scales of decoding path. On the lung tumor ^{18}FDGPET/CT multimodal medical images dataset, experiments' results show that the DSC, Recall, VOE, and RVD coefficients of APU-Net are 96.86%, 97.53%, 3.18%, and 3.29%, respectively. APU-Net can improve the segmentation accuracy of the adhesion between the lesion of complex shape and the normal tissue. This has positive significance for computer-aided diagnosis.

1. Introduction

Lung cancer has become one of the most common cancers, and it is a cancer with the highest fatality rate at present. When lung cancer patients are diagnosed, 70% patients has been already in the middle or advanced stage [1], the five-year survival rate for lung cancer patients is often less than 15% [2]. Hence, there are great significance for the early intervention and treatment of lung cancer patients. In the early stage, the main form of lung cancer is pulmonary nodules. When checked with Computed Tomography (CT), its imaging manifestations are mainly circular shadows with a diameter of no more than 30 mm [3]. With the improvement of medical image imaging technology, the scale of medical imaging data has been increased rapidly, which brings great challenges to the clinical work, and the workload of clinicians is greatly increased by judgment of large amount of medical imaging data. Due to the small area of pulmonary nodules in the whole medical image and the low contrast between the lesion and the background, it is easy to miss a diagnosis. Therefore, the computer-aided diagnosis system for automatic segmentation of pulmonary



(a) (b)

Figure 1: (a) Lung CT image. (b) Lung PET/CT image.

nodules [4–6] has become a research hotspot. There are outstanding achievements of deep learning methods in the field of image semantic segmentation, such as FCNs [7] and DeepLabV3 [8]. Hence, it can be widely used in medical image segmentation, of which the typical model is U-Net [9]. U-Net is an encoder and decoder structure, which can achieve semantic segmentation of the input images [10] and realize a good segmentation effect in small medical image datasets. Aiming at the problem that the segmentation performance of lung CT images is not high, Khanna et al. [11] added residual blocks to the U-Net network to improve the segmentation performance. Liu et al. [12] input two CT images of different scales into a residual blocks based on dual-path network, and the network encoder extracts the global and local features of the image from residual blocks and rich contextual information of pulmonary nodules. Aiming at the problem of the loss of spatial information caused by U-Net pooling operation, Gridach [13] proposed a pyramid expansion network, which integrates multiple dilated convolutions with different dilated rates to capture the tiny details of the image. Channel attention mechanism is added to skip connection of U-Net for lung parenchymal segmentation by Wang et al. [14], and the last layer of the network used a hybrid dilated attention convolution layer. Some researchers make full use of the target slices and its continuous slices to provide sequence information of lesions and improving the segmentation accuracy. Cao et al. [15] proposed a dual-path residual network for the 3D segmentation of pulmonary nodules. The target slice and two adjacent slices are input into an encoder with an improved residual structure, and a weighted sampling strategy is used for unbalanced training labels. Lee et al. [16] proposed the Mu-net network to be used for 3D image denoising, the downsampling is used to generate images of different scales in the input image, and different scales' images are input different scales' U-Net to extract image features of different scales. However, most of the current segmentation methods only use single-modal medical images, which ignores the complementarity of multimodal medical images to lesions. Therefore, in this paper, a parallel U-Net network is proposed for the lung lesion segmentation in combination with the different abilities of multimodal medical images to characterize lesions and uses a hybrid attention mechanism in this paper.

2. Methods

In this study, the paper proposed an attention mechanism parallel U-Net. It could be used for doctors to segment lung lesions and reduce missed diagnoses.

2.1. Dataset and Preprocessing. A total of 90 clinical patients with lung tumors are collected, confirmed by surgical pathology, including 32 females and 58 males, between January 2018 and June 2019 at the nuclear medicine department of a hospital in Ningxia. The patients ranged in age from 20 to 87 years, with a mean age of 61.2 ± 13.6 years, the patient's blood glucose is controlled to normal, patients are forbidden to eat for 6 hours and are injected with ^{18}F-FDG (^{18}F-Nuorodeoxyglucose) 0.11~0.13mci/kg (^{18}F-FDG is automatically synthesized by the HM-10 accelerator of Sumitomo, Japan, with a radiochemical purity greater than 95%), and the imaging is performed after approximately 60 minutes of lying down. Before imaging, drink 500 ml of water. To ensure the lesions are annotated correctly, the ground truth is annotated by radiologists with 30 years of experience. After data augmentation processing, such as rotation and mirror image, the final sample number of the three modal image datasets is 1026, respectively, including 909 PET/CT, CT, and PET images as the training set, and 117 PET/CT, CT, and PET images as the testing set. The image labels are manually drawn by clinicians. Figure 1 shows the contrast between PET/CT and CT images.

To solve the problem that pulmonary nodules occupy too few pixels of the whole image in the original image, ROI extraction based on Hough transform [17] is used to cut the original image to a pixel size of 50×50. And an image enhancement method based on exposure fusion [18] is used to improve the contrast between pulmonary nodules and background. The cross-entropy loss function is used for both subnetworks. Adam optimizer is used, training time is 150, the learning rate is initialized to 0.005, and batch size is 8. To prevent network overfitting, dropout is added to the network layer.

2.2. APU-Net. Two parallel U-Net (UNet_A and UNet_B) networks are used to extract multimodal medical image features of PET/CT (Positron Emission Tomography/Computed Tomography), CT (Computed Tomography), and PET (Positron Emission Tomography). UNet_A is composed of two encoders (encoder1_A and encoder2_A), which extracts medical image features of PET/CT and PET. UNet_B network extracts the features of CT images. Master network, which is composed of UNet_B, provides rich anatomical structure information for lesion segmentation. These features, that are extracted from the two subnetwork encoder paths, are concatenated and input into the decoder paths through the hybrid attention mechanism. Hence, a mutual image fusion method is realized about multimodal medical image features, to take full advantage of the complementarity of multimodal medical images. PET/CT and PET images are input into the UNet_A encoders, as shown in Figure 2. CT images are input into the UNet_B encoder, as shown in Figure 2. A hybrid attention mechanism is proposed in the skip connection of

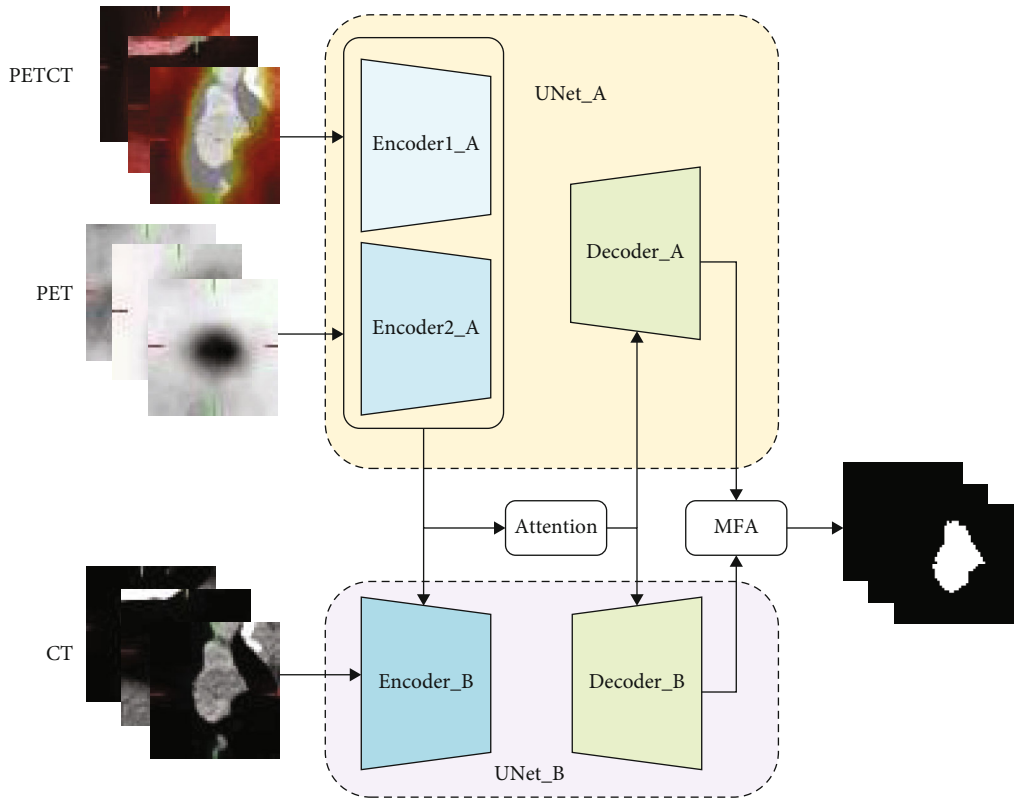


Figure 2: APU-Net network.

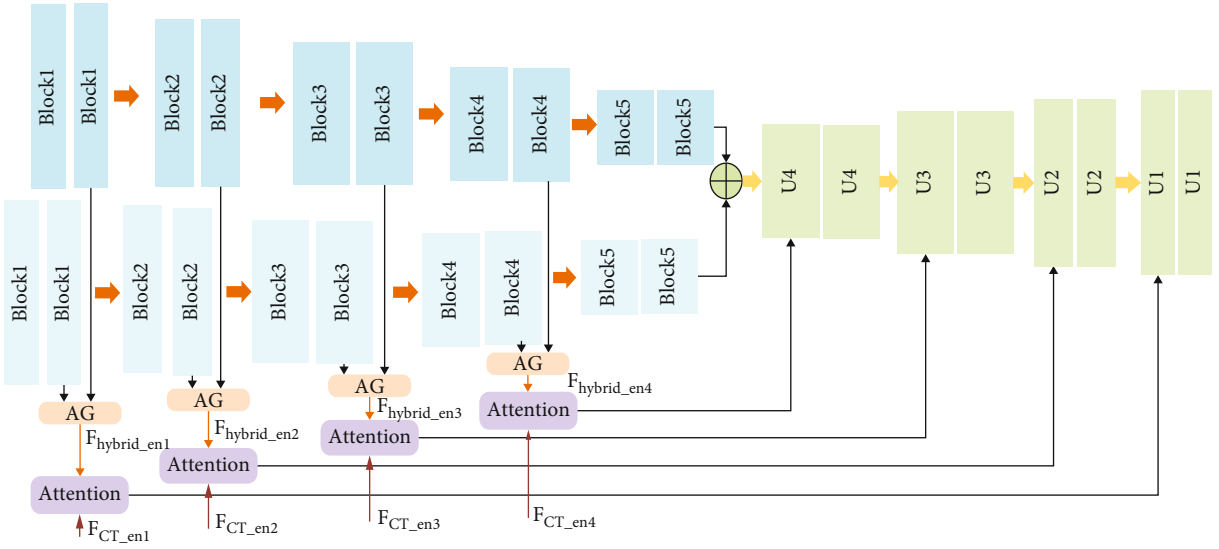


Figure 3: UNet_A network.

the network. Input features of UNet_A and UNet_B encoder paths are processed by the attention mechanism and input into decoder_A and decoder_B, respectively. Multiscale feature maps of two decoder paths are aggregated by multiscale feature aggregation blocks, and segmentation results are better than others.

The UNet_A consists of two encoders and one decoder. The UNet_A structure is shown in Figure 3. PET images and PET/CT images are input Encoder1_A and encoder2_A, respectively. The two encoders have the same parameters, and there are five layers. Each encoder layer of UNet_A includes two convolutional blocks, which consist of 3×3 convolutional blocks, Batch Normalization (BN), and ReLU activation function. To reduce the parameters of parallel U-Net, the number of convolutions at the first layer of the network is 16. With the deepening of the network, the number

Table 1: Network parameters of UNet_A.

Encoder path	Feature size	Kernel size	Decoder path	Feature size	Kernel size
Block1	50×50	$[3 \times 3, 16] \times 2$	U1	50×50	$[3 \times 3, 16] \times 2$
Block2	25×25	$[3 \times 3, 32] \times 2$	U2	25×25	$[3 \times 3, 32] \times 2$
Block3	12×12	$[3 \times 3, 64] \times 2$	U3	12×12	$[3 \times 3, 64] \times 2$
Block4	6×6	$[3 \times 3,128] \times 2$	U4	6×6	$[3 \times 3,128] \times 2$
Block5	3×3	$[3 \times 3,256] \times 2$			

Table 2: Network parameters of UNet_B.

Encoder path	Feature size	Kernel size	Decoder path	Feature size	Kernel size
Block1	50×50	$[3 \times 3, 32] \times 2$	U1	50×50	$[3 \times 3, 32] \times 2$
Block2	25×25	$[3 \times 3, 64] \times 2$	U2	25×25	$[3 \times 3, 64] \times 2$
Block3	12×12	$[3 \times 3,128] \times 2$	U3	12×12	$[3 \times 3,128] \times 2$
Block4	6×6	$[3 \times 3,256] \times 2$	U4	6×6	$[3 \times 3,256] \times 2$
Block5	3×3	$[3 \times 3,512] \times 2$			

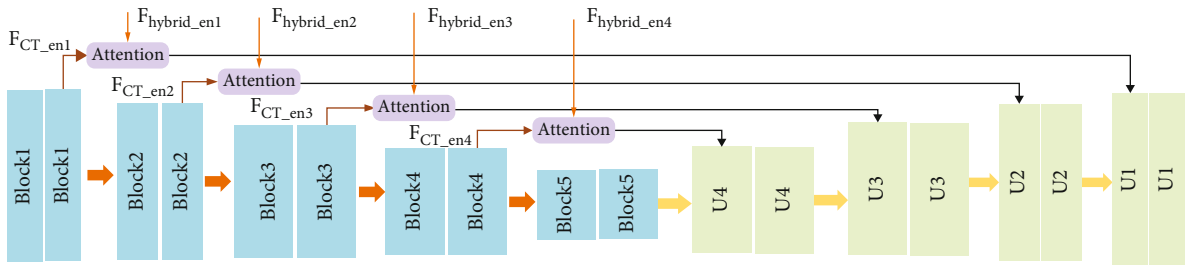


Figure 4: UNet_B network.

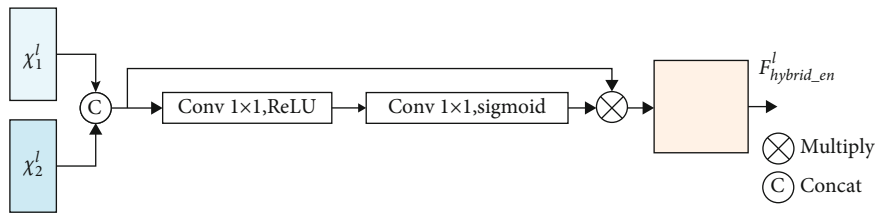


Figure 5: Two-modality medical image feature extraction block.

of convolution kernels at each layer increases by two times. The convolutional kernel number of the encoder1_A and encoder2_A is 16, 32, 64, 128, and 256, respectively. After each convolution layer, the downsampling operation is connected, and the downsampling operation is the 2×2 maximal pooling, and the image size is reduced to half comparing with the beforeimage when each downsampling. At the last layer of the UNet_A network, two-modal medical image features of extracting by the two encoders are summed, and the summed feature maps are input into the decoder_A. Decoder_A consists of four layers, and there are two convolution blocks in each layer, the convolution block is same as the encoder1_A. The convolution kernel number is 128 at the first upsampling, and the number of convolution kernels is reduced to half and the image size is doubled at each upsampling. The upsampling operation of the decoder_A is a 2×2 transpose convolution. A two-modal medical image feature extraction block is used for each layer of PET feature maps and PET/CT feature maps of the UNet_A encoder. This feature extraction block is used to extract relevant information by an attention gate from the PET image and PET/CT image, and these feature maps are transmitted to the hybrid attention mechanism of skip connection. The input feature maps of skip connection not only

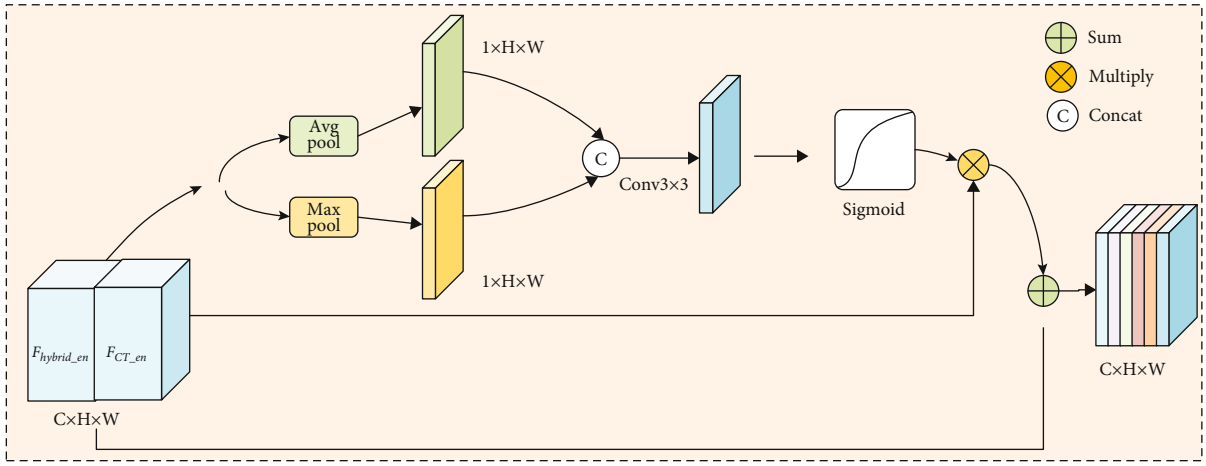


Figure 6: Spatial attention mechanism.

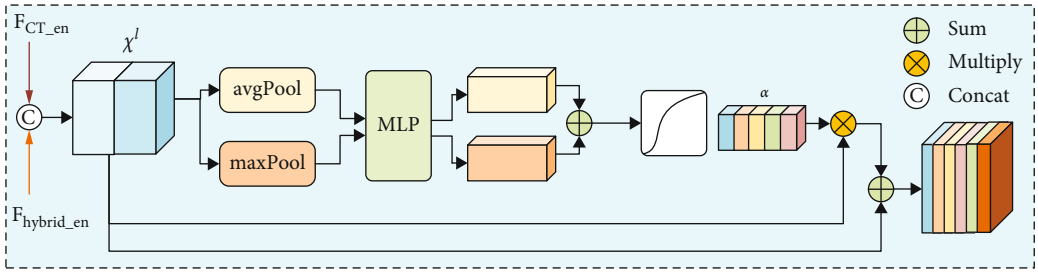


Figure 7: Channel attention mechanism.

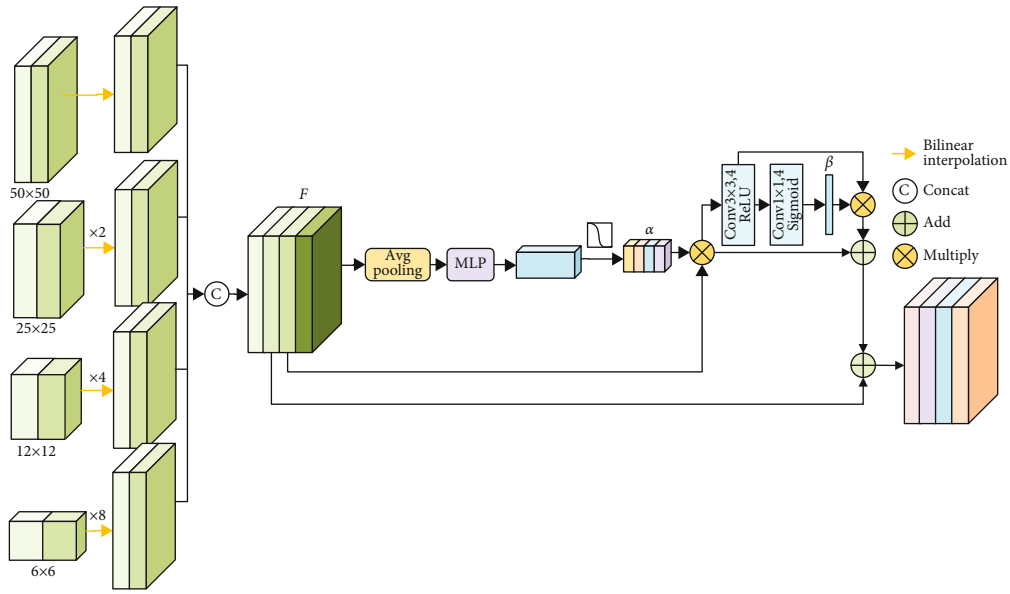


Figure 8: Multiscale feature aggregation block.

contain the feature maps F_{hybrid_en} of the UNet_A encoder but also receive the feature maps F_{CT_en} of UNet_B which makes the underlying features fuse with each other and realize feature reuse. The UNet_A parameters are shown in Table 1.

The UNet_B has five layers, and the numbers of convolutional kernels for encoder are 32, 64, 128, 256, and 512, respectively. Each layer of the network consists of two 3 × 3 convolutional blocks, Batch Normalization (BN), and activation function ReLU. Downsampling uses a 2 × 2 maximal

Table 3: Model segmentation index results.

Model	DSC (%)	Recall (%)	VOE (%)	RVD (%)
U-Net	95.16	94.99	2.53	2.74
MEU-Net	95.20	95.13	2.59	2.76
PU-Net	95.48	95.91	2.72	2.81

pooling, and the size of the feature maps is halved after downsampling. There are four layers in decoder_B, and convolutional kernel numbers of each layer are 256, 128, 64, and 32, respectively. Hybrid features F_{hybrid_en} and the UNet_B encoder feature maps F_{CT_en} are input into the decoding paths of two subnetworks through the attention mechanism. Table 2 describes the UNet_B network parameters, and the UNet_B structure is shown in Figure 4.

2.2.1. Two Modal Medical Image Feature Extraction Block. In the UNet_A, the two-modal medical image feature extraction block is used to extract complementary feature maps of PET and PET/CT. This block includes two input feature maps, namely, χ_1^l and χ_2^l, as shown in Figure 5.First, χ_1^l and χ_2^l are concatenated, then 1×1 convolution and ReLU are performed on the concatenated feature maps, and then, 1×1 convolution and sigmoid functions are performed to compress the weight of the feature maps to between 0 and 1. Finally, the output weights after sigmoid function are multiplied with the concatenated feature maps. The feature maps are inputted into the skip connection after multiplication, and the formula is expressed as equation (1).

$$F_{hybrid_en}^{l} = \left(\chi_1^l + \chi_2^l\right) \otimes \sigma\left(Conv_{3\times3}\left(Conv_{1\times1}\left(\chi_1^l + \chi_2^l\right)\right)\right), \tag{1}$$

where χ_1^l represents the layer l PET/CT feature maps extracted by UNet_A, χ_2^l represents layer l PET feature maps extracted by UNet_A, $Conv_{1\times1}(\cdot)$ represents 1×1 convolution operation and ReLU function, $Conv_{3\times3}(\cdot)$ represents 3×3 convolution operation, "σ" represents sigmoid function, and $F_{hybrid_en}^{l}$ represents the two-modality medical image feature representation of the l layer $F_{hybrid_en}^{l}$ extracted by UNet_A.

2.2.2. Hybrid Attention Mechanism. Hybrid attention mechanism [19] includes spatial attention mechanism and channel attention mechanism. The spatial attention mechanism focuses on the lesion in the feature map and suppresses irrelevant information such as the background information. Channel attention can assign larger weight coefficients to important channel feature maps, so it is necessary to process the attention mechanism of the three modal image features of the skip connection.

(1) Spatial Attention Mechanism. First, the UNet_A feature maps F_{hybrid_en} and UNet_B feature maps F_{CT_en} are concatenated and perform the maximal pooling and average pooling on the concatenated feature maps, respectively. The average pooling denoise the lung tumor image, and the maximal pooling highlights the lung tumor in the medical image. Then, the two pooled feature maps are concatenated, 3×3 convolution operation is performed on the concatenated feature maps, using sigmoid function to compress the feature coefficients range from 0 to 1, and the original feature maps are multiplied by the weight coefficients, as shown in Figure 6.

Specifically as shown in equation (2), where "σ" represents the sigmoid function, "+" represents the concatenating the feature maps. $F_{hybrid_en}^{l}$ is the two modal medical images feature maps of layer 1 extracted by UNet_A, and F_{CT}^{l} is the feature maps extracted by UNet_B. The two-modality medical image feature maps $F_{hybrid_en}^{l}$ and the CT feature maps $F_{CT_en}^{l}$ are concatenated, named as χ^l. "$\oplus$" represents the summed of the channels feature maps, "$\otimes$" represents the multiplication, and $SA(\cdot)$ represents the spatial attention mechanism operation.

$$SA^l(F) = \chi^l \oplus \left[\chi^l \otimes \left(\sigma\left(Conv_{3\times3}\left(AvgPool\left(\chi^l\right) + MaxPool\left(\chi^l\right)\right)\right)\right)\right]. \tag{2}$$

(2) Channel Attention Mechanism. First, to utilize more feature information, the average pooling and the maximal pooling are used to process the concatenate feature maps χ^l to obtain two $1 \times 1 \times C$ weight coefficients. Then, using the MultiLayer Perceptron (MLP) which composed of two fully connected layers and ReLU function to implement nonlinear transformation of features, the neuron numbers are 1/3 of the number of feature maps channel in the first fully connected layer, activation function is the ReLU function, and the neuron numbers are the feature map channels number in the second fully connected layer. Finally, the two results obtained by MLP are summed, and "α" weight coefficient is obtained through sigmoid function, and "α" weight coefficients and χ^l feature maps are multiplied and obtained the weighted feature maps, named as R. Then channel attention feature maps are obtained through sum weighted feature maps R and feature maps χ^l. These processing are as shown in Figure 7, and the formula is expressed as

$$CA^l(F) = \left(\chi^l \otimes \sigma\left(MLP\left(avgPool\left(\chi^l\right)\right) \oplus MLP\left(\max Pool\left(\chi^l\right)\right)\right)\right) \oplus \chi^l. \tag{3}$$

2.3. Multiscale Feature Aggregation Block. The network decoder contains low-level semantic feature information, and this information plays a vital role in lesion segmentation. Due to their different sizes in different scale feature maps, feature maps have different relevance in different scales to the object. Inspired by the reference [20], a multiscale feature aggregation (MFA) block is used to automatically determine the scale-wise weight for each pixel. The MFA block is illustrated in Figure 8. We concatenate each layer feature maps of decoder_A and decoder_B, and then, bilinear interpolation is used to resample the feature maps

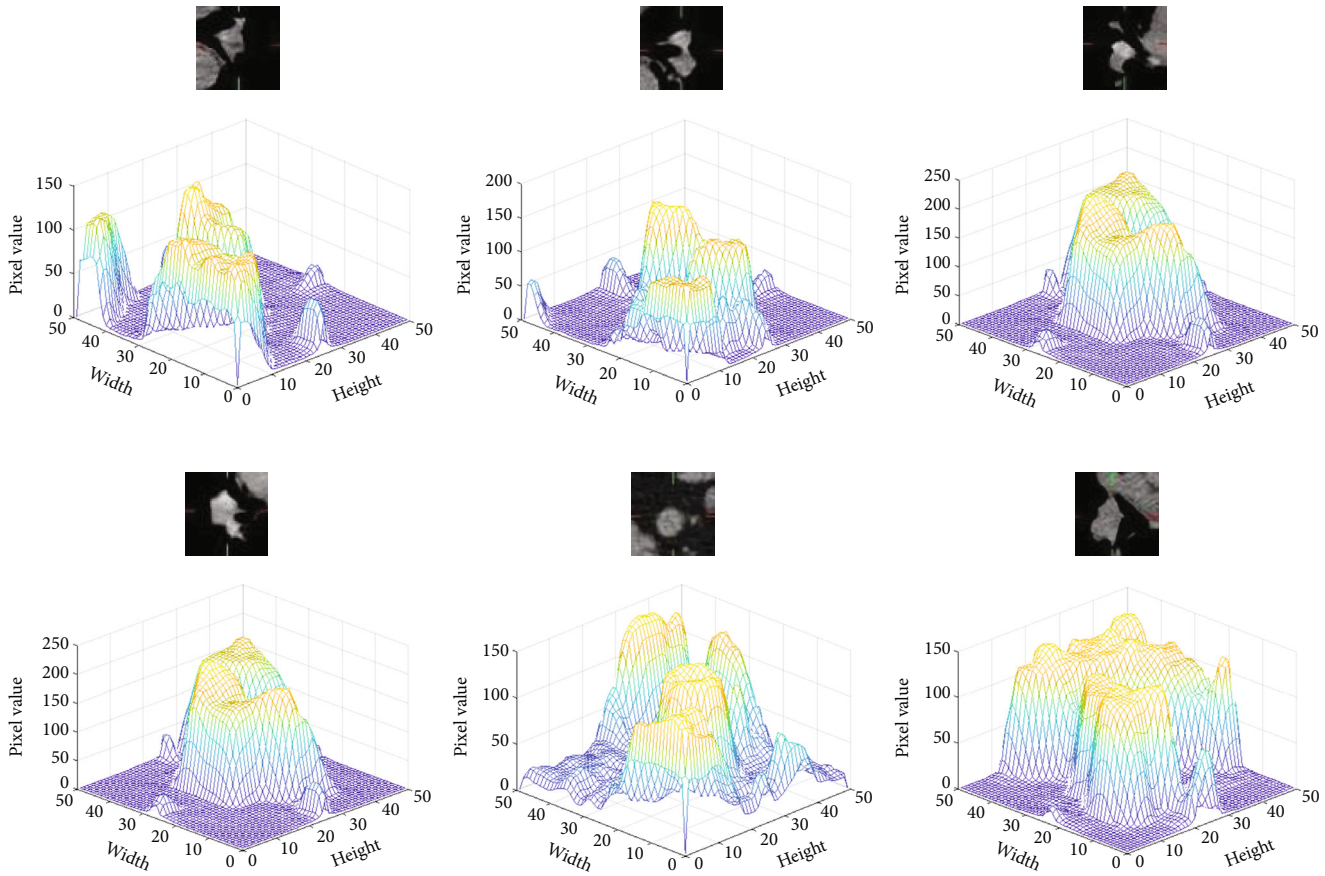


Figure 9: CT image three-dimensional gray value.

of different scales to a size of50pixel × 50pixel, and1 × 1 convolution operation is used to compress the four scale feature map channels into 16, and the feature maps were denoted as F. Avg-pooling and MLP are performed to obtain the channel coefficients, similar to the channel attention mechanism. The formula is expressed as equation (4). Where "σ" represent sigmoid function, and "α" represent channel coefficient.

$$\alpha = \sigma(MLP(avgPool(F))). \tag{4}$$

After the channel coefficients "α" are multiplied by the concatenated feature maps F, 3×3 convolution operation, ReLU function, 1×1 convolution operation, and sigmoid function are performed to obtain the coefficients "β", as shown in

$$\beta = \sigma(Conv_{1\times1}(\mathrm{Re}\,LU(Conv_{3\times3}(F \cdot \alpha)))). \tag{5}$$

Finally, the residual connection is used to connect the features. The specific process is shown in

$$F_{MFA} = F \cdot \alpha \cdot \beta + F \cdot \alpha + F. \tag{6}$$

3. Experimental Results

3.1. Implementation and Evaluation Models. All models are based on the PyTorch framework, cross entropy loss is used for training of UNet_A and UNet_B, Adaptive Moment Estimation (Adam) is used to train the model, and the value of the initial parameter is the following: initial learning rate is 0.005, batch size is 8, weight decay is 10^{-7}, iteration epochs are 150, and the model is implemented on one NVIDIA Geforce GTX 1080 Ti GPU.

In order to evaluate the network performance, the DSC, Recall, Volumetric Overlap Error (VOE), and Relative Volume Difference (RVD) are used. The specific formula are as follows. In this paper, the positive values of VOE and RVD are taken, the smaller the value of these two variables are the better. In order to be able to unify these four evaluation indicators, the two indicators of VOE and RVD use the method of calculating the absolute value with 1. The calculation methods of VOE and RVD indicators are as follows:

$$DSC = \frac{2 \times |P \cap G|}{|P| + |G|}, \tag{7}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN}, \tag{8}$$

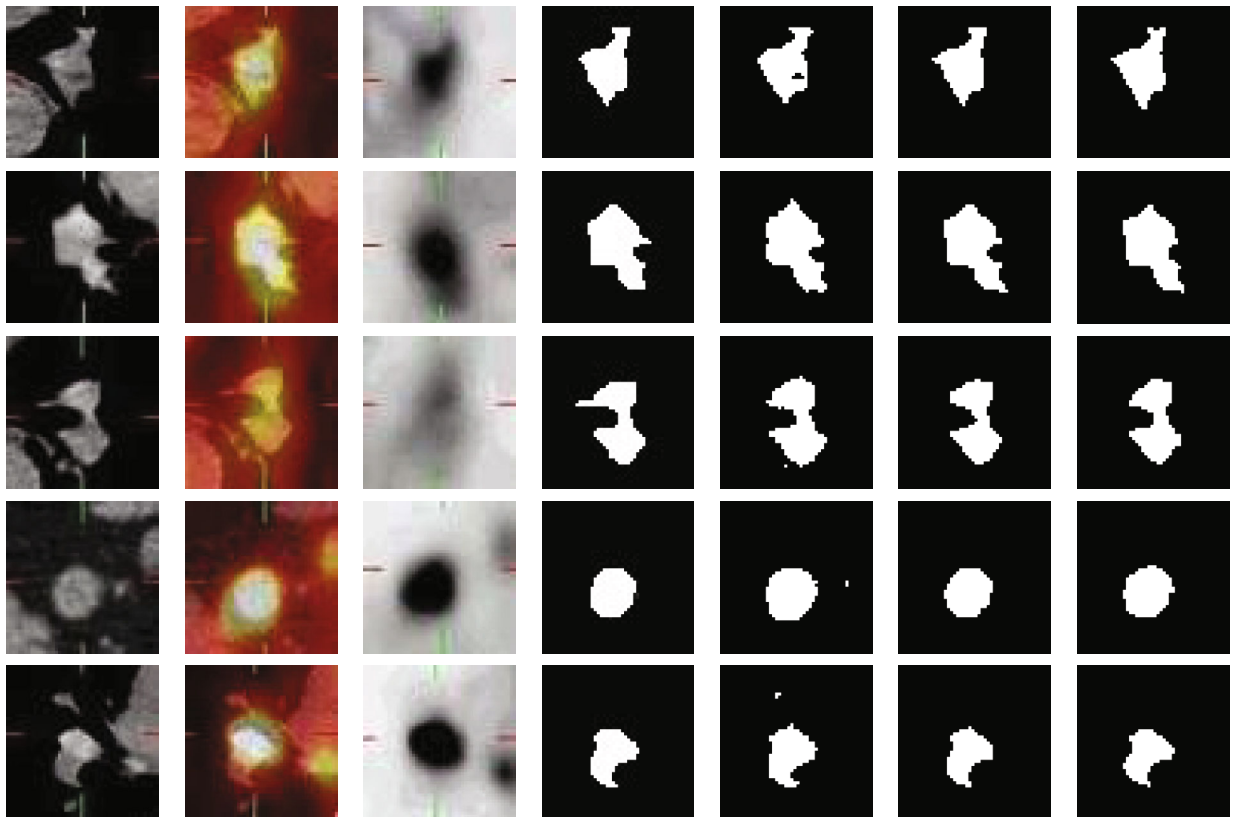


Figure 10: Network segmentation results. (a) CT image. (b) PET/CT image. (c) and (d) PET image. (d) Ground truth. (e) U-Net segmentation result. (f) MEU-Net segmentation result of the three encoders. (g) PU-Net segmentation result.

Table 4: Segmentation index results.

Model	DSC (%)	Recall (%)	VOE (%)	RVD (%)
PU-Net	95.48	95.91	2.72	2.81
SAPU-Net	95.79	96.35	2.87	2.9
CAPU-Net	95.78	96.47	2.88	2.89
MFPU-Net	95.81	96.92	2.9	2.94
APU-Net	96.86	97.53	3.18	3.29

$$VOE = abs\left(1 - \left|\frac{P \cap G}{P \cup G}\right|\right), \tag{9}$$

$$RVD = abs\left(\frac{P}{G} - 1\right), \tag{10}$$

where the correct segmentation of the lesion area is defined as True Positive (TP), the normal tissue area is segmented as the lesion area is defined as False Positive (FP), and the normal area is segmented as True Negative (TN). Segmentation of the focal area into a normal area is defined as False Negative (FN). P represents the target pixel predicted by the model, and G represents the target pixel in the ground truth.

3.2. Performance of the Model

3.2.1. Network Architecture. In experiment 1, the network model is U-Net. 909 CT images are used as the training sets, and 117 CT images are used as the testing set. In experiment 2, the network model is multiencoder U-Net (MEU-Net), this network is based on U-Net. PET, CT, and PET/CT medical images are input into three encoders, respectively. In the last layer of the network, the feature maps of the three encoders are summed, and the parameters of MEU-Net every layer are the same as U-Net, each network layer uses two 3×3 convolution blocks, and convolutional kernels' numbers of each layer are 1024, 512, 256, 128, and 64, respectively. In MEU-Net, we sum the feature maps of the three encoding paths and transmit them to the decoding path through the skip connection. 909 PET/CT, CT, and PET images are used as the training sets, and 117 PET/CT, CT, and PET images are used as the testing set. In experiment 3, the network model is a parallel U-Net network (PU-Net), PU-Net is a network in which APU-Net removes the two-module medical image extraction block, hybrid attention mechanism, and multiscale feature aggregation block. The segmentation results of the UNet_A and UNet_B are summed as the final segmentation result of the PU-Net network. 909 PET/CT, CT, and PET images are used as the training set, and 117 PET/CT, CT, and PET images are used as the testing set. PET/CT and PET images are input into the encoder1_A and encoder2_A, respectively, and CT images are input by the encoder_B. The feature

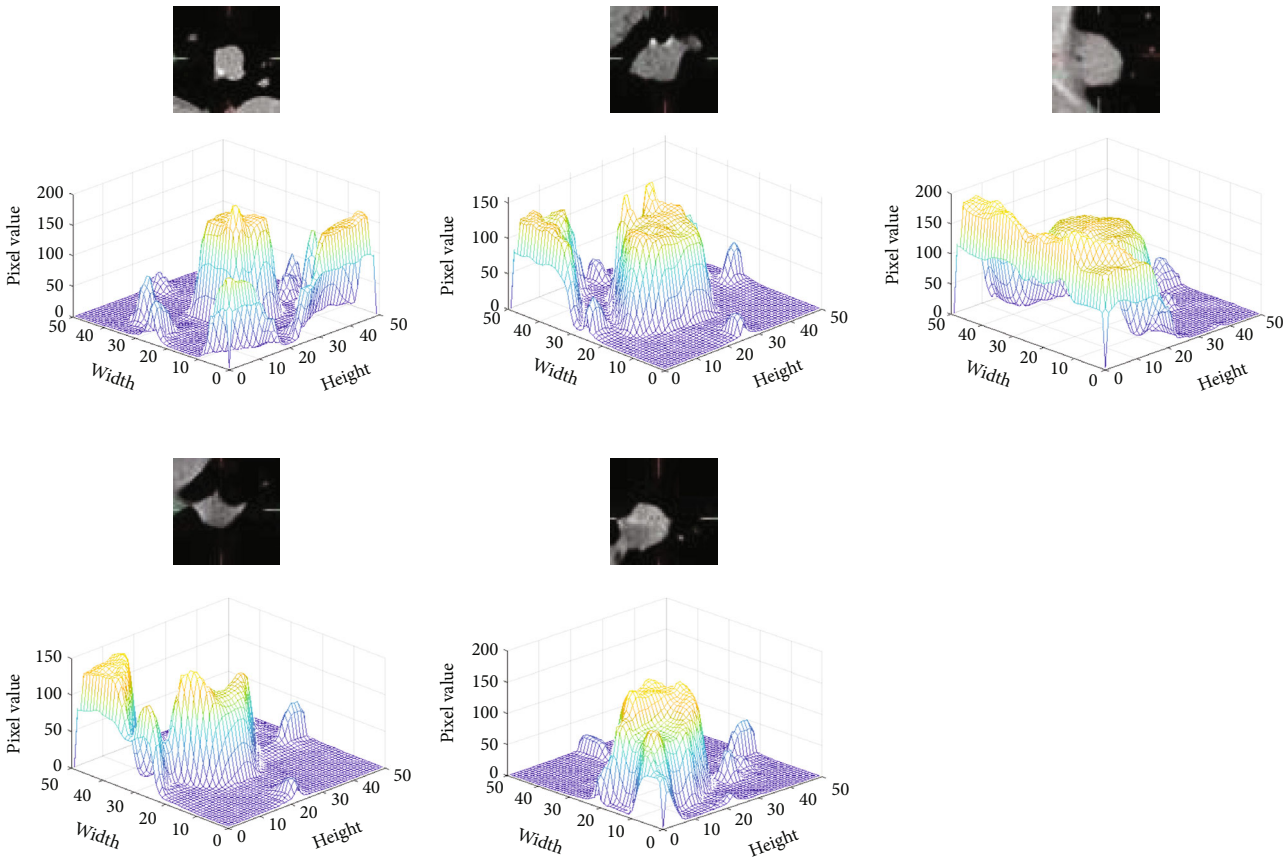


Figure 11: CT image three-dimensional gray value.

maps of encoder_A and encoder_B are concatenated and transmitted to the corresponding decoder_A and decoder_B through skip connection. The segmentation results of the two subnetworks UNet_A and UNet_B are summed as the segmentation result of the PU-Net. The evaluation indexes of segmentation result are shown in Table 3. Figure 9 is the CT image three-dimensional gray value and Figure 10 is the network segmentation results.

Figure 10 shows the segmentation results of U-Net, MEU-Net, and PU-Net. It can be seen that the lesion segmentation results of PU-Net are more accurate than U-Net and MEU-Net. In the third row of Figure 10, comparing with the ground truth of lesion in CT image, PU-Net is superior to the other two networks in delineating the lung tumor lesion shape. It is proved that the parallel U-Net is effective in improving the performance of multimodal medical images segmentation.

3.2.2. Attention Mechanism Module. Four experiments are conducted to evaluate segmentation performance of attention mechanisms and multiscale feature aggregation blocks on APU-Net. All experiments are based on the PU-Net, and 909 PET/CT, CT, and PET images are used as the training sets, and 117 PET/CT, CT, and PET images are used as the testing set. In experiment 1, the network model is PU-Net. In experiment 2, the network model is PU-Net based on the spatial attention mechanism and named as SAPU-NET. The feature maps F_{hybrid_en} and F_{CT_en} are concatenated and input into the spatial attention mechanism of skip connection. In experiment 3, the network is PU-Net based on the channel attention mechanism and named as CAPU-Net. The feature maps F_{hybrid_en} and F_{CT_en} are concatenated and input to the channel attention mechanism of skip connection. In experiment 4, the parallel U-Net is based on multiscale feature aggregation block and named as MFPU-Net, and the feature maps of the UNet_A and UNet_B decoder are concatenated and input into the multiscale feature aggregation block. The evaluation indexes of segmentation result are shown in Table 4. Figure 11 is the CT image three-dimensional gray value and Figure 12 shows the segmentation results of different networks.

As can be seen from Figure 12, segmentation performances are improved by spatial attention mechanism, channel attention mechanism, and multiscale feature aggregation block. SAPU-Net, CAPU-Net, and MFPU-Net are superior to PU-Net in the lung cancer segmentation with complex shapes. As can be seen from the second and fourth row of the Figure 12, the there are some under-segmented phenomenon in PU-Net network, and the segmentation accuracy of APU-Net for the complex shape lesions is better than the network with a single attention mechanism. As can be seen from Table 4, comparing with those of the benchmark network PU-

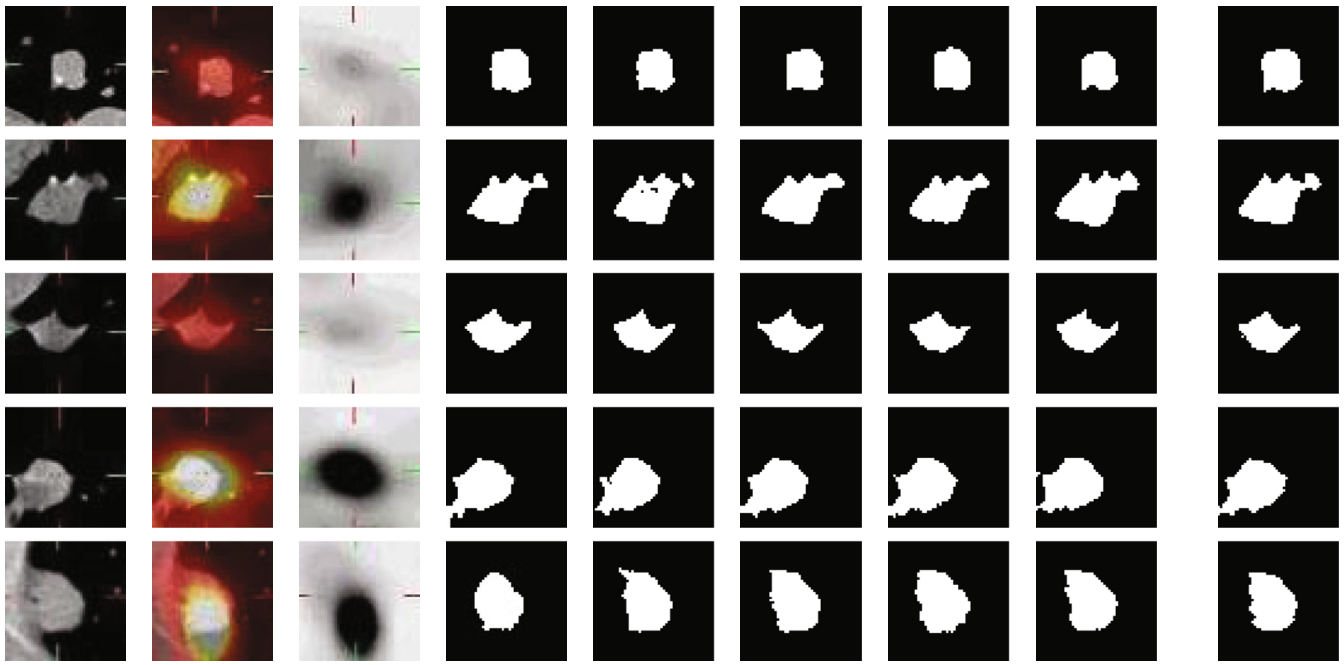


Figure 12: Segmentation results. (a) CT image. (b) PET/CT image. (c) PET image. (d) Ground truth. (e) PU-Net segmentation result. (f) SAPU-Net segmentation result. (g) CAPU-Net segmentation result. (h) MFPU-Net segmentation result. (i) APU-Net segmentation result.

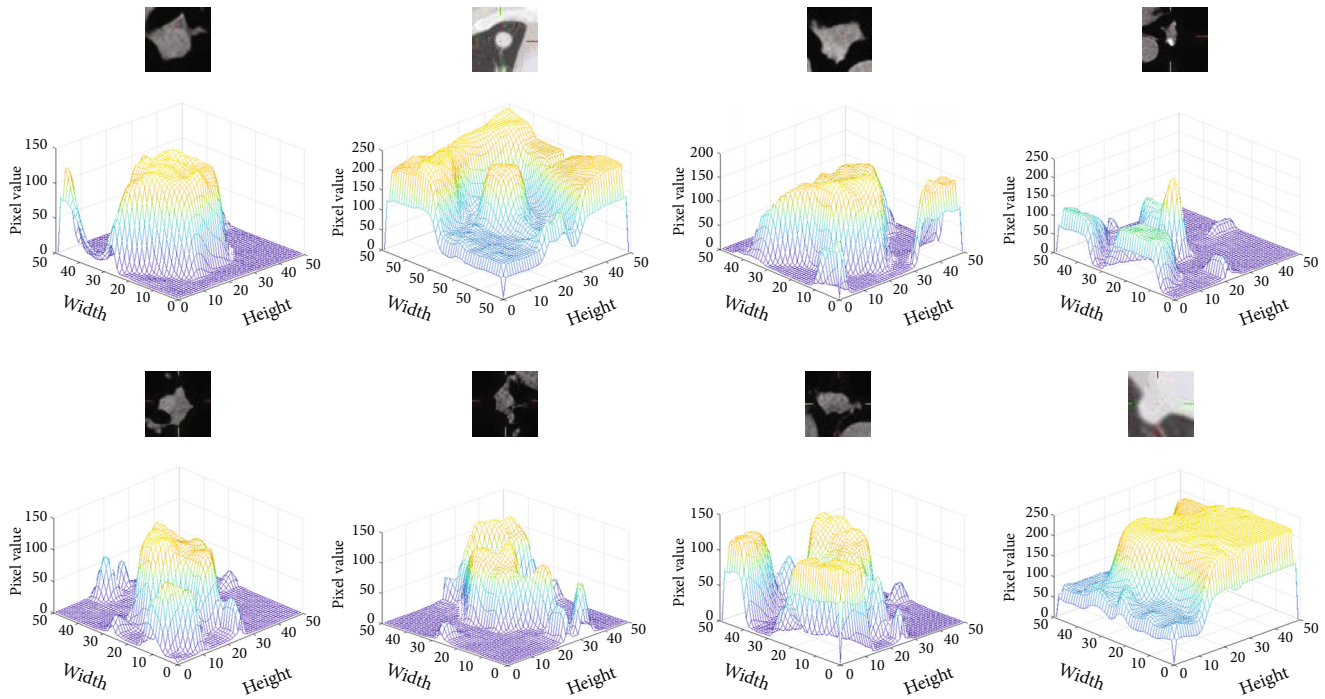


Figure 13: CT image three-dimensional gray value.

Net, the DSC, Recall, VOE, and RVD coefficients of SAPU-Net, CAPU-Net, and MFPU-Net are improved. Compared with spatial and channel attention mechanism, the evaluation indexes of MFPU-Net are the highest among the three mechanisms, these are 95.81%, 96.92%, 2.9%, and 2.94%, respectively, and because the multiscale feature aggregation block is to aggregate the feature maps of decoder_A and decoder_B, MFPU-Net automatically learns the image-specific weight of different scales. In conclusion, this experiment shows that the hybrid attention module and multiscale feature

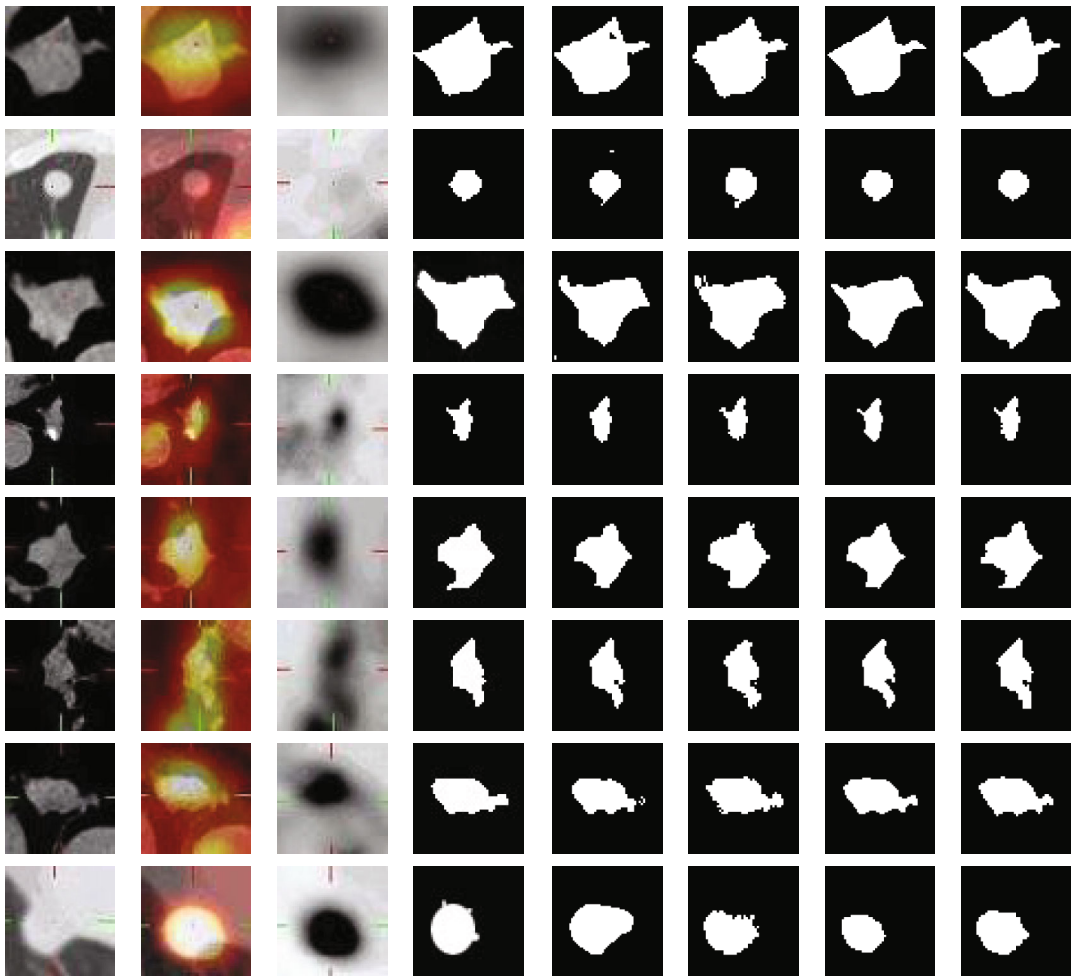


Figure 14: Comparison of segmentation results of different algorithms (a) CT image. (b) PET/CT image. (c) PET image. (d) Ground truth. (e) Segmentation results of SegNet [21]. (f) Segmentation results of Wnet [22]. (g) Segmentation results of Attention Unet [23]. (h) APU-Net segmentation results.

Table 5: Comparison results between APU-Net and other networks.

Model	DSC (%)	Recall (%)	VOE (%)	RVD (%)
SegNet [21]	94.82	95.11	1.81	2.04
Wnet [22]	94.73	95.98	2.08	2.17
Attention Unet [23]	95.69	96.17	2.64	2.73
APU-Net	96.86	97.53	3.18	3.29

aggregation block used in the network are effective in improving the segmentation performance of the lung tumor.

3.2.3. Comparison with Other Networks. In order to the effectiveness of the APU-Net, we compared it with SegNet [21], Wnet [22], and Attention Unet [23]. The encoder network in SegNet [21] is the convolution layer of VGG-16, decoder network is to map the low resolution encoder feature maps to full input resolution feature maps for pixel-wise classification, and the pooling indices are used to compute nonlinear upsampling in decoder. Wnet [22] is that two cascade U-Net networks are used for the segmentation of systemic bone lesions in PET/CT images of myeloma. The first U-Net network has 5 layers, and CT images are the input, while the second U-Net network has 3 layers, and the segmentation results which are the first U-Net and PET images acted as input of the second U-Net network. Attention Unet [23] is used to segment pancreatic CT images lesions, it is an unembedded attention gates in the skip connection, the size and shape of the segmentation target are automatically learned through a self-attention gate, and feature maps in encoder path and decode path are input into self-attention gate. Figure 13 is CT image three-dimensional gray value, the segmentation results of APU-Net and other models are shown in Figure 14, and the evaluation indexes of segmentation result are shown in Table 5.

As can be seen from Table 5, the DSC, Recall, VOE, and RVD coefficients of SegNet [21] are 94.82%, 95.11%, 1.81%, and 2.04%, respectively. SegNet only uses CT single-mode images for the lesion segmentation, and the segmentation

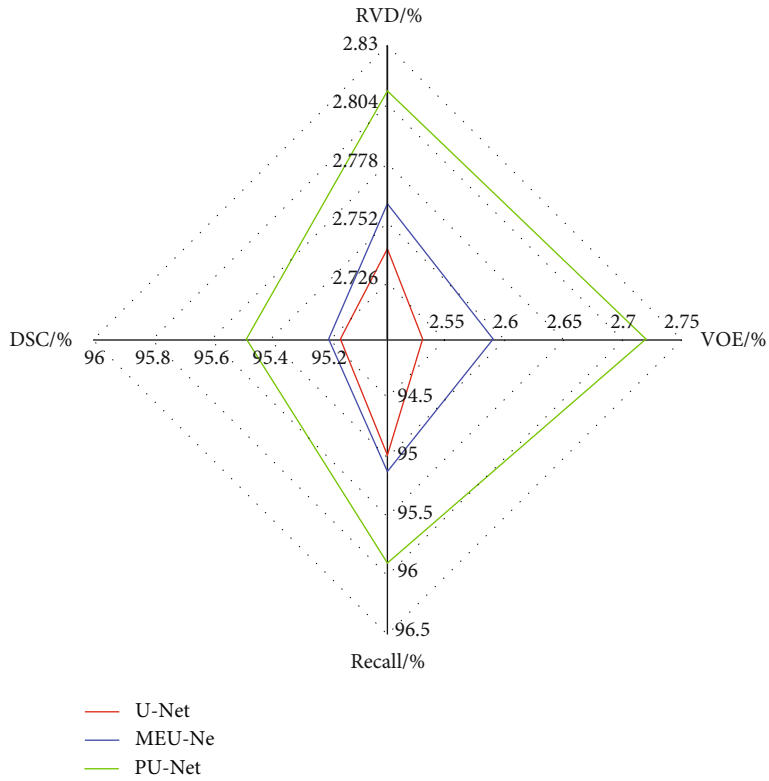


Figure 15: Radar chart of evaluation index of segmentation result.

effect is not good for lesions adhered to normal tissues. As can be seen from Figure [14], there are some under-segmentation in SegNet, such as the 1st row and eth column, there are some oversegmentation, such as last row and eth column, there are some tissue adhesions. DSC, Recall, VOE and RVD coefficients of Wnet [22] are 94.73%, 95.98%, 2.08%, and 2.17%, respectively, and its most coefficients are better than SegNet [21], such as Recall, VOE, and RVD coefficients. Final output feature maps of the first network are input into the second network by this Wnet, by making full use of the details of CT images which helps locate systemic bone lesions and PET images which can provide the metabolic information of systemic bone lesions. However, since the second network of Wnet is very shallow, and medical image feature information cannot be extract well, and lesions' edge is fuzzy in PET images, segmentation result is not good of Wnet comparing with attention mechanism. In the 2nd row and fth column in Figure 14, there are some under-segmentation in Wnet.

For lesions with complex shapes in the 3rd and 7th rows, the Wnet segmentation results are not good. As Attention Unet [23] automatically focuses on the lesion by attention gates in the skip connection, the segmentation effect is better than SegNet [21] and Wnet [22]. However, there is no metabolic information provided by PET images for lesions with complex shapes, such as the 6th row and ath column, and lesions adhered to normal tissues, such as the last row and ath column.

4. Discussion

It is well known that, for patients with lung cancer, delineation of lung lesions plays an important role in the customization of treatment plan and recovery of prognosis [24]. In this study, we proposed an attention mechanism parallel U-Net which enables accurate segmentation for lung cancer. We make full use of multimodal medical images to complement lesion features and use CT images to provide anatomical information of lesions and PET images to provide functional information of lesions. We use the radar chart to show the various segmentation result indicators of each experiment. The radar chart can clearly show the various comparison result indicators of different networks. Figure 15 is the radar chart of the segmentation index of the basic network architecture. From the chart, it can be seen that PU-Net's DSC, Recall, VOE, and RVD coefficients are all higher than U-Net and MEU-Net, and the parallel network architecture can extract more features than U-Net.

Figure 16 is a radar chart of the segmentation index of each attention mechanism network. It can be seen from the figure that except for APU-Net, the coefficients of MFPU-Net are higher than other networks. DSC and Recall coefficients increased from 95.48% to 95.91% and 95.81% to 96.92%, and VOE and RVD coefficients increased from 2.72% to 2.9% and 2.81% to 2.94%. The parallel networks with the attention mechanism are all higher than the

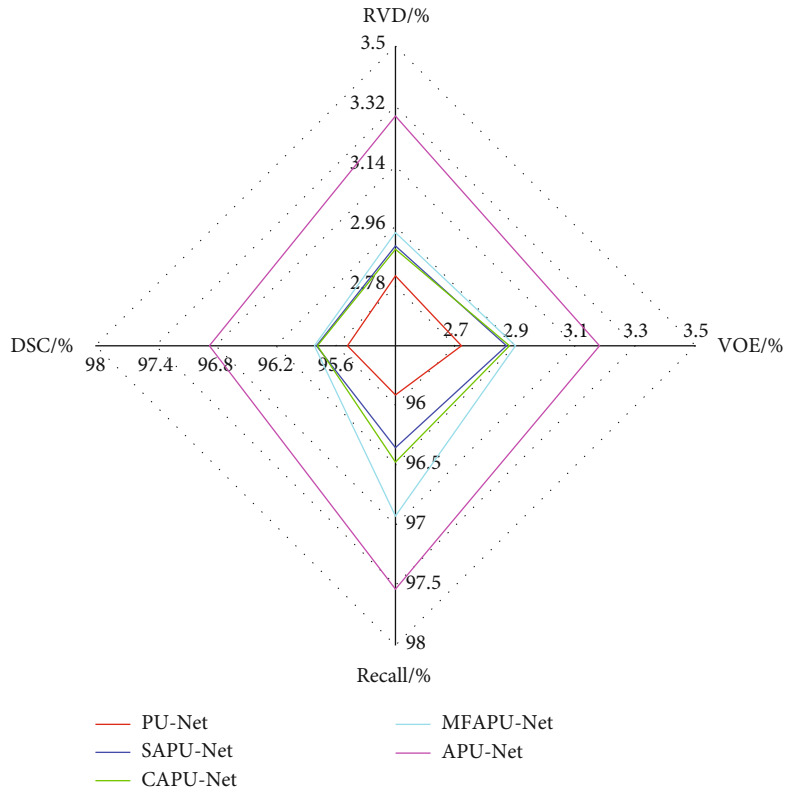


Figure 16: Radar chart of segmentation index.

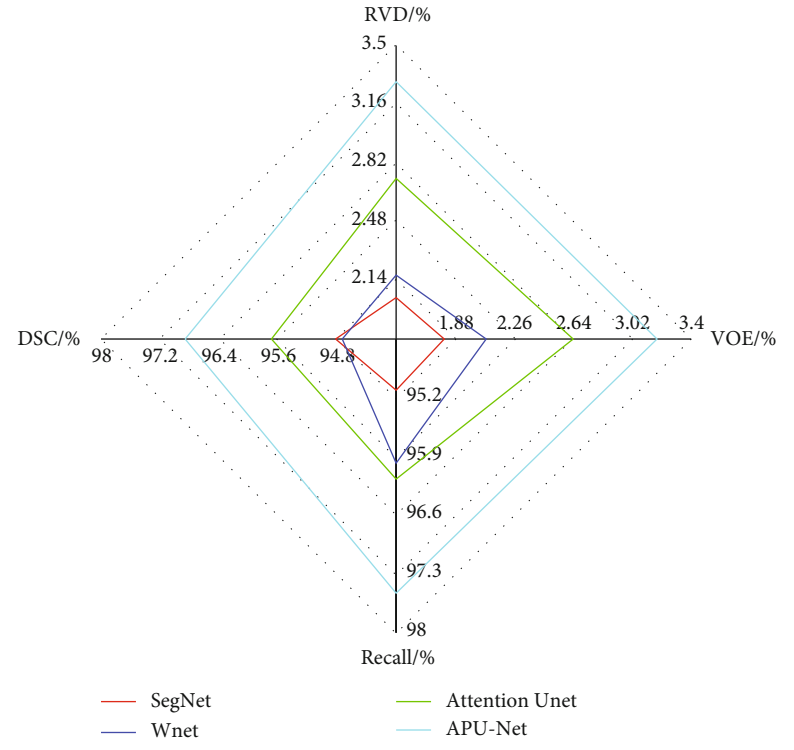


Figure 17: Radar diagram of segmentation index.

benchmark PU-Net, so it can be concluded that the attention mechanism is effective and improves the segmentation performance of the network.

Figure 17 is a radar chart of different model segmentation indicators, and we compared SegNet, Wnet, Attention Unet, and APU-Net, respectively. The DSC, Recall, VOE, and RVD coefficients of Attention Unet [23] are 95.69%, 96.17%, 2.64%, and 2.73%, respectively. Compared with Wnet, the DSC and Recall coefficients are increased by 0.96% and 0.19%, and VOE and RVD coefficients are increased by 0.56%, respectively. It can be seen from the Figure 14 that the evaluation index coefficients of APU-Net are superior to other networks. PET, CT, and PET/CT multimodal medical image features are not used in the networks of SegNet [21] and Attention Unet [23], and single-modal medical images are only used, ignoring the complementary advantages of multimodal medical images. It is feasible to improve the segmentation performance by using multimodal medical images, and the DSC and Recall coefficients of APU-Net are increased by 0.46% and 0.26%, respectively, and VOE and RVD coefficients are increased by 0.18% and 0.23%, respectively. It is very necessary to use a parallel network to learn the features of medical images. In addition, the hybrid attention mechanism selects important feature maps and focuses on the lesions in the feature maps, which can improve the performance of network segmentation and provide a more accurate medical images segmentation, and the segmentation results are closer to the ground truth.

The paper explored two network architectures for extracting complementary information from multimodal medical images to lung tumor. The network based on multi-encoder U-Net and the network based on parallel U-Net. After the adjusting parameters of the two networks, the segmentation performance shows that the based on the parallel network is superior to the network based on the three encoders. It is guessed that the parallel U-Net network has more semantic features on the decoding path than the three-encoder network, and the network has a certain impact on the performance of segmentation. Therefore, the model for extracting multimodal medical image information needs further exploration.

Due to the noninvasive characteristics of medical imaging equipment, it has become a tool for doctors to diagnose diseases [25, 26]. At present, medical image data is increasing explosively, due to their own ethical issues, medical images are difficult to obtain, and the parameters of different hospital imaging equipment are different, resulting in inconsistent medical images obtained [27]. The abovementioned problems lead to certain difficulties in feature extraction of many medical images. Therefore, the standardization of medical image data is one of the important development directions in the future.

In the future, the paper will further expand the dataset and extend the model to the segmentation of 3D lung tumors. The paper will further standardize the prescanning and scanning procedures of ^{18}F-FDG PET/CT and optimize the postprocessing of data reconstruction, in order to maximize the clinical application of ^{18}F-FDG PET/CT in lung malignant tumors.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing financial interests.

Authors' Contributions

Zhou Tao and Yali Dong designed the research; Huiling Lu and Xiaomin Zheng are responsible for the multimodal medical images collection and labeling; Qiu Shi and Senbao Hou analyzed the data; Zhou Tao and Yali Dong cowrote the manuscript. All authors discussed the results and commented on the manuscript.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 62062003), Key Research and Development Project of Ningxia (Grant no. 2020BEB04022), Natural Science Foundation of Ningxia (Grant no. 2022AAC03149), North Minzu University Research Project of Talent Introduction (Grant no. 2020KYQD08), Jiangsu Province Innovation and Entrepreneurship Program (JSSCRC2021569), and Jiangsu Province Medical Distinguished Expert, Wuxi "Taihu Talent Team" project.

References

[1] S. M. Bao, Q. H. Hu, W. T. Yang, Y. Wang, Y. P. Tong, and W. D. Bao, "Targeting epidermal growth factor receptor in non-small-cell-lung cancer: current state and future perspective," *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, vol. 19, no. 8, pp. 984–991, 2019.

[2] C. T. T. Nguyen, F. Petrelli, S. Scuri, B. T. Nguyen, and I. Grappasonni, "A systematic review of pharmacoeconomic evaluations of erlotinib in the first-line treatment of advanced non-small cell lung cancer," *The European Journal of Health Economics*, vol. 20, no. 5, pp. 763–777, 2019.

[3] M. Astaraki, Y. Zakko, I. Toma Dasu, Ö. Smedby, and C. Wang, "Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features," *Physica Medica*, vol. 83, pp. 146–153, 2021.

[4] J. Zhang, Y. Xia, H. Zeng, and Y. Zhang, "NODULe: Combining constrained multi-scale LoG filters with densely dilated 3D deep convolutional neural network for pulmonary nodule detection," *Neurocomputing*, vol. 317, pp. 159–167, 2018.

[5] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109–119, 2019.

[6] G. Tong, Y. Li, H. Chen, Q. Zhang, and H. Jiang, "Improved U-NET network for pulmonary nodules segmentation," *Optik*, vol. 174, pp. 460–469, 2018.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440, Boston, 2015.

[8] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected *CRFs*," *Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Cham, 2015.

[10] L. Liu, J. Cheng, Q. Quan, F. X. Wu, Y. P. Wang, and J. Wang, "A survey on U-shaped networks in medical image segmentations," *Neurocomputing*, vol. 409, pp. 244–258, 2020.

[11] A. Khanna, N. D. Londhe, S. Gupta, and A. Semwal, "A deep residual U-Net convolutional neural network for automated lung segmentation in computed tomography images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1314–1327, 2020.

[12] H. Liu, H. Cao, E. Song et al., "A cascaded dual-pathway residual network for lung nodule segmentation in CT images," *Physica Medica*, vol. 63, pp. 112–121, 2019.

[13] M. Gridach, "PyDiNet: Pyramid dilated network for medical image segmentation," *Neural Networks*, vol. 140, pp. 274–281, 2021.

[14] Z. Wang, Y. Zou, and P. X. Liu, "Hybrid dilation and attention residual U-Net for medical image segmentation," *Computers in Biology and Medicine*, vol. 134, article 104449, 2021.

[15] H. Cao, H. Liu, E. Song et al., "Dual-branch residual network for lung nodule segmentation," *Applied Soft Computing*, vol. 86, article 105934, 2020.

[16] S. Lee, M. Negishi, H. Urakubo, H. Kasai, and S. Ishii, "Mu-net: multi-scale U-net for two-photon microscopy image denoising and restoration," *Neural Networks*, vol. 125, pp. 92–103, 2020.

[17] Y. Takahiro and M. Tsukasa, "Detection and localization of manhole and joint covers in radar images by support vector machine and Hough transform," *Automation in Construction*, vol. 126, article 103651, 2021.

[18] Z. Ying, L. Ge, Y. Ren, R. Wang, and W. Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *International Conference on Computer Analysis of Images and Patterns*, pp. 36–46, Cham, 2017.

[19] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *European Conference on Computer Vision-ECCV*, pp. 3–19, Munich, Germany, 2018.

[20] R. Gu, G. Wang, T. Song et al., "CA-Net: comprehensive attention convolutional neural networks for explainable medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 699–711, 2021.

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[22] L. Xu, G. Tetteh, J. Lipkova et al., "Automated whole-body bone lesion detection for multiple myeloma on 68Ga- pentixafor PET/CT imaging using deep learning methods," *Contrast Media & Molecular Imaging*, vol. 2018, pp. 1–11, 2018.

[23] O. Oktay, J. Schlemper, L. L. Folgoc et al., "Attention u-net: learning where to look for the pancreas," https://arxiv.org/pdf/1804.03999v3.pdf.2021.8.26.

[24] Z. Tao, L. Hui-ling, Y. Zaoli, S. Qiu, B. Huo, and Y. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied soft computing*, vol. 98, article 106885, 2021.

[25] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-Ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[26] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense," *Applied Soft Computing*, vol. 75, pp. 323–332, 2019.

[27] Y. Li, C. Li, X. Li et al., "A comprehensive review of Markov random field and conditional random field approaches in pathology image analysis," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 609–639, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 7088063, 7 pages
https://doi.org/10.1155/2022/7088063

Research Article

The Reaction Mechanism Study for the F_3 System

Dequan Wang,[1] **Nan Gao**,[2] **Hongmei Yu**,[3] **Yuxuan Bai**,[4] **Jing Cao**,[1] **Chunmei Hu**,[5] **Yanchun Li**,[1] **Huiling Liu**,[1] **and Xuri Huang**[1]

[1]*Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Collage of Chemistry, Jilin University, Changchun, China*
[2]*Department of Thoracic Surgery, China-Japan Union Hospital of Jilin University, Changchun, China*
[3]*High Performance Computing Center of Jilin University, China*
[4]*College of Life Science, Sichuan Agricultural University, Ya'an, China*
[5]*Hematology and Oncology Department, The Second Hospital, Jilin University, Changchun, China*

Correspondence should be addressed to Dequan Wang; dequan_wang@jlu.edu.cn and Chunmei Hu; hucm@jlu.edu.cn

Dequan Wang and Nan Gao contributed equally to this work.

Received 2 March 2022; Accepted 25 March 2022; Published 28 April 2022

Academic Editor: Chen Li

Copyright © 2022 Dequan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to study the F_3 system, an accurate global adiabatic potential energy surface is reduced in the present work. The high-level *ab initio* (MCSCF/MRCI level) methods with big basis set aVQZ are used to calculate 27690 potential energy points in the MOLPRO quantum chemistry package using the Jacobi coordinate. Meanwhile, the B-spline fit method is used to reduce the global potential energy surface in this present work. The shallow well complexes are found in the present work when the angles $\theta = 30°, 60°,$ and $90°$. Analysing the global potential energy surfaces can get the conclusion that reactants should overcome at least 0.894 eV energy to cross the transition state and reach products. This study will be helpful for the analysis in histopathology and for the study of biological and medical mechanisms.

1. Introduction

As the most active element of halogens, fluorine often plays an important role in the chemical reaction process. In the chemical industry, fluorine is widely used in the fields of semiconductors and new energy, which is particularly critical for future development. Therefore, the three-dimensional potential energy surface of F_3 is very interesting. Duggan and Grice have described the approximate shape of the potential energy surface of F_2+F [1]. Artau et al. [2] determined the dissociation energy of the F_2 bond by the collision-induced dissociation energy method.

Although many studies have been made on fluorine [1–8], while exchange reactions between halogen atoms and halogen atoms have been studied by many scholars [9–16], the exact understanding of the three-dimensional potential energy surface for F_3 is so limited. Currently, quantum chemistry calculations are able to provide sufficiently reliable data for this system, and potential energy surfaces (PESs) can help us to understand possible reaction mechanisms and estimate reaction rates. The potential energy surface of F_3 was studied using the classical orbital method combined with direct dynamic electronic structure calculation. By now, a train of scholars have studied the exchange reaction of trifluoride ions [2–4], but there is no precise dynamic study for this system. In order to accurately describe this system, three-dimensional PESs for the title system are constructed in this paper.

In the current work, the accurate three-dimensional PESs for this system are deduced. The details of this paper

are shown as follows: the computational method will be introduced in the second chapter and the detailed description of PESs will be introduced in the third section. In the last part, the conclusion and discussion will be demonstrated.

2. Computational Methods

Three adiabatic potential energy points for F_3 are calculated with MCSCF/MRCI method with aVQZ [17] basis sets. The lowest potential energy surface is described in this study. All the calculations are performed in the MOLPRO 2012 package [18]. A total of 12 ($9A' + 3A''$) active orbitals are considered in the present work. And 225 ($138A' + 87A''$) external orbitals are used in this work. Thus, the MCSCF function included 360 determinants and 792 intermediate states for the triatomic molecule system. In the MRCI calculated progress, 3 orbitals ($3A' + 0A''$) are put in the core orbital, and 24 electrons are set in the valence space. The total number of contracted configurations is 595861.

The Jacobi coordinate (r, R, θ), which can express the three-dimensional PESs well in our previous works on triatomic systems [3, 19–23], is used in the present work. r is used to show the distance between two fluorine atoms, R demonstrates the distance from the reduced mass of the two fluorine atoms to the third fluorine atom, and the angle between these two vectors is presented by θ (see Figure 1). R is used in the range of 0.0 Å-20.0 Å with 69 points, r is fixed from 0.4 Å to 5.0 Å with 39 points using different step sizes, and the range of angles θ is 0°-90° with a 10° scan step. Therefore, in total, 27690 geometries are calculated. So the large number of scanning points that can warrant the quality of the following fitted the PESs. The three-dimensional B-spline method [24, 25] is used to get the procedure interpolate surface in the whole scan field.

3. Results and Discussion

The three-dimensional adiabatic potential energies of the lowest energy state for the $F_3(^2A'')$ system is deduced in the present work; for a clearer understanding of the character of this system, the two-dimensional (2D) PES is shown in the following part. For a clearer discussion of these PESs, the energy of F+F_2 is shifted to 0.00 eV.

The one-dimensional potential energy surface for F_2 is studied in the present work. The dissociation energy is 1.40 eV, which compares well to the former theoretical and experimental results (1.25 eV and 1.60 eV, respectively) [2, 26].

The 2D lowest energy state adiabatic potential energy surfaces (2D-PESs) have been shown in different angles for the $F_3(^2A'')$ energy surface. The four different angles ($\theta = 0°, 30°, 60°$, and 90°) are used to describe the vivid change of the PES. The upper part of each figure is the 2D image PES, and the corresponding contour plots of each 2D-PES are shown in the lower part of each figure. In the contour plotted part, the difference of the energy between two adjacent curves is 0.20 eV; the short distance between two adjacent curves means big energy changes with distance (R or r) and vice verse. Some important parts (isomers or transition state parts)are presented in enlarged panel in each figure.

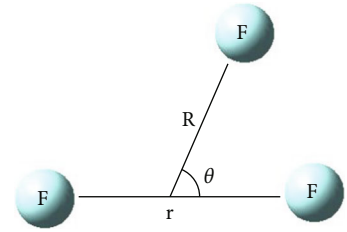


Figure 1: Jacobi coordinate for the $F_3(^2A'')$ system.

The 2D-PES for $F_3(^2A'')$ in linear symmetry is plotted in Figure 2. From this figure, it can be found that the potential energies increase with the decrease of the distance of the R. There is no minimum in this figure. The PES of the $F_3(^2A'')$ system in linear symmetry is flat when the r is in the range of $1.3\,Å < r < 1.6\,Å$ and $R > 3.0\,Å$. In this area, the energy is lower than 0.2 eV and larger than 0.0 eV, so this area is embraced by the blue curve (the blue curve is the contour line of the energy of 0.2 eV). An enlarged figure is plotted to show the transition state (TS) area. The geometry of this TS is $r = 1.68\,Å$ and $R = 2.46\,Å$. The linear symmetry of reactants F_a+F_bF_c overcomes 1.089 eV energy crossing the TS then transforming to the products F_aF_b+F_c.

The character of the PES for the $\theta = 30°$ is shown in Figure 3. There is a minimum in this angle, the geometry of the minimum is $r = 1.40\,Å$, and $R = 3.65\,Å$. The energy of this minimum is -0.025 eV, which is slightly lower than that of reactants. The difference of these two energies is so small that the minimum is not stable. The reaction pathway along the $\theta = 30°$, and the reaction barrier is 0.88 eV.

The character of the PES with $\theta = 60°$ (see Figure 4) is similar to that of $\theta = 30°$: Firstly, both of these two figures have a shallow minimum; the energy of the minimum for $\theta = 60°$ is -0.03 eV and the geometry of this minimum is $r = 1.40\,Å$ and $R = 3.11\,Å$. Secondly, the reactants want to change to the products through these two-angles reaction pathway and the additional energies should be given to this system to overcome the transition state. The energy of the TS for $\theta = 60°$ is 1.255 eV, which is higher than that of $\theta = 30°$. The geometry of the TS is $r = 1.63\,Å$ and $R = 1.84\,Å$.

Figure 5 shows the adiabatic PES of the title system for $\theta = 90°$. When the F atom attacks the F_2 molecule along $\theta = 90°$, and for the same r, the energy is nearly unchanged with the function of R when $R > 3.5\,Å$; when R is decreasing to 2.95 Å and $r = 1.40\,Å$, the system reaches the shallow well complex; the energy of this complex is -0.026 eV. After this complex, the energy of this system becomes higher and higher, decreasing the R distance further.

The figures for $\theta = 10, 20, 40, 50, 70$, and 80° also have been plotted in Figure S1-S6 (see the Supporting Information).

The 2D-PES for $F_3(^2A'')$in the function with $r(F_aF_b)$, r (F_bF_c) and $\alpha = 180°$ is plotted in Figure 6. The reaction

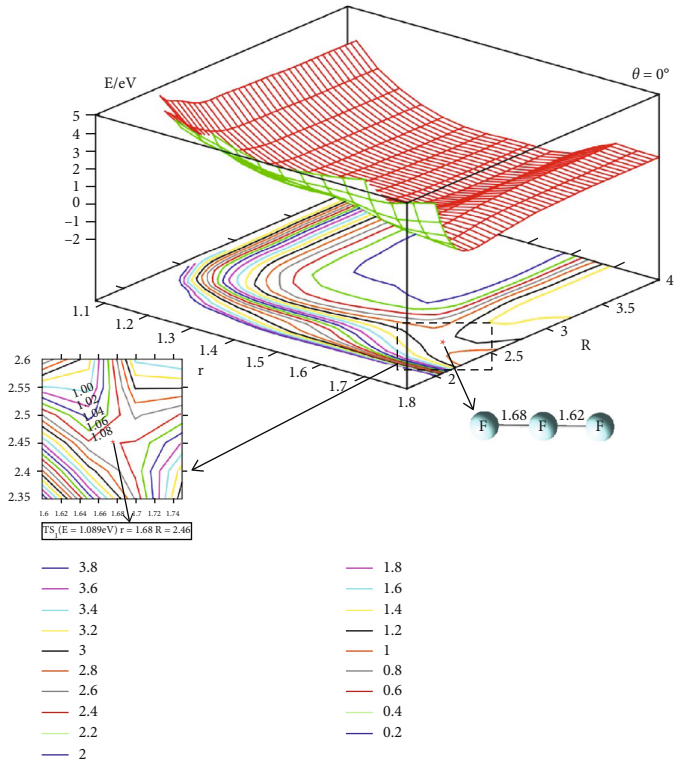


FIGURE 2: Adiabatic potential energy surface (in eV) and contour plots of the potential energy surface for the lowest energy state of $F_3(^2A'')$ as the function of r and R (in Å) in C_{∞} symmetry.

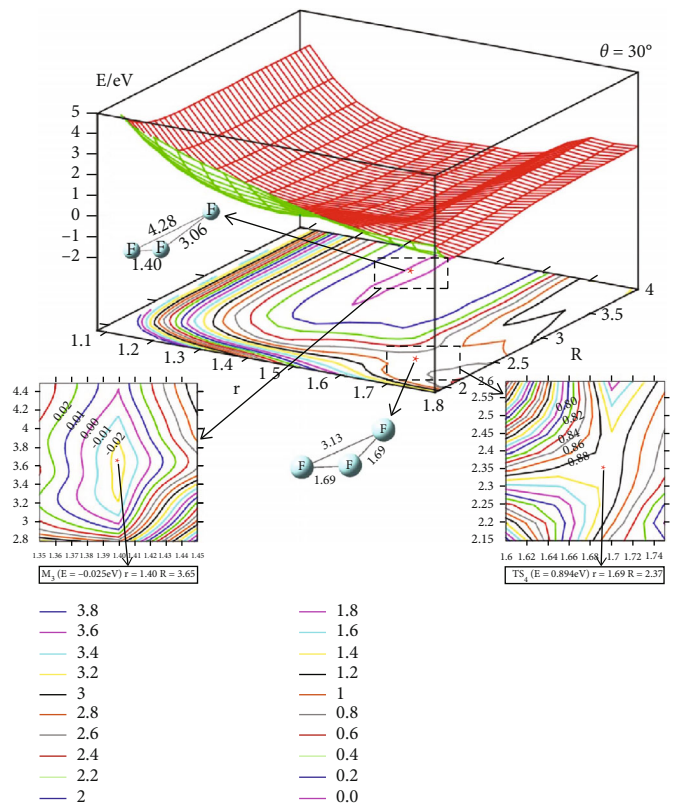


FIGURE 3: Adiabatic potential energy surface (in eV) and contour plots of the potential energy surface for the lowest energy state of $F_3(^2A'')$ as the function of r and R (in Å) for $\theta = 30°$. The important isomer and transition sate are shown in the lower panels.

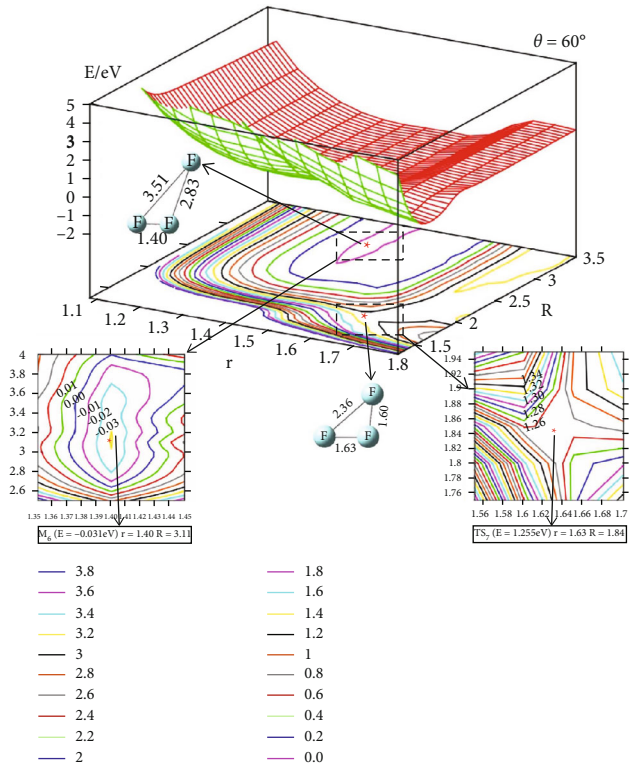


Figure 4: Adiabatic potential energy surface (in eV) and contour plots of the potential energy surface for the lowest energy state of $F_3(^2A'')$ as the function of r and R (in Å) for $\theta = 60°$. The important isomer and transition sate are shown in the lower panels.

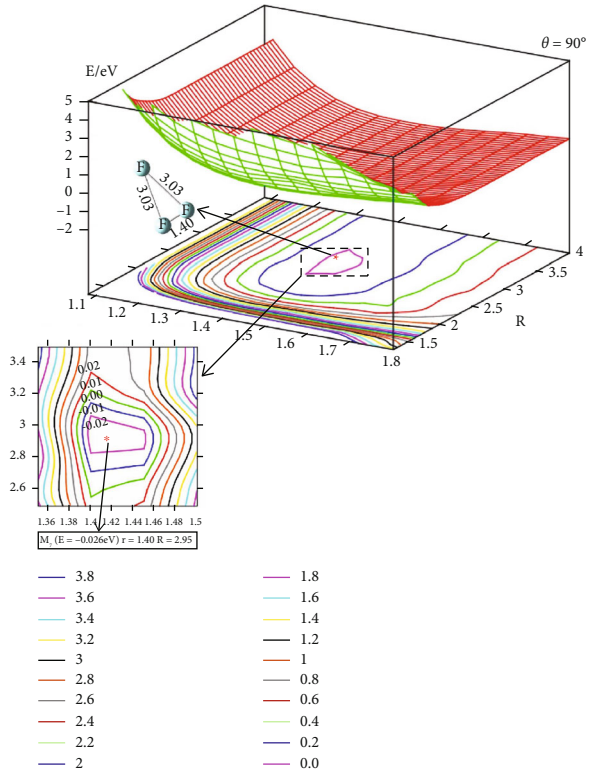


Figure 5: Adiabatic potential energy surface (in eV) and contour plots of the potential energy surface for the lowest energy state of $F_3(^2A'')$ as the function of r and R (in Å) for C_{2v} symmetry. The importance isomer is shown in the lower panel.

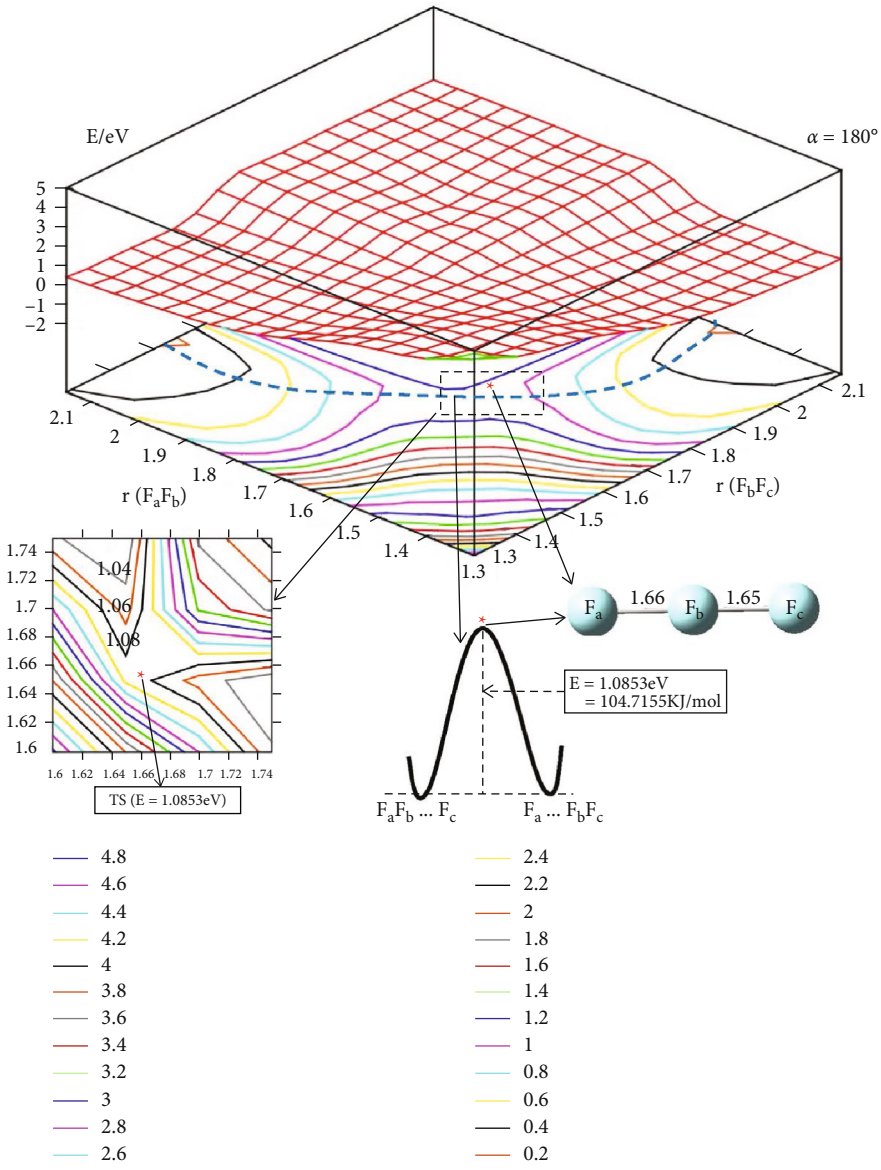


FIGURE 6: Adiabatic potential energy surface (in eV) and contour plots of the potential energy surface for the lowest energy state of $F_3(^2A'')$ as the function of $r(F_aF_b)$ and $r(F_bF_c)$ (in Å) in C_{∞} symmetry.

barrier is 104.715 kJ/mol, this is good agreement with Duggan and Grice's work(108 Kj/mol). The important region is a magnified plot in the same figure. The blue dotted line shows the easiest reaction pathway for the title reaction. Meanwhile, the sketch map of this dotted line is plotted in the lower panel, which shows that the reaction barrier is 104.715 kJ/mol; compared with Duggan and Grice's work (108 kJ/mol) [1], the results are almost consistent. The geometry of the transition state for the reaction pathway is $r(F_aF_b) = 1.66$ Å and $r(F_bF_c) = 1.65$ Å.

4. Conclusions

The electronic adiabatic state PESs for the lowest energy state of F_3 are presented with the high-level *ab initio* (MCSCF/MRCI level) method by the MOLPRO quantum chemistry package in the Jacobi coordinate. The potential energies are calculated using the aVQZ basis set over a larger region of configuration space. 27690 potential energy points are calculated for getting the globally accurate PESs, which is deduced by an accurate B-spline fit method. There is no minimum in the PES with $\theta = 0.0°$, but there is a transition state in this PES. The process of $F_a+F_bF_c \rightarrow F_aF_b+F_c$ along the angle $\theta = 0.0°$ should be given additional energy of 1.089 eV, but for $\theta = 30.0°$, this energy is reduced to 0.894 eV; when the angle increases to 60°, the reaction barrier will reach 1.255 eV. There are shallow complexes in the PES for $\theta = 30°, 60°,$ and $90°$ and the energies are -0.025 eV, -0.031 eV, and -0.026 eV, respectively. The proposed method in this study will be helpful for the development of biological and medical mechanisms.

It is worth performing the full dynamical study with this global potential energy surface. We will continue this work in the following study.

Data Availability

Data are available, which are included in the code.

Disclosure

A preprint has previously been published by Wang et al. [27].

Conflicts of Interest

The authors declare no competing interests.

Authors' Contributions

Dequan Wang and Nan Gao contributed equally to this work.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 21573090) and Jilin Province Science and Technology Development Plan (20150101005JC). This work completed under cooperation between the Institute of Theoretical Chemistry and High Performance Computing Center of Jilin University, China.

Supplementary Materials

The supporting information is available free of charge at BioMed Research International. The figures for $\theta = 10, 20, 40, 50, 70,$ and 80° also have been plotted in Figure S1-S6 (PDF). *(Supplementary Materials)*

References

[1] J. J. Duggan and R. Grice, "Topography of potential-energy surfaces. DIM (diatomics-in-molecules) surface for F + F2," *Journal of the Chemical Society, Faraday Transactions 2*, vol. 80, no. 7, pp. 795–808, 1984.

[2] A. Artau, K. E. Nizzi, B. T. Hill, L. S. Sunderlin, and P. G. Wenthoid, "Bond dissociation energy in trifluoride ion," *Journal of the American Chemical Society*, vol. 122, no. 43, pp. 10667–10670, 2000.

[3] D. Q. Wang, D. G. Wang, L. W. Fu et al., "An accurate ground state potential surface for the scattering reaction $F^- + F_2(v,j) \rightarrow F_2(v',j') + F^-$," *RSC Advances*, vol. 9, pp. 1929–1932, 2019.

[4] A. A. Tuinman, A. A. Gakh, R. J. Hinde, and R. N. Compton, "The first direct observation of the trifluoride anion (F_3^-) in the gas phase," *Journal of the American Chemical Society*, vol. 121, pp. 8397-8398, 1999.

[5] F. Mota and J. J. Novoa, "The symmetry breaking problem in the triflouride anion: a multireference approach," *The Journal of Chemical Physics*, vol. 105, no. 19, pp. 8777–8784, 1996.

[6] T. Takayanagi and Y. Kurosaki, "Theoretical calculations of potential energy surface and thermal rate constants for the H(Mu) + F2Reaction," *The Journal of Physical Chemistry A*, vol. 101, no. 38, pp. 7098–7104, 1997.

[7] B. S. Ault and L. Andrews, "Matrix reactions of alkali metal fluoride molecules with fluorine. Infrared and Raman spectra of the trifluoride ion in the M+F3- species," *Journal of the American Chemical Society*, vol. 98, no. 6, pp. 1591–1593, 1976.

[8] G. G. Balint-Kurti, "Potential energy surfaces for simple chemical reactions: $Li+F_2 \rightarrow LiF+F$," *Molecular Physics*, vol. 25, no. 2, pp. 393–435, 1973.

[9] Y. T. Lee, J. D. Mcdonald, P. R. Lebreton, and D. R. Herschbach, "Molecular-beam kinetics: evidence for short-range attraction in halogen atom-molecule exchange reactions," *The Journal of Chemical Physics*, vol. 49, no. 5, pp. 2447-2448, 1968.

[10] D. Beck, F. Engelke, and H. J. Loesch, "Reaktive streuung in molekularstrahlen: $Cl+Br_2$," *Bet Bunsenges Phys Chem*, vol. 72, 1968.

[11] Y. T. Lee, P. R. LeBreton, J. D. McDonald, and D. R. Herschbach, "Molecular beam kinetics: evidence for preferred geometry in Interhalogen exchange reactions," *The Journal of Chemical Physics*, vol. 51, no. 1, pp. 455-456, 1969.

[12] N. C. Blais and J. B. Cross, "Molecular beam kinetics: the differential cross section of the reaction $Cl + Br_2$," *The Journal of Chemical Physics*, vol. 52, no. 7, p. 3580, 1970.

[13] J. J. Valentini, Y. T. Lee, and D. J. Auerbach, "A crossed molecular beam study of the translational energy dependence of $Cl + Br_2 \rightarrow BrCl + Br$ reaction," *The Journal of Chemical Physics*, vol. 67, no. 11, pp. 4866–4872, 1977.

[14] J. J. Duggan and R. Grice, "Topography of potential-energy surfaces. DIM (diatomics-in- molecules) surfaces for F + HF and Cl + HCl," *Journal of the Chemical Society, Faraday Transactions 2*, vol. 80, 1984.

[15] J. J. Duggan and R. Grice, "Topography of potential-energy surfaces. DIM (diatomics-in-molecules) surface for $Cl + Cl_2$," *Journal of the Chemical Society, Faraday Transactions 2*, vol. 80, no. 7, pp. 809–816, 1984.

[16] N. C. Firth and R. Grice, "Topography of potential-energy surfaces. Spin-orbit interaction for $F + F_2$, $Cl + Cl_2$," *J. Chem. Soc., Faraday Trans. 2*, vol. 83, no. 6, pp. 1011–1022, 1987.

[17] T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *The Journal of Chemical Physics*, vol. 90, no. 2, pp. 1007–1023, 1989.

[18] H. J. Werner, P. J. Knowles, G. Knizia, and F. R. Manby, "Schtz and others; MOLPRO, version 2012.1, a package of ab initio programs," http://www.molpro.net.

[19] D. Q. Wang, G. Shi, L. W. Fu, R. L. Yin, and Y. B. Ji, "Accurate Potential Energy Surfaces for the Three Lowest Electronic States of N(2D) + H2(X1$\sum$g+) Scattering Reaction," *ACS Omega*, vol. 4, no. 7, pp. 12167–12174, 2019.

[20] D. Q. Wang, L. W. Fu, Z. X. Qu, Y. K. Chen, and X. R. Huang, "Accurate potential surfaces for the ground state of $H + C_2$ reaction," *The European Physical Journal D*, vol. 71, no. 10, pp. 252–258, 2017.

[21] L. W. Fu, D. Q. Wang, Y. K. Chen, and X. R. Huang, "Accurate potential surfaces for the first two lowest of $Li + H_2$ reaction," *RSC Advances*, vol. 8, no. 28, pp. 15595–15602, 2018.

[22] D. Q. Wang, D. G. Wang, L. W. Fu, and X. R. Huang, "Accurate potential surfaces for the first three lowest states of reaction $O(^3P) + C_2(a^3\Pi_u) \rightarrow CO(X^1\sum) + N(^1D)$," *Chemical Physics*, vol. 517, pp. 228–236, 2019.

[23] R. L. Yin, N. Gao, R. M. Zhang, D. Q. Wang, and X. R. Huang, "Accurate potential energy surfaces for the excited state of CF_2 molecule," *Chemical Physics*, vol. 538, pp. 110906–110913, 2020.

[24] N. Sathyamurthy and L. M. Raff, "Quasiclassical trajectory studies using 3D spline interpolation of ab initio surfaces," *The Journal of Chemical Physics*, vol. 63, no. 1, article 464473, pp. 464–473, 1975.

[25] M. Patrício, J. L. Santos, F. Patrício, and A. J. C. Varandas, "Roadmap to spline-fitting potentials in high dimensions," *Journal of Mathematical Chemistry*, vol. 51, no. 7, pp. 1729–1746, 2013.

[26] K. P. Huber and G. Herzberg, "Constants of diatomic molecules (data prepared by Gallagher JW and Johnson RD, III)," in *In NIST chemistry web book, NIST standard reference database 69*, W. G. Mallard and P. J. Linstrom, Eds., National Institute of Standards and Technology, Gaithersburg, MD, 2000, http: //webbook http://nist.gov.

[27] J. Cao, C. M. Hu, H. M. Yu et al., *Three dimensional potential energy surface for the F3*, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 2384830, 22 pages
https://doi.org/10.1155/2022/2384830

Review Article

Dense Convolutional Network and Its Application in Medical Image Analysis

Tao Zhou,[1,2] **XinYu Ye**,[1,2] **HuiLing Lu**,[3] **Xiaomin Zheng**,[4] **Shi Qiu**,[5] **and YunCan Liu**[1,2]

[1]*School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China*
[2]*Key Laboratory of Image & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China*
[3]*School of Science, Ningxia Medical University, Yinchuan 750004, China*
[4]*Research Institute for Reproductive Medicine and Genetic Diseases, Wuxi Maternity and Child Health Hospital, Wuxi, Jiangsu 214002, China*
[5]*Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China*

Correspondence should be addressed to XinYu Ye; 3303626778@qq.com and HuiLing Lu; lu_huiling@163.com

Received 28 February 2022; Accepted 23 March 2022; Published 25 April 2022

Academic Editor: Chen Li

Copyright © 2022 Tao Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dense convolutional network (DenseNet) is a hot topic in deep learning research in recent years, which has good applications in medical image analysis. In this paper, DenseNet is summarized from the following aspects. First, the basic principle of DenseNet is introduced; second, the development of DenseNet is summarized and analyzed from five aspects: broaden DenseNet structure, lightweight DenseNet structure, dense unit, dense connection mode, and attention mechanism; finally, the application research of DenseNet in the field of medical image analysis is summarized from three aspects: pattern recognition, image segmentation, and object detection. The network structures of DenseNet are systematically summarized in this paper, which has certain positive significance for the research and development of DenseNet.

1. Introduction

Deep learning is an end-to-end approach to extracting and abstracting image features layer by layer and implementing recognition functions. Krizhevsky et al. [1] used deep learning to win the champion of computer vision challenges in ImageNet. Therefore, deep learning methods represented by convolutional neural networks have become a popular research topic in the field of pattern recognition. GoogLeNet, ResNet, and DenseNet have been proposed one after another. Among them, DenseNet [2] has achieved better results among many deep learning models due to the new architecture of dense connectivity, DenseNet achieves interconnection between arbitrary layers and skip connection mode that transfers information from shallow layers directly to deep layers, and enhanced feature transfer and feature reuse between network layers, resulting in a more compact network representation with less feature redundancy. Under the condition of same layer depth, the convergence performance of network is better, the network degradation and gradient disappearance problems caused by the deepening of convolutional network are better alleviated, and the number of network parameters and computational efficiency are significantly reduced.

Litjens et al. [3] point out that deep learning algorithms, especially convolutional networks, have been successful in medical image analysis. Xie et al. [4] point out that deep learning is widely used in medicine for disease diagnosis, lesion and abnormality detection, lesion and organ segmentation, etc. In recent years, deep learning has developed rapidly in the medical field, represented by companies such as Watson, Microsoft Nuance, Google Health in the world and iFlytek, Tencent Miying, and Ali Health in China. Eagle Pupil Airdoc [5] fundus image-assisted diagnosis system for diabetic retinopathy (DR) has high sensitivity and specificity in detecting, DR. Dr. Wise [6], launched by DeepMed, can

achieve rapid identification, accurate measurement, diagnostic analysis, and scientific follow-up of pneumonia signs, providing clinicians with a basis for treatment, thus realizing precision medicine.

DenseNet has made a breakthrough in medical image analysis tasks [7]; Li and Liu [8] used DenseNet to learn local block features of MRI brain image clusters and achieved a better Alzheimer's disease classification accuracy of 89.5%; Jégou et al. [9] extended densely connected structure to fully convolutional neural network by introducing different depth paths to produce different scale nonlinear mappings for semantic segmentation, thus avoiding gradient disappearance and enabling deeper networks to be trained with fewer parameters; Ke et al. [10] based on 3D DenseNet performed automatic detection and segmentation of nasopharyngeal carcinoma in MRI images, which can obtain higher overall accuracy, sensitivity, and specificity than radiologists. DenseNet can provide clinical aid diagnostic solutions for major diseases such as benign and malignant tumors, cardiovascular and cerebrovascular diseases, and respiratory diseases. Liu et al. [11] based on DenseNet extracted spatiotemporal features of patients' cardiac magnetic resonance imaging (MRI) to predict the ray fraction of left ventricle; Kishan and Janardhan [12] aimed at the problem of small intracerebral hemorrhage and high square difference and used a combination of DenseNet and Inception V3 models to accurately detect and identify early intracranial hemorrhage; Li et al. [13] used transfer learning and DenseNet to classify benign and malignant COVID-19 CT images, reaching the state-of-the-art in terms of accuracy and F1 score.

This paper summarizes DenseNets, and the rest of the structure is as follows: Section 2 introduces the basic principles of DenseNet; Section 3 reviews the development of DenseNet models and summarizes five aspects of broaden network structure, lightweight network structure, dense unit, dense connection mode, and attention mechanism; Section 4 summarizes the application of DenseNet in field of medical image analysis and explores three aspects of pattern recognition, image segmentation, and object detection.

2. Basic Principles of DenseNet

In 2017, Huang et al. [2] proposed DenseNet, a convolutional neural network with densely connected structure, and DenseNet structure is shown in Figure 1.

DenseNet has the characteristics of feature sharing and arbitrary interlayer interconnection. The advantages of DenseNet are feature map reuse through dense connection, reducing interdependence between layers by reusing feature maps from different layers, providing compact and differentiated input features by shortcut connections of different lengths, and effectively reducing the gradient disappearance problem that is difficult to optimize in deep networks. The final prediction is to use features from all layer to obtain better performance and model robustness on a standard dataset with smaller model size and computational effort. The disadvantage of DenseNet is that the feature maps of each layer are spliced with the previous layer, and the data is replicated multiple times. As the number of network layers increases, the number of model parameters grows linearly, eventually leading to explosive growth in computation and memory overhead during training.

3. Development of DenseNet

There are five improvement methods of DenseNet, such as broaden DenseNet structure, lightweight DenseNet structure, dense unit, dense connection mode, and attention mechanism. This section summarizes the development of DenseNet, as shown in Figure 2.

In 1998, LeCun et al. [14] proposed LeNet model, in which some new operators such as convolutional operator, pooling operator, and full connection operator were used, and LeNet become a pioneer model in deep learning models.

In 2012, Krizhevsky et al. [1] proposed AlexNet, in which ReLU was used for activation function; network architecture uses dual GPUs, reduced pooling steps, enhanced training data, and dropout function to avoid overfitting; this network achieved 10.9% better result than second place in ILSVRC competition.

In 2015, He et al. [15] proposed ResNet for the problem of network with more stacked structural layers but fast performance degradation, using residual connections to transfer shallow features directly to deeper layers, so that the originally fitted output $H(x)$ becomes the residual output $H(x) - x$. Even when learning structural layer with 0 features, ResNet only does constant mapping and learns new features based on input features; ResNet learns new features based on input features, alleviating the degradation problem that occurs in deep models and giving network better performance.

In 2017, Huang et al. [2] further improved feature reuse capability based on ResNet and proposed DenseNet network with dense connection operation, but in the case of dense connection as the number of layers of dense blocks increases, input feature map dimension at the proximal end becomes larger and larger, to alleviate this problem Huang et al. [2] added 1×1 convolution (bottleneck unit) into dense unit to reduce the dimension of input feature maps, reducing the computational effort while incorporating different dense unit features. DenseNet-B uses 1×1 convolution to reduce input feature map dimension of each layer to a fixed multiple of growth rate; DenseNet-C adds compression rate in transition layer to reduce output feature dimension of dense blocks; DenseNet-BC adds both bottleneck unit and variable compression rate transition layer.

Broaden DenseNet structure is multibranch improvement of network single layer, network structure block, and whole network using multibranch extraction function, including multipath structure, parallel dense block, parallel transition layer, and parallel network.

Lightweight DenseNet structure significantly reduces the number of network parameters and computation from two aspects: network model architecture design and network model compression. Network model architecture design which includes grouped convolution, depth-separable convolution, simplified dense block, one-time aggregation, and network model compression has automatic sparseness.

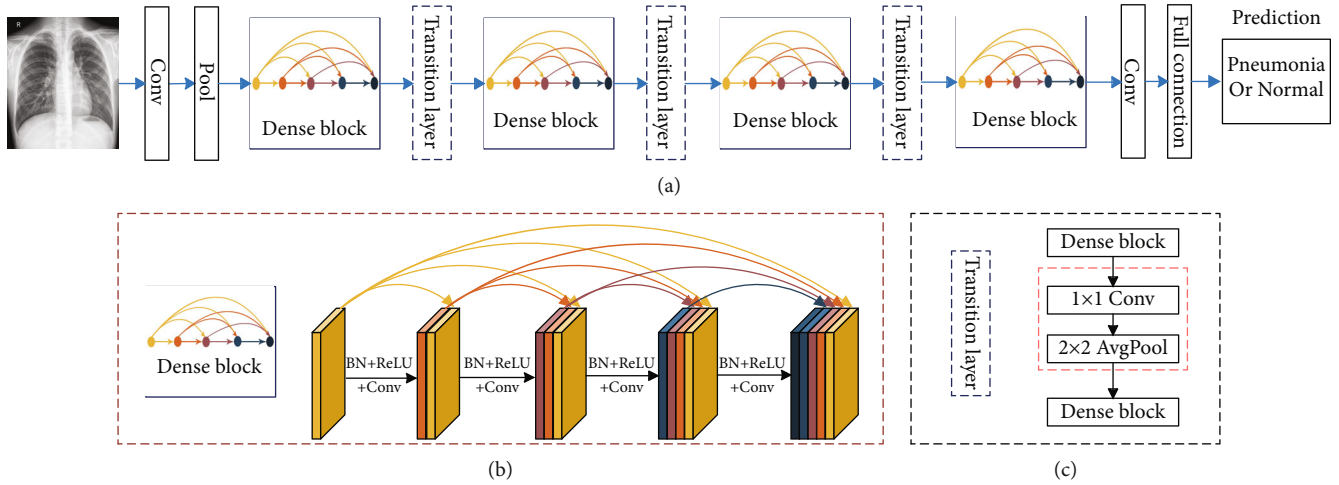


FIGURE 1: DenseNet structure. (a) The basic structure of DenseNet consists of dense block, transition layer, convolutional layer, and fully connected layer. (b) Denseblock consists of densely connected dense units with nonlinear mapping functions of BN, ReLU, and Conv, which are designed with preactivation strategy to make network training easier and generalization performance better. Dense unit input is spliced and merged with all outputs of the previous dense units, and new features generated also need to be passed to subsequent dense units, so that shallow features of dense block are repeatedly reused and effectively utilized, which can alleviate gradient disappearance to a certain extent, and a large number of features can be generated with a small number of convolution kernels; final DenseNet model is relaticely in scale. (c) Transition layer is the structure between adjacent dense blocks, which consists of 1×1 convolution and 2×2 average pooling layer, compressing dense block input and all extracted feature information, reducing feature map size and dimensionality, which can effectively reduce the number of dense block parameters and prevent network from overfitting. The fully connected layer is classification prediction layer, which reduces the influence of feature location on classification by integrating category feature information in network features, and classifies feature information after weighting.

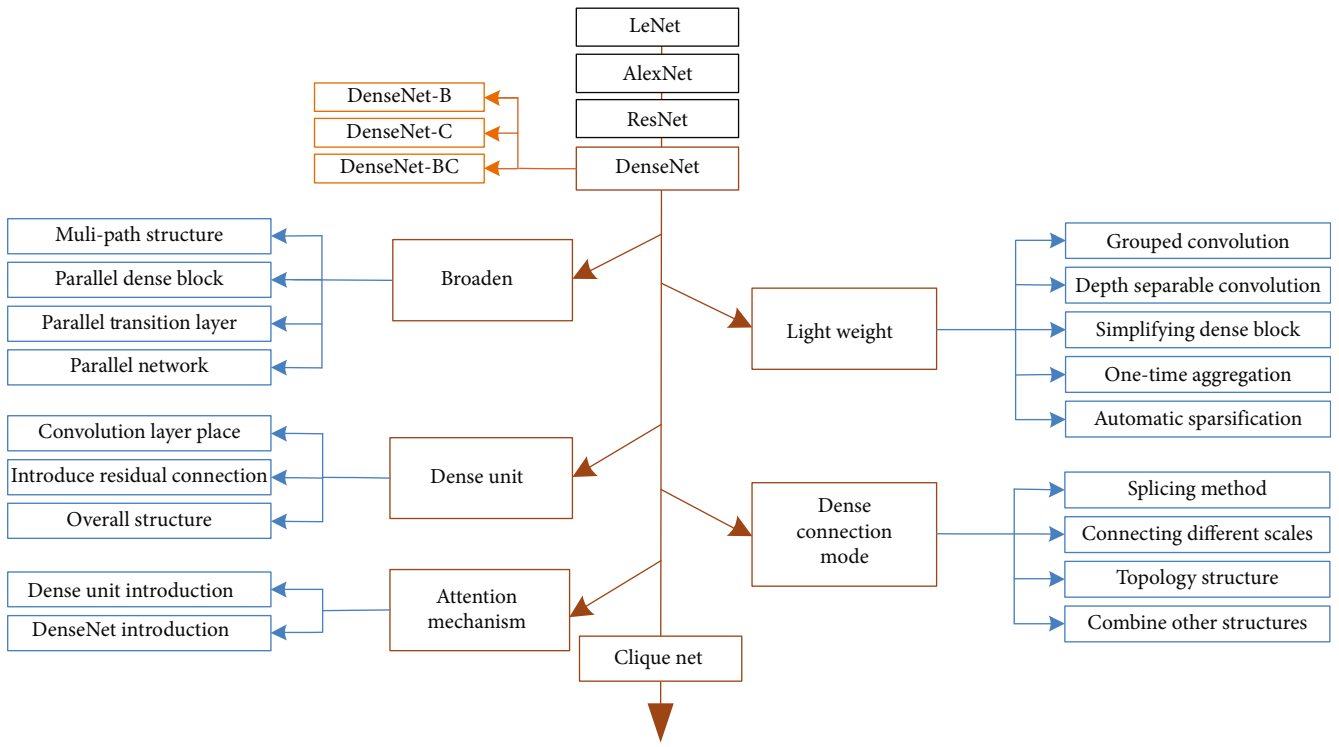


FIGURE 2: Development of DenseNet.

Dense unit is composite function improvement of DenseNet, optimizing feature extraction module dense unit for specific applications, summarized in three aspects: convolution layer placement, introduce residual connection, and overall structure.

Dense connection mode is a study of the way dense blocks are composed, dense connection improvement has four aspects: splicing method, connect dense unit of different scales, topological structure, and combine with other structure.

Attention mechanism is to emphasize or select important features in DenseNet and suppress some irrelevant detailed information, which can be divided into attention mechanism in dense unit and DenseNet.

Chen et al. [16] indicated that densely connected path of DenseNet increases linearly with increasing network depth, and the parameters increase dramatically; for this reason, Yang et al. [17] proposed CliqueNet, a convolutional neural network with alternating update mechanism, Clique block performs same dense propagation as DenseNet, and then, all layers except the top layer are used as the bottom layer for alternating updates, stitching feature maps of each layer into the output of Clique block, passing only fixed-dimensional feature maps, with forward and backward connections between any layer, cycling, and alternating updates to maximize the flow of information between layers. Final network uses fewer convolutional kernels to generate a large number of features, which greatly improves the parameter efficiency, and parameter propagation process generates multilevel features by cycling; experiments show that the second stage of Clique block helps to suppress noise and improve performance.

3.1. Broaden DenseNet Structure. Depth is not necessarily good, and width is not necessarily bad. On the one hand, in 2014, Ba and Caruana [18] pointed out that a well-trained deep network can usually find a suitable shallow network instead, and in 2019, Hanin and Rolnick [19] pointed out that expressiveness of network in practical applications is independent of network depth and only grows linearly with the number of neurons; on the other hand, deep learning models currently are trained by GPUs. GPU parallel processing makes it easier to train widened networks. Therefore, depth and width are not completely opposed to each other, increasing both depth and width is increasing the number of learnable parameters and improving network fitting ability, increasing depth can obtain a larger sensory field and capture more pixel-like features, and increasing width can obtain more subtle and richer features.

Broaden DenseNet structure is using multiple branches to improve network single layer, structure block, and whole, so that network learns more rich features per layer, such as texture features with different directions and frequencies. Although network deepening can extract knowledge layer by layer abstraction and continuous refinement, too narrow network can capture limited patterns per layer, and the network is no deeper to extract information. Broaden DenseNet structure approach is shown in Figure 3; there are multipath structure, parallel dense block, parallel transition layer, and parallel network.

First, multipath structure is an improvement of multipath feature extraction layer on single-path feature extraction, using multiple scales to achieve different perceptual fields, increasing the adaptability of network to features of different sizes and enriching features. Multipath convolution also uses the computational principle of decomposing sparse matrices into dense matrices to speed up convergence, aggregating output features with strong relevance, decomposing feature sets into multiple densely distributed subsets, aggregating features with strong relevance, and finally outputting less redundant information and faster convergence. In 2014, Google's Szegedy et al. [20] proposed typical multipath structure inception, which is parallel to commonly used 1×1, 3×3, 5×5 convolutions and 3×3 max-pooling to achieve different perceptual fields for extracting rich features and making classification judgments more accurate. Li et al. [21] proposed DenseNet-II for benign and malignant classification of mammogram images; the inception structure was used to improve adaptability of network to different scales and maintain the sparsity of network structure, while efficiently expanding network depth and width, resulting in better classification performance with good generalization and robustness. Arega and Bricq [22] used cascaded segmentation network FC-DenseNet for automatic segmentation of cardiac scars in multisequence cardiac MRI; similar inception structure captures relevant features such as scars and edema of different sizes, significantly improving segmentation results with the same computational overhead.

Second, parallel dense block is original single-branch dense block multiple times in parallel to increase network width, combining the benefits of network depth and width to improve model robustness; although deep networks have higher performance, they are limited in terms of gradient disappearance, reduced forward flow, and slower training time. Lodhi and Kang [23] proposed multipath-DenseNet with multiple dense blocks parallel learning to find valid paths from some or all of the blocks, which allows efficient gradient flow and achieves higher accuracy than other baselines using fewer parameters on the classification dataset. Chen et al. [24] proposed MFR-DenseNet with multipath feature recalibration, where SE module assembles multipath dense blocks from different channel feature maps to build interchannel feature correlation model to achieve strong feature extraction, which achieves high accuracy in classification while maintaining training stability.

Third, parallel transition layer is a branching supplement based on the single-branch transition layer to realize broadening. Liu et al. [25] proposed dense binary tree DenseBT-Net for lung nodule classification, which introduces central cropping operation into transition layer and parallelizes it to create indirect shortcuts between different dense blocks, preserving detail features of nodule morphology and filtering irrelevant background information; binary tree-like singular network model can generate rich multiscale features, making the parameter scale lighter and with higher parameter efficiency.

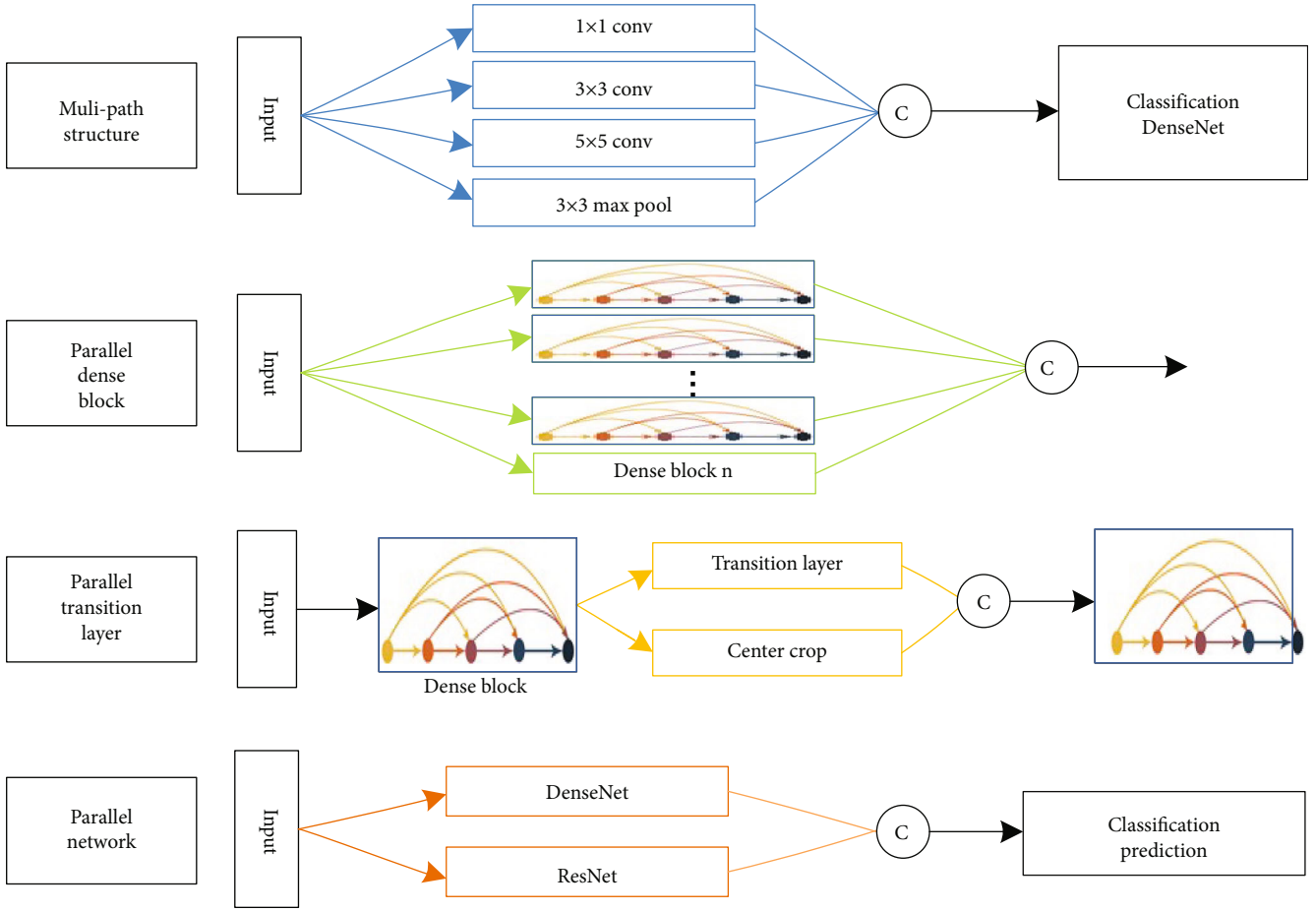


Figure 3: Broaden DenseNet structure.

Fourth, parallel network is multiple networks in parallel, and the fusion of multiple network features can ensure more stable and accurate prediction. Chen et al. [16] argued that ResNet is essentially DenseNet with shared connections, and ResNet implicitly has residual path reuse features, and dense connections that can be mined for new features, based on this joint perspective dual-channel network (DPN) were proposed, by sharing same features of network through dual-path structure; high accuracy is achieved with flexibility, small models, less computation, and low resources. Nguyen et al. [26] proposed MSDENSE-DAT to predict regions in crowd scenes that attract human attention, designed two-branch homography network based on DenseNet201 to extract multiscale features, extracted as many features as possible from original size and half-size images, and cascaded two-branch features along channel axis, and also designed self-attention blocks to emphasize interfeature correlation; MSDENSE-DAT extracts the best features with low density in the crowd.

3.2. Lightweight DenseNet Structure. Lightweight of CNN is to reduce the number of parameters and computation without degrading the performance of existing models and to achieve storage and computation on mobile platforms with limited storage space and power consumption. CNN is widely used in image analysis and has achieved great success, greatly reducing the number of parameters and computation through weight sharing and sparse connection; however, two aspects, model storage and model prediction speed, limit the application of CNN in mobile. Lightweight of CNN is important research in deep learning, and two main directions are studied: lightweight model architecture design and model compression. This section summarizes four improvements of lightweight DenseNet structure in model architecture design: grouped convolution, depth separable convolution, simplified dense block and one-time aggregation, and automatic sparseness of model compression, as shown in Figure 4.

First, grouped convolution is to group feature maps and then convolve them separately, the number of model parameters decreases as the number of groups decrease, and model structure and training are more efficient. In 2012, AlexNet [1] proposed grouped convolution and utilized dual GPU processing. To apply DenseNet to mobile devices with limited computational resources, Huang et al. [27] proposed CondenseNet based on grouped convolution, where the network is connected layer by layer, preserving dense connectivity of DenseNet, achieving feature reuse, sharing input features using self-learning grouped convolution modules, eliminating layer-to-layer connections, and avoiding

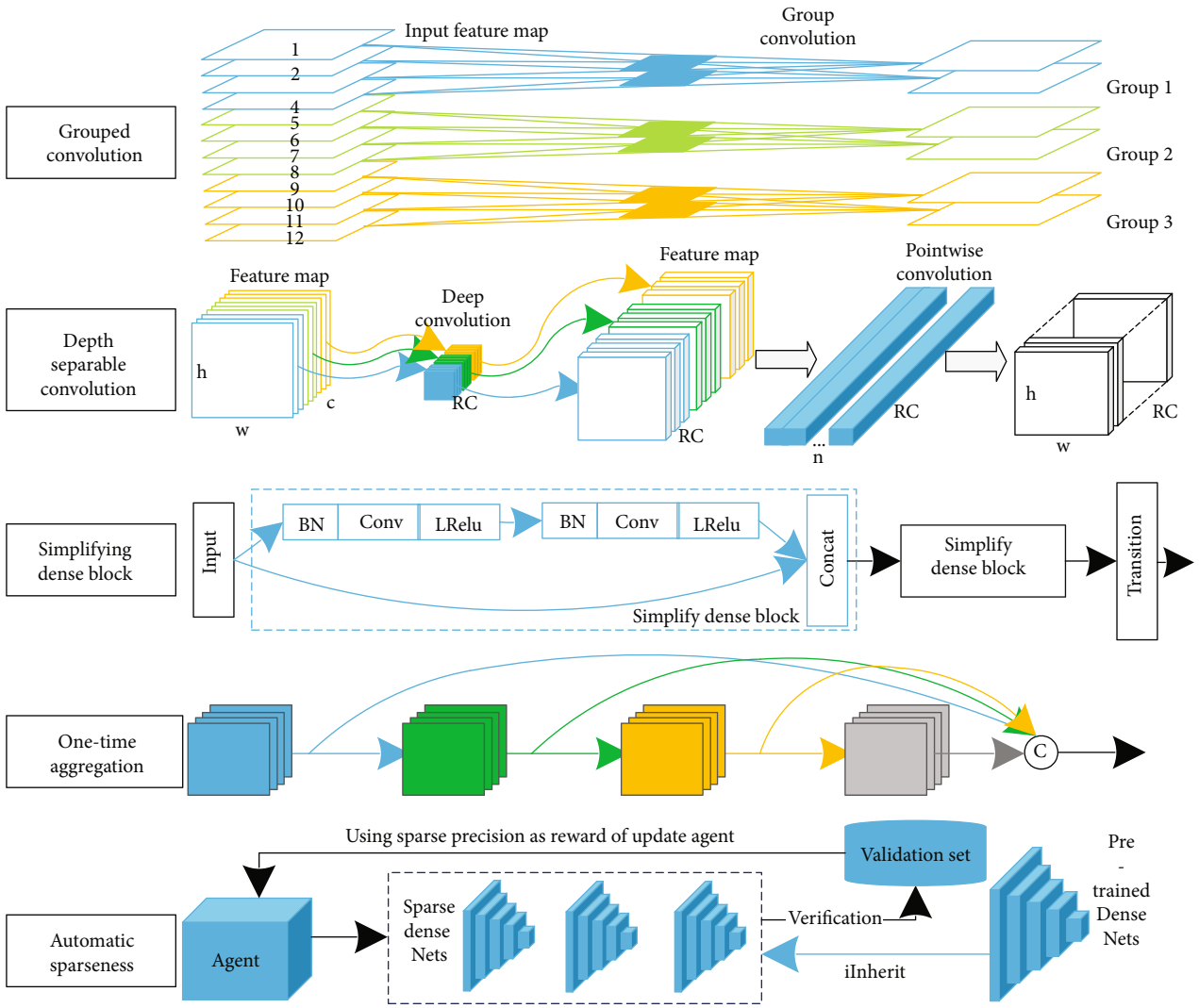


Figure 4: Lightweight DenseNet structure.

accuracy branching process that affects too much; CondenseNet uses only 1/10th of computation at an accuracy level comparable to that of DenseNet. Li et al. [28] proposed Dense2Net with efficient multilevel group convolution, using channel mixing to rearrange features, tiling features into multiple groups and randomly receiving the previous group of features, while fusing all groups of different scales into output using 1×1 convolution to improve parameter efficiency and information transfer; Dense2Net has higher accuracy and better information transfer than DenseNet.

Second, depth separable convolution (DWSC) is a decomposition of the convolution strategy into spatial convolution followed by channel convolution, replacing convolution layer to achieve a lightweight network, using a filter for each input channel in spatial directional convolution (DWC), and applying 1×1 point-by-point convolution (PWC) on the channel side to linearly combine the depth directional convolution outputs. MobileNet [29] uses depth separable convolution to build a lightweight deep network that greatly reduces the computational effort with a small reduction in accuracy, achieves a balance between model efficiency and classification accuracy, and performs well in task such as target detection, fine classification, and face attribute. Chen et al. [30] proposed Mobile-DANet, an attention-embedded lightweight network for corn disease recognition, using depth-separable convolution instead of convolutional layers in dense blocks and depth convolution using input mapping filters to achieve channel-by-channel convolution and combining channel-by-channel results through 1×1 convolution, while using spatial and channel attention mechanism to detect and highlight targets in multifeature map, with the final model featuring small size, high accuracy, and excellent performance.

Third, simplified dense block is to replace original dense block with multiple simplified dense blocks in series, fixing the number of dense units to be smaller and multiple simplified dense blocks in series to replace original dense blocks, reducing output feature map dimension while maintaining

feature reuse. s-DenseNet was proposed by Yu et al. [31] for pathological grading of breast cancer on large-size mammograms, using fewer dense units to prevent network overfitting, extracting high-level semantic features, and combining logistic regression to select image histology feature values, the trained network achieves high accuracy at lower complexity. Huang et al. [32] designed a lightweight neural network using simplified dense block consisting of two dense units, stitching adjacent blocks of mammogram and osteosarcoma histological images as input; feature vectors were mapped to simplified dense block for kernel learning; final network was easier to train and better in terms of classification accuracy, sensitivity, and specificity.

Fourth, one-time aggregation is to aggregate outputs of all dense units to the end of dense block, rather than two-by-two interconnection of dense units. Lightweight network design considers two main factors, model size and FLOPs, to reduce inference time and energy consumption; memory access cost (MAC) and GPU computational efficiency are also considered. Densely connected aggregations of intermediate features with different sensory fields can produce powerful features using a small number of parameters and triggers, but dense connections can lead to a large memory access overheads, and 1×1 convolution using a small number of operations can also lead to inefficient GPU computation. To address the slower and inefficient DenseNet detector, a fast and efficient architecture VoVNet consisting of one-time aggregation (OSA) was proposed by Lee et al. [33], which reduces memory access cost and performs efficient computation by using diverse feature representations with multiple reception rates; dense aggregation is aggregating all previous features to subsequent layers; while OSA is aggregating all features at once to final unit, it also removes bottleneck unit and keeps intermediate layer input constant, greatly improving MAC and GPU efficiency while maintaining cascade strength; VoVNet applied to target detector achieves twice detection speed, and energy consumption is 1.6×–4.1× lower than DenseNet.

Fifth, automatic sparsification method for model compression to achieve lightweight is to prune DenseNet by reinforcement learning. Li et al. [34] proposed automatic sparsification (ADS) method to prune redundant jump connections in DenseNet and ADS compression based on reinforcement learning (RL) method to prune dense connections representation by the original adjacency matrix; sparse DenseNet verification accuracy is used as a reward and punishment for updating RL agents, and high-performance sparse networks are generated by iterating RL agents with fewer parameters, less computation, and portability after sparse.

3.3. Dense Unit. DenseNet uses dense cells interconnected two by two to form dense blocks; this section summarizes three aspects of convolution layer placement, and introduces residual connection and overall structure.

3.3.1. Convolution Layer Placement. Dense unit of DenseNet uses preactivation structure, in which BN and ReLU layers achieved preactivation before convolution, and also uses the structure of 1×1 convolution followed by 3×3 convolution. Dense unit is improved by placement of convolution layer, which can be divided into preactivation and nonpreactivation, where ReLU and BN can interchange their positions, and the specific improved form is shown in Figure 5.

First, convolutional layer is easier to train and has better generalization performance after BN algorithm and preactivation structure of activation function. Zhu et al. [35] achieved image superresolution using dense jump-connected network, where dense unit provides large sensory fields using 3×3 convolution, and then, 1×1 convolution deepens network to learn robust feature representations, and small filters with a small number of parameters are used to deepen network, making full use of hierarchical feature deepening structure, and network outperforms new algorithm with robust fitting capability. Wei and Liu [36] introduced dense blocks in SSD network to detect hazardous substances in X-ray images. Dense unit connects two 1×1 convolutions in series to avoid destruction of learned feature map region information, and the method has good transfer learning capability.

Second, convolutional layer does not use preactivated dense units before BN and activation function. Wei et al. [37] added dense blocks of first convolution in YoloV3 backbone, and transition layer also used dense units in parallel with maximum pooling layer, and the model was effective in multiobjective defect detection of railroad routes. Lei et al. [38] used densely connected dilated convolution for codec connection, and dense blocks were used for first convolution of dilated dense unit, which can enhance effective feature transfer and retain finer structural information for accurate segmentation of skin lesions.

Third, convolutional layer distributes BN-normalized data to the nonsaturated region before activation function and BN, and ReLU can then control the degree of saturation of activation; if the order of BN and ReLU is switched, ReLU will lead to deactivation of some of BN neuron, which causes BN instability and affects the model performance, practically different application scenarios have different role in the order. Hasan et al. [39] proposed an automatic semantic segmentation network (DSnet) for skin lesion segmentation, in which the encoder uses dense blocks and transition layers to avoid learning redundant features and uses BN of dense units after activation function, and DSnet outperforms existing baseline segmentation networks in several metrics.

Fourth, convolutional layer precedes activation function, and BN is more suitable for denoising, but it leads to slow convergence in superresolution and large fluctuation in the loss function, which also increases the memory and computational burden. Mustafa et al. [40] proposed multilevel dense MLDNet for multifocus image fusion (MFIF), which uses parallel dense blocks to extract multilevel local dense and global features in the source image; dense units are connected using convolution followed by ReLU to retain as many dense features as possible, and MLDNet has significant performance improvement over the latest methods. Li et al. [41] proposed multiresolution dense network to connect mutually different structures dense units densely, with different size convolution for feature extraction of three central consistent lung field blocks with different resolutions,

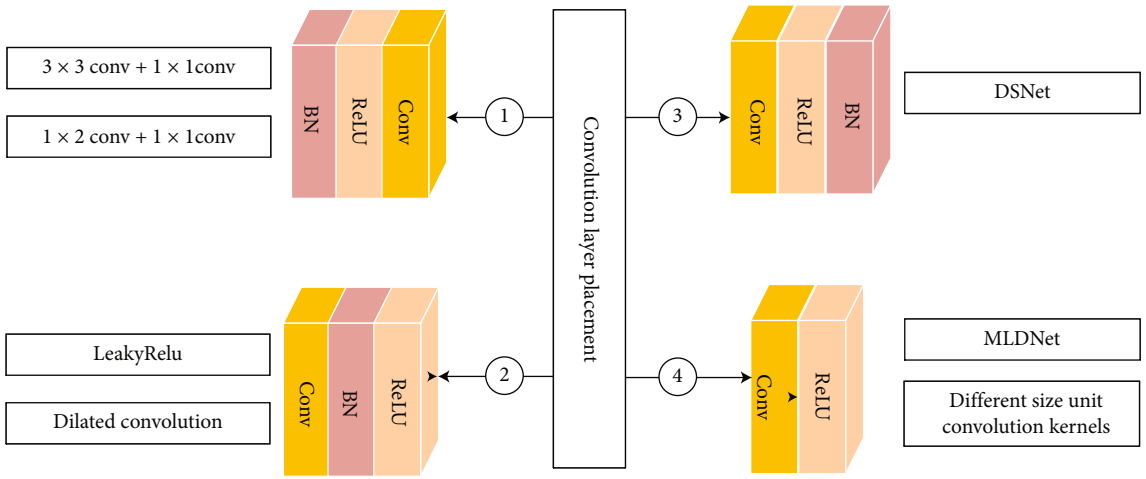


Figure 5: Convolution layer placement.

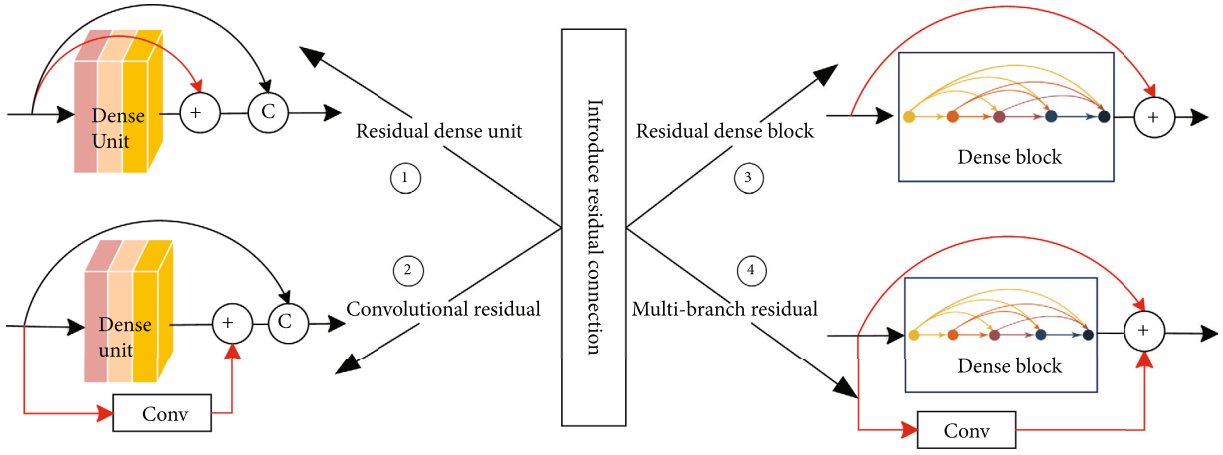


Figure 6: Introduce residual connection.

and fusion to detect lung nodules, which outperformed general JSRT database for radiologists by up to 99%.

3.3.2. Introduce Residual Connection. Residual connection uses skip connection to connect shallow layers of information directly to deep layers, and output is expressed as a linear superposition of input and its nonlinear transformations, using a new expression to achieve a constant mapping between layers. Deep learning relies on error chain derivative back propagation for parameter update; too small derivatives will become smaller and smaller gradients after multiple series multiplications, leading to gradient dispersion; residual connection adds constant terms to derivatives, so that error signal is propagated directly to the shallow layer without intermediate weight matrix changes, which alleviates the problem of gradient dispersion to a certain extent, and information is propagated back and forth more smoothly. Introduction of residual connection in DenseNet can solve network degradation problem, thus deepening network model and improving performance. Four structures are introduced for residual connectivity: residual dense unit, convolutional residual dense unit, residual dense block, and multibranch residual dense block, as shown in Figure 6.

First, residual dense unit is residual connection between inputs and outputs of dense unit, Feng et al. [42] proposed recurrent network DCRN to cope with single-image superresolution (SISR) task, which performs local residual learning within dense block and dense unit to extract semantic features, and experiments on Set5, Set14, BSD100, and Urban100 datasets to demonstrate DCRN efficiency and model component effectiveness.

Second, convolutional residual dense unit is a residual connection using convolutional improvements to prevent gradient disappearance or explosion problems while speeding up convergence. Yang et al. [43] introduced convolutional residual in DenseNet by parallelizing multiple multiresolution convolutional residuals in dense block, effectively reducing the number of superresolution network parameters, and increasing model hierarchy and network depth, with prediction speed reaches 25 ms per image under certain accuracy loss. Single-branch convolutional residual block can only extract single-level semantic information;

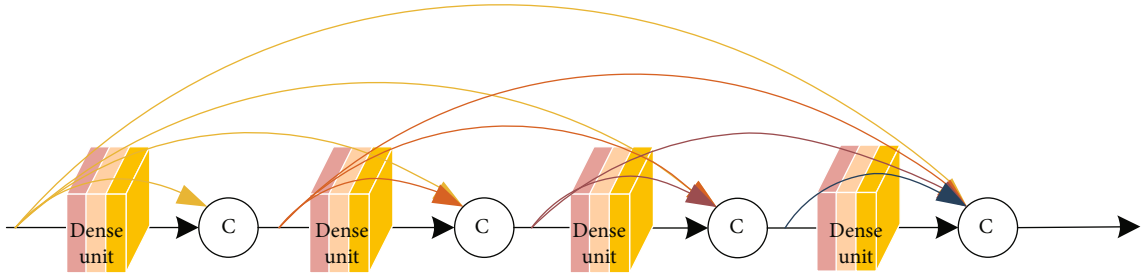


Figure 7: Dense connection mode.

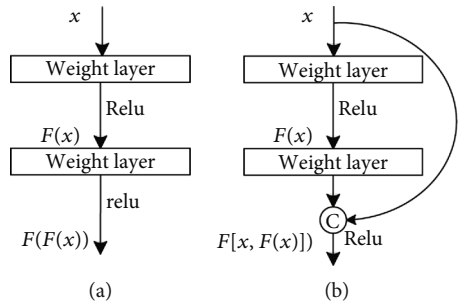


Figure 8: (a) Traditional CNN unit mode and (b) DenseNet unit mode. (a) Input gradient information in conventional unit module goes through many layers. (b) DenseNet unit module uses dense jump connection to splice input features with learned features as subsequent layer inputs thus achieving feature reuse, which is easy to train and more parameter efficient in DenseNet. Specific structural design details can be expressed in following equation with some notations defined in advance: input picture is X_0, l is index of neural network layers, $H(\bullet)$ indicates nonlinear operations combination, and $H(\bullet)$ in DenseNet refers to composite function consisting of BN, ReLU, and Conv.

Wang et al. [44] designed multiscale residual network (EMRN) in image superresolution task and densely connected multiscale residual blocks extracted image hierarchical features to achieve multilevel semantic information in different perceptual domains.

Third, residual dense block is dense block using residual connection between inputs and outputs, and residual dense block inherits advantages of ResNet and DenseNet, improves information flow of network, and retains more structural information. Wu et al. [45] proposed dense pyramid residual network (DPRnet) for restoring fog-free images, residual dense block extracts network features and quadruple downsampling features, global residual connection is applied to whole network, allowing the feature maps to be heavily reused and deepening network at different scales, and ultimately, DPRnet has better performance.

Fourth, multibranch residual dense block is dense block that uses both residual connection and convolutional residual connection, and multibranch residual structure improves diversity of feature extraction for hierarchical feature extraction. Long et al. [46] proposed aggregated residual dense network (RXDNFuse), which fuses residuals and convolutional residuals from infrared and visible images into parallel dense blocks to extract hierarchical features, and qualitative evaluation showed that rich texture details and prominent thermal radiation information were effectively retained.

3.3.3. Overall Structure of Dense Unit. DenseNet is connected with dense units in a dense connection, and overall structure of dense unit is replaced with whole structure; five overall structure improvements are as follows.

First, multirate extended convolutional structure MDCS replaces dense unit, 1×1 convolution, and four different extended rate 3×3 convolutions to perform input feature mapping in parallel, which was used in training DenseNet by Zhang et al. [47], using original image and marker point set, adding MDCS to increase network width laterally to fully utilize multiscale vascular spatial feature information at lower computational cost, and dense connection alleviates gradient disappearance and accelerates network, the average dice score reached 93.20% with good integrity and sensitivity for cerebral vessel segmentation, especially for slender vessels. Second, Wang et al. [48] proposed dual residual attention model (DRAM) for SISR task, in which dense units tandem two 3×3 convolutions capture more global information, channel attention features and convolutional layer features share information in tandem, spatial attention features and convolutional path features are fused, and dual feature fusion enables the model to capture important high-frequency components; the superiority of model in benchmark dataset was verified. Third, Zeng et al. [49] used densely connected cascade network DCCN for MRI reconstruction of K-space undersampled data, using dense unit tandem residual dense block (RDB) and data consistency layer (DC), with RDB combining dense connection and residual learning to fully utilize features of different layers, and DC used data consistency to improve MRI image reconstruction under different sampling trajectories and sampling

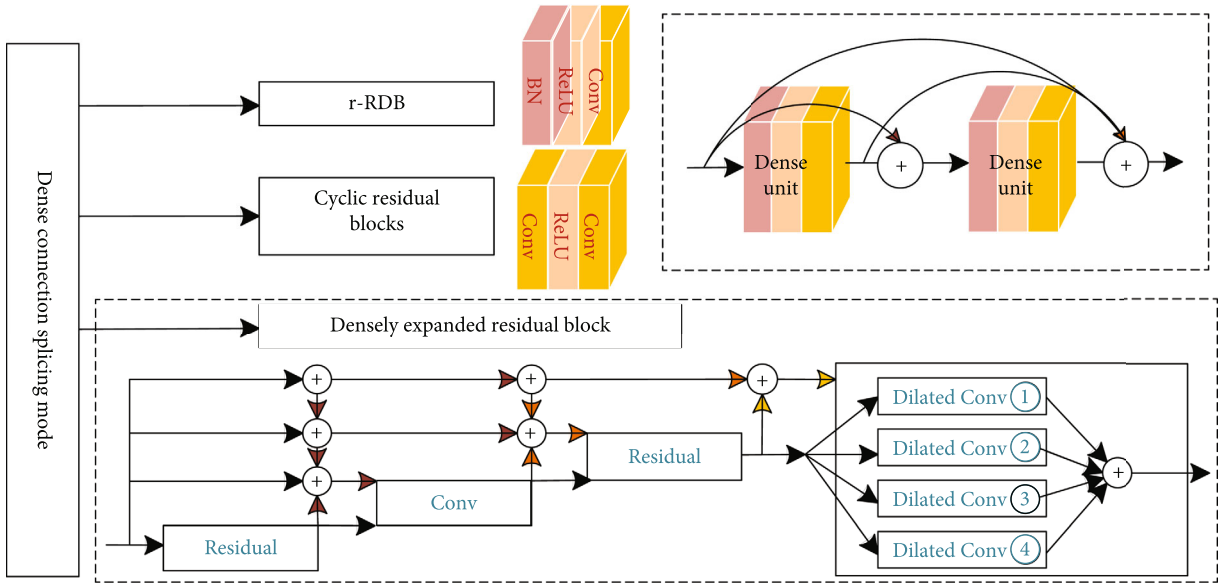


Figure 9: Dense connection splicing mode.

rates. Fourth, Zhang et al. [50] used densely connected inception-Res module for segmentation of medical images, with branches of different kernel sizes aggregating feature maps, widening network to learn more features, and residual connection referencing input feature mapping learning function to make the module easier to learn; final model achieved better performance on lung and brain tumor segmentation datasets. Fifth, Qamar et al. [51] proposed encoder dense block with residual connection for skin lesion segmentation architecture, using residual connection to obtain more robust features, dense unit also uses expanded space pyramid pools (ASPP) to efficiently capture global multiscale contextual information while maintaining network parameters, ASPP also aggregates image space information, and the results of study showed that architecture achieved the most advanced performance.

3.4. Dense Connection Mode. Dense connection mode is two-by-two interconnection of all dense units in same dense block, and the feature maps of same scale are connected using channel stitching, as shown in Figure 7.

Dense connection mode [9] maximums information flow and reduces the parameters, while alleviating network gradient disappearance and overfitting phenomena, dense connection has following features: first, dense connection uses output as input to subsequent layers, enhancing feature transfer and using features more efficiently; second, dense unit in dense connection is narrower, adding only a small number of feature maps to network each time without relearning redundant feature maps, and final classifier makes predictions based on all features, maintaining network performance while reducing the number of parameters; third, each layer in dense connection directly uses gradient of loss function and initial input information, which is equivalent to invisible depth supervision and helps alleviate gradient disappearance; fourth, dense connection also has a regularization effect, which has a suppressive effect on overfitting, and network also can mitigate overfitting with fewer parameters. Dense connection improvement has four aspects: splicing method, connect dense unit of different scales, topological structure, and combine with other structure.

3.4.1. Splicing Method. Current unit input in dense connection is stitched from previous unit feature maps, and stitching method is to perform channel stitching at same feature map size. Traditional convolutional neural network unit module is different from dense connection neural network unit module, as shown in Figure 8.

Output of the first layer of conventional neural network can be expressed as

$$X_l = H(X_{l-1}). \tag{1}$$

DenseNet splices different layer features, output of the first layer is expressed as:

$$X_l = H([X_0, X_1, \cdots \cdots, X_{l-1}]), \tag{2}$$

where $[X_0, X_1, \cdots \cdots, X_{l-1}]$ represents $0 \sim (l-1)$ layer feature map splicing; dense connection mode is shown in Figure 9.

First is residual dense connection mode for channel fusion. Zhang et al. [52] proposed improved residual dense blocks (r-RDB) in image string and handwritten character recognition tasks, maintaining RDB local feature fusion and local residual learning capabilities, cascading multiple r-RDBs to form global dense block using dense connection of residual summation, and adaptively learning global dense residual features to ensure interblock information flow, r-RDB feature fusion, and residual learning refine block structure and reduce inner computational cost. Adegun and Viriri [53] proposed DenseNet framework for cyclic residual block classification based on dense block, where dense connection within cyclic residual block is combined using

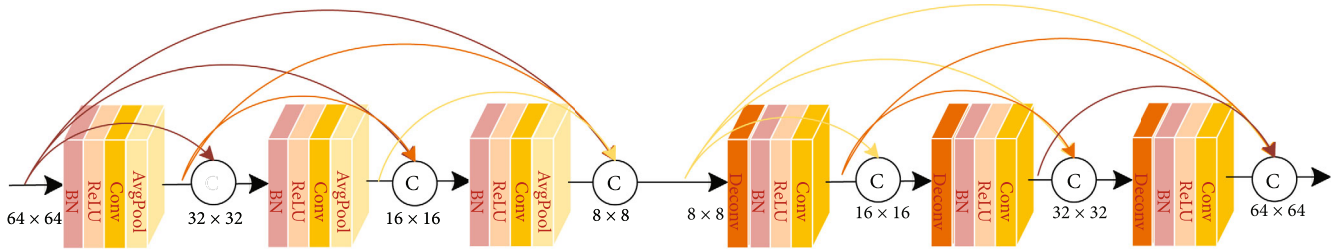


Figure 10: Connect dense units of different scales [55].

summation operations to reduce computational effort, contour refinement, and localization of lesion boundaries for edge potential; results showed that effective preprocessing and segmentation of skin lesions can improve classification performance.

Second, dense-expanded residual block improves dense connectivity as a method for multiplexed residual fusion, Hong et al. [54] proposed an end-to-end network MMCL-Net for simultaneous detection, segmentation, and classification of spinal structures, establishing spatial dynamic connectivity for three structures and three tasks, dense connectivity for multilevel aggregation of expanded residual unit to extract important radiological features, and dense connectivity as residual mapping; shared learning among multiple tasks represents adaptive optimization of model, and densely expanded residual module outputs are received in parallel by dilated convolutional layers with different expansion rates; MMCL has been validated on MR images of more than 200 clinical patients, and it can accurately identify structurally relevant lesions in spine and show superiority in grading lumbar spinal neural foraminal stenosis.

3.4.2. Connect Dense Units of Different Scales. Dense connection in DenseNet is to connect same scale dense unit feature maps, and connecting dense units of different scale is shown in Figure 10.

Dense connection is used between different scale structures of encoders or decoders. Wang et al. [56] used dense connection within encoder and decoder networks, average pooling between convolutional layers to eliminate redundant features, and dense connectivity of units with different feature sizes to adequately extract biomarkers associated with recurrence of high-grade plasmacytic ovarian cancer from CT data; the model provides a new prognostic analysis approach. Pemasiri et al. [57] used multimodal semantic segmentation model to segment human body parts in visible and X-rays, where densely connected features were adjusted to dense units of same size, and features of different modalities were integrated using channel splicing to achieve single model for multimodal semantic segmentation. There is only one set of cascade layers between U-Net codec blocks, and Jun et al. [58] designed nested codec architecture (T-Net) with pooling and upsampling inserted in codec blocks, feature mappings interconnected, encoder low-level and high-level features connected to whole decoder, and feature mappings at different scales to produce complex semantic segmentation, and experiments showed that Dice similarity coefficient of T-Net reached 83.77%, which is 10.69% higher than that of U-Net.

3.4.3. Topological Structure. Conventional L-layer convolutional neural network has L connections, while DenseNet has $L(L+1)/2$ connections, densely connected Lth layer has L inputs, which are feature maps of its predecessors, and its feature maps also need to be passed to all subsequent layers, as shown in Figure 11. Due to this two-by-two interconnected connection, DenseNet has constant mapping, deep supervision, and deep diversity characteristics, densely connected approach has a simple internal representation and can reduce redundant features.

First, dense connection topological structure is to aggregate outputs of all dense unit to the end of dense block. In superresolution (SR) domain, deeper network helps to improve image reconstruction quality and introduce residual connection to avoid gradient disappearance, but deep network is computationally intensive and inefficient, Yang et al. [59] used a new dense connection approach to extract low-resolution image features; model topology is to merge outputs of unit into the end of module and obtain deeper network and richer feature map by reducing the number of channels and network parameters; model was experimentally shown to be effective in SR with different magnifications. Mostefa et al. [60] proposed DenseMultiOCM for extracting MRI image features in brain tumor segmentation, improved dense connection by connecting only unit output to final output of dense block, dense units in series to two 3×3 convolutions, and final unit used pooling and upsampling to extract half-scale features; experiments showed model segmentation results were improved.

Second, residual connection is introduced in dense connection mode, two methods of SR are to learn features in high-resolution (HR) and low-resolution (LR) image space, and HR method is to interpolate LR information into HR, but it increases computational complexity and may introduce additional noise; LR method is to extract features and then increase spatial resolution by deconvolution, which is less computationally intensive, but difficult to learn multiscale nonlinear LR-HR mapping. Both methods learn mapping relationships in single-scale image space and cannot provide SR information across different scale of features, and although there is a Laplace pyramid structure to progressively upsample LR information to generate multiscale features, it also uses only highest-level features and ignore low-level features, Zhou et al. [61] designed dense convolutional autoencoding

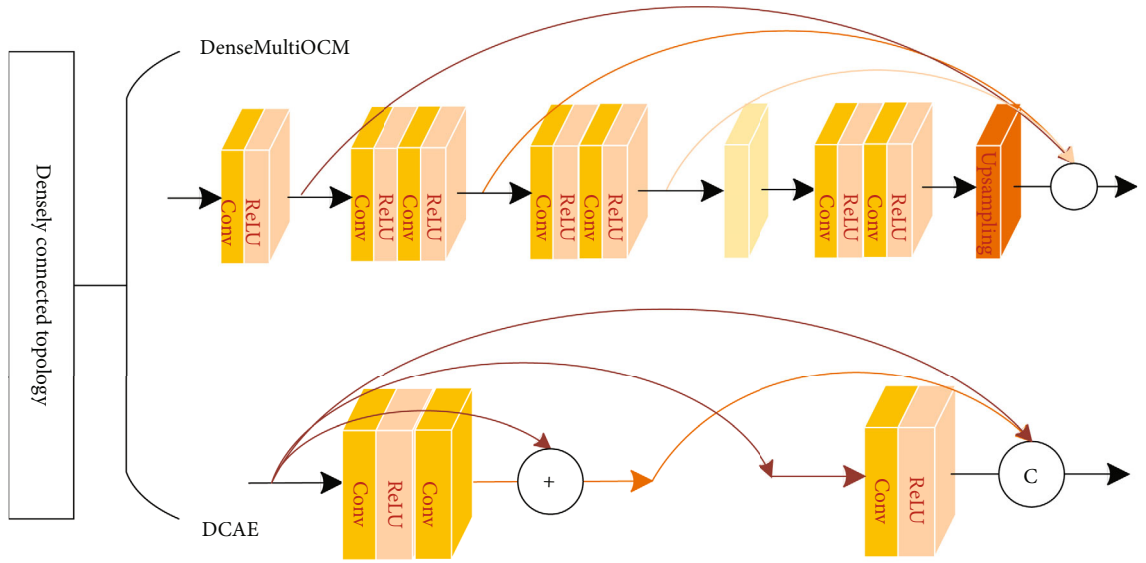


Figure 11: Densely topological structure.

(DCAE) block to extract multispatial scale features with different spatial resolution and temporal scales to establish multilevel feature reuse mechanism, and DCAE units were stitched with dense concatenation and jump concatenation to achieve long-term temporal feature reuse, and benchmark evaluations showed that this method outperforms existing methods, especially for structured images.

3.4.4. Combine with Other Structures. Dense connectivity method connects other structures with the following six densely connected structures.

First, Signoroni et al. [62] proposed BS-Net for predicting pneumonia in chest X-ray images, segmenting and predicting disease severity with dense connection to encode stage feature maps and also using densely connected deconvolution layers instead of code and decode jump connections, and model exceeded radiologists' scoring accuracy and consistency on the public COVID-19 dataset. Second, deep learning rarely focuses on both accuracy and speed in medical image segmentation and extends to mobile environments, where multiple downsampling loses significant information, Tseng et al. [63] densely connected U-nets at different scales to form dense blocks, dense connection improves feature reuse and preserves features at each step, and deconvolution enables upsampling to retain as much information as possible; DNetUnet achieves better segmentation performance and can be a mobile platform. Third, Liu et al. [64] designed DRN-DCMB model for reducing motion artifacts in brain MRI, where dense units of multiresolution block consist of U-net-like structure; four resolution features of U-net are convolved and then stitched together, where inputs and outputs of residual connectivity model predict image artifacts; learning residual mapping can significantly reduce motion artifacts and maintain image contrast, while improving image quality and sharpness. Fourth, Wang et al. [65] proposed multipath connectivity network (MCNet) for medical segmentation, where the encoder extracts different size features using multiscale feature blocks and densely connects same size feature maps in codec, and rich contextual information improves segmentation effect; MCNet has strong robustness and better performance for different scale targets. Fifth, for performance improvement, network structure improves from simple to complex, but simple stacking of capsule layers does not improve performance; too much stacking will lead to too small coupling coefficients and inhibit gradient, capsule network dynamic routing algorithm requires a lot of computation and storage costs, parameter increase will also lead to serious overfitting, so reasonable architecture is the key of deep capsule network, and lightweight of deep capsule network is also important research direction; Sun et al. [66] designed structure based on multilevel separable convolution and dense capsule layers, with smaller performance loss while effectively reducing parameters, dynamic road algorithm is lightweight with separable matrix, densely connected capsule layers increase network depth and complexity, competitive performance can be obtained with fewer parameters, and model is quantitatively demonstrated to have optimal parameter utilization through accuracy and parameters number ratios. Sixth, variation in image scale limits performance of target detection and semantic segmentation models to some extent, and crowd counting domain is prone to significant variation in crowd image size, leading to more obvious model performance limitations, Wang et al. [67] proposed single-column scale-invariant network (ScSiNet) for crowd counting with densely connected scale-invariant (SiT) layers forming dense block and SiT using different feature channel groups to encode different receiver domain images and fusing fine-grained multiscale information, using dense connection for multilayer feature interlayer fusion, and model combines intralayer scale aggregation with interlayer multiscale to obtain more accurate performance and computational precision with good metastability and scale invariance.

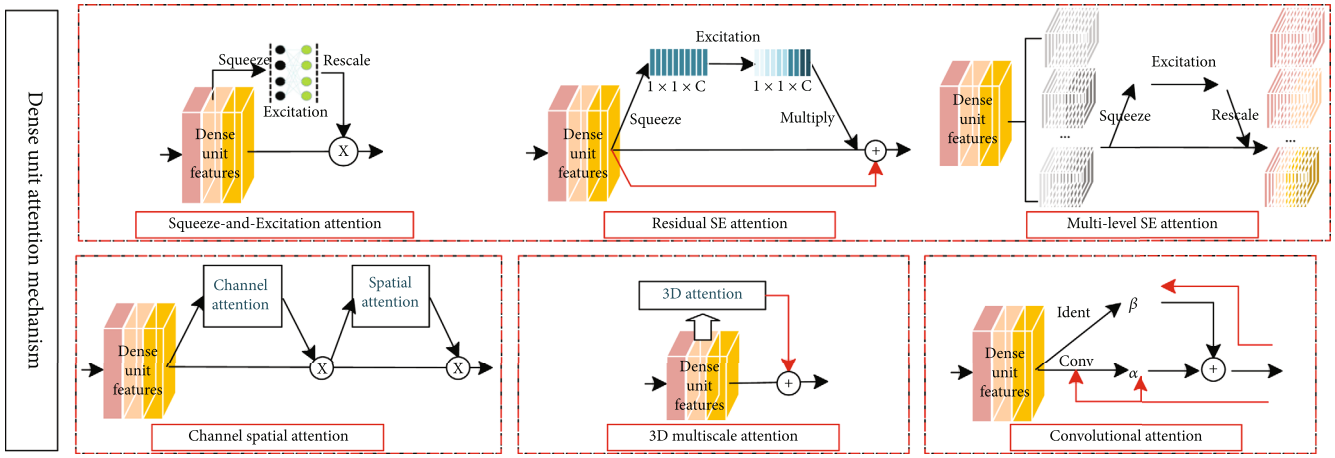


Figure 12: Attention mechanism in dense unit.

3.5. Attention Mechanism. Attention mechanism [68] is introduced into codec framework, which makes it easier to learn the intermapping relationships between multiple contents and different modal data; thus, allowing better representation and overcoming its uninterpretable and hard to design drawbacks, attention mechanism does not require supervised signals and appears to be extremely effective for problems with little a priori cognition. Attention module reduces computational burden of high-dimensional data by structurally selecting a subset of input and improves the quality of output by "eliminate false and retain true," allowing model to focus more on useful information in input that is significantly relevant to current output. Introduce attention mechanism into dense unit, and DenseNet is based on three main considerations: first, allow model to achieve better performance in text image analysis tasks; second, attention mechanism can calculate weights for each input and improve the interpretability of network model operates way [69]; third, it can alleviate some defects in recurrent neural networks, such as performance degradation due to increasing sequence input length and computational inefficiency due to input ordering.

3.5.1. Attention Mechanism in Dense Unit. Dense unit introduces attention mechanisms and mainly squeezes excitation (SE), residual SE, multilevel SE, channel spatial, 3D multiscale, and convolutional attention mechanism, as shown in Figure 12.

First, SE attention mechanism is to increase skip connection of attention mechanism in adjacent layers, focuses on interchannel relationships and learns the importance of different channel features, and improves interchannel information interaction, thus improving network performance. Feng et al. [42] proposed densely connected recurrent network (DCRN) for single-image superresolution tasks, introducing SE for dense unit and dense block features, rescaling and extracting contextual information, and global pooling of SE attention mechanism to extract global structural information, and fully connected layer emphasizes information channels and suppresses useless channels to improve model representation and reconstructed image consistency. Huang e al. [55] proposed a preprocessing method to optimize diagnostic performance of meningioma or glioma and narrow intraclass gap, including multidirectional brain region extraction (MDBRE) method and bimodal iterative gamma correction, and combined SE and DenseNet classifier to extract brain regions; the accuracy of triple classification normalized luminance images was improved by 9.7%.

Second, residual SE attention mechanism is to add residual connection to SE branch. Shi et al. [70] combined dense connection with residual SE to propose myopic eye detection network (MDNet), which introduced residual SE correction for dense unit feature mapping channels, reduced the number of parameters, and improved network performance; network detected spherical equivalent mean absolute error up to 1.1150 d (Diopter); feasibility and applicability were verified in fundus images.

Third, multilevel SE attention mechanism is introduced after feature map grouping; personnel reidentification task is critical to capture appearance changes in different viewpoints; localization and alignment of body part are inaccurate, and it is difficult for CNN to capture distinguished information to generate robust representation of pedestrians, for which Yan et al. [71] based on interdependence between different channels and horizontal feature maps, using spatial attention mechanism to learn distinguished local features and construct multilevel SE attention blocks for different subnets, adaptively discover body parts and recalibrate features to solve misalignment problems caused by inaccurate pedestrian detectors.

Fourth, channel spatial attention mechanism is introduced in dense unit, and tandem channel attention and spatial attention constitute convolutional block attention module (CBAM), which can be widely used to improve convolutional neural network representations. Xue et al. [72] proposed dual-attention dense ATP-DenseNet for handwritten font recognition, in which CBAM channel attention aggregates spatial features of average and maximum pools; channel attention map is computed using shared multilayer perceptron; operations along channel axis can effectively highlight information regions; average and maximum pools process features to map channel features and generate a

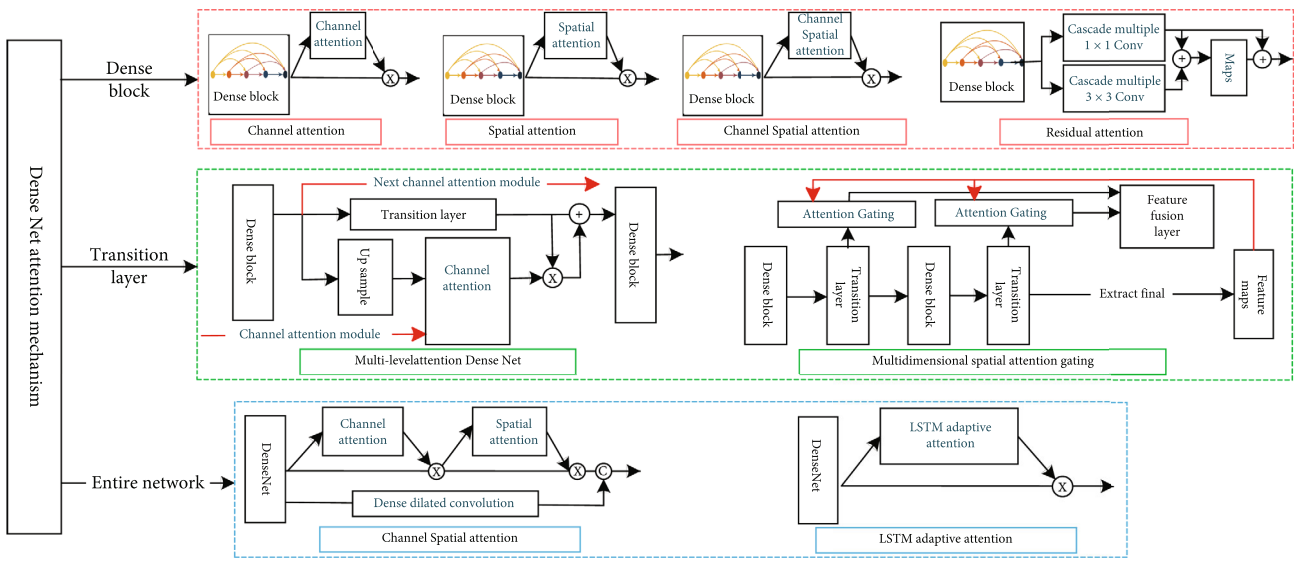


Figure 13: Attention mechanism in DenseNet.

two-dimensional spatial attention map in tandem; CBAM is introduced into dense unit to facilitate interlayer information flow and refine features across layers, while target-scale features are given more attention.

Fifth, 3D multiscale attention mechanism is introduced in dense unit. Zhang et al. [73] used 3D densely connected convolutional neural network (CAM-CNN) to extract brain MRI multilevel features for classification of Alzheimer's disease and mild cognitive impairment, densely connected difference at different unit levels; 3D dense unit introduces attention mechanism to generate attention maps and sum transformed MRI hierarchical data into more compact high-level; model has high classification prediction accuracy, and classification performance is at the highest level.

Sixth, convolutional attention mechanism is introduced in dense unit, considers interlayer relationships, and makes full use of interlayer information in convolutional and skip connection. In superresolution tasks, the main difference between high-resolution and low-resolution images lies in edges and textures, which are mainly based on high-frequency feature information to guide image texture recovery, residual superresolution networks have similar convolution and skip processing, resulting in some high-frequency information being missed in network. Bai et al. [74] proposed dense convolutional attention networks (DCAN) to introduce convolutional attention mechanism, which introduces weighted skip connection in convolution, considering interdependence between layers and adaptively calibrating convolutional features to learn richer details and sharper edges by crossing layers; DCAN spends less time to obtain superior performance in both quantity and quality.

3.5.2. Attention Mechanism in DenseNet. DenseNet mainly consists of dense block and transition layer, introducing attention mechanism to enhance feature propagation and reduce the number of network parameters, and introducing attention mechanism in DenseNet is shown in Figure 13.

(1) Attention Mechanism Introduced in Dense Block. Attention mechanism introduced in dense block has channel, spatial, channel and spatial, and residual attention mechanisms.

First is attention mechanism in dense block for feature enhancement. Zhou et al. [75] proposed 3D SE-DenseNet for dynamically enhancing hepatocellular carcinoma grading in MR images, introducing SE to enhance key features while suppressing redundant features, feature maps are easier to fuse, and network obtains better performance and accuracy. Squeeze excitation simulates interdependencies between channels but has limited processing of contextual information, Wang et al. [76] used dilated dense unit to extract multiscale features for crowd counting, channel attention mechanism to guide dense block multiscale contextual feature fusion, adjustment of feature channel weights at each fusion stage, recalibrate of fused feature responses, and its convolution layer to capture channel correlation and global average pooling aggregated output feature spatial information; bootstrap channel fusion to adjust feature method outperforms other methods in population density map estimation.

Second, spatial attention mechanism is introduced in dense block. Wei et al. [77] proposed Att-DenseUnet for skin lesion segmentation; spatial attention module was added to encoding and decoding skip connection, using advanced semantic contextual information in encoding stage to focus on lesion region features and suppressing irrelevant artifact region features in dense block, and final model output had high shape accuracy, clear boundary, and high recall segmentation results; Zhang et al. [78] used a multiscale model to automatically classify traffic patterns and speeds, and multiscale spatio-temporal granularity grid data features were extracted in dense block; spatial attention map is obtained after element-by-element dot product of 1×1 convolution and sigmoid functions, which reduces the number of network parameters and enhances interchannel feature

fusion and further enhances information flow in dense block using residual connection; final model outperforms existing models in terms of accuracy, recall, and $F1$ scores.

Third, channel spatial attention is introduced in dense block, which adaptively adjusts cross-channel and spatial region weight to improve network performance using interchannel and interfeature relationships. Qin and Gu [79] introduced channel and spatial attention to residual dense block and proposed deep multilevel residual attention network (MRAN) with enhanced image high-frequency features. Attention module aggregates channel information with feature statistics and multilayer perceptron compress channel information and sums it pixel by pixel, emphasizes information-rich channels, computes feature statistics to obtain spatial mapping, and multiplies element-by-element, and spatial attention maps along channel axis improve network perception, focusing more on informative rich features to recover more accurate details, and extensive evaluations show that MRAN is quantitative results and visual performance state-of-the-art methods.

Fourth, residual attention mechanism is introduced in dense block to maintain top-down spatial attention without interrupting bottom-up convolutional feature extraction process, and the residual connection attention mechanism fuses two different convolutional features and enhances feature representation. Bi et al. [80] proposed residual attention densely connected RADC for aerial image scene classification, residual attention mechanism highlights local semantics in output features, and experiments show that RADC with fewer parameters outperforms some existing methods.

(2) Attention Mechanism Introduced in Transition Layer. Attention mechanism introduced in transition layer has multilevel attention mechanism and multidimensional spatial attention mechanism with spatial attention gating.

First, multilevel attention mechanism is introduced in transition layer to fuse adjacent dense block output features and generate channel spatial attention map to enhance transition layer features. In diagnosing gastric cancer in gastric pathology section images, although DenseNet rich semantic information can detect larger sensory field data, it is still difficult to obtain spatial information and identify hidden gastric features in feature maps, Liu et al. [81] proposed multilevel attention dense network (MSA-DenseNet) to detect gastric cancer in 20x magnified section images; transition layer introduced attention providing attention vectors to enhance gastric features, selecting more semantic information about cancer, global average pool, and fully connected layer to provide nonlinearity for channel attention, using adjacent dense block features to enhance relevant semantic information, combining lower-order features with more spatial prediction and higher-order features with more semantic prediction features; final model obtains better detection results than existing manual detection methods.

Second, multidimensional attention with spatial attention gating is introduced between dense block at different scales to adjust multiscale feature fusion and improve network prediction capability. In noninvasive clinical image lung cancer diagnosis research, which lacks reliable extraction of fine-grained features in different imaging modalities, Qin [82] proposed multidimensional attention-based lung cancer classification architecture for PET and CT images, with 3D DenseNet processing different imaging modalities in parallel, embedding spatial attention gating in dense block, generating corresponding spatial mappings by high-level semantic features and SE, implementing multidimensional attention mechanism, enhancing fine-grained feature extraction, and effectively reducing feature noise; quantitative evaluation metrics and deep learning visualization analysis shows that architecture obtains 0.72 ROC curve area and 0.92 accuracy, which can effectively extract fine-grained features of different imaging modalities.

(3) Attention Mechanism Introduced to Whole Network. Attention mechanism introduced in whole network has channel spatial and LSTM adaptive attention mechanism. First, skin lesion regions vary in color, size, and shape; some light lesions are highly similar to normal skin, and boundaries of deeper lesion regions are complex, to segment model with sufficiently dense feature resolution, Ren et al. [83] used serial attention network (SANet) to segment skin lesion regions and introduced channel first followed by a spatial attention mechanism to aggregate global, local, and interchannel information using interdependencies between channel mappings to improve representation of semantic features and using spatial attention to select global contextual information and contextual relationships to make semantic features more compact and consistent; SANet captures interpixel and interchannel feature dependencies and achieves 0.7692 average Jaccard index in ISIC2017 dataset. Second, it is difficult for image caption generation tasks to correctly extract image global features and perform image regions for each word without ignoring partial words, Deng et al. [84] used DenseNet to extract features required for LSTM generation of sentences and introduced LSTM adaptive attention mechanism to improve correspondence problem of forced text and image regions, which was demonstrated on Flickr30k and COCO datasets showed that flexibility of caption generation was improved, with significant improvements in the BLEU and METEOR evaluation criteria.

4. Application of DenseNet in Medical Image Analysis

Reviewing the past 10 years, deep learning has seen a proliferation of improvements in models and algorithms, and availability and computational power of large-scale image data has continued to improve, solving many challenges in computer vision. In 2022, Li et al. [85] provided a comprehensive summary of deep learning research in medical image classification, detection and segmentation, alignment, and retrieval; applications of DenseNet for medical images mainly include image classification, segmentation, detection, alignment, reconstruction, retrieval, generation, enhancement, and fusion, but research and improvement of its technology mainly focus on classification and recognition,

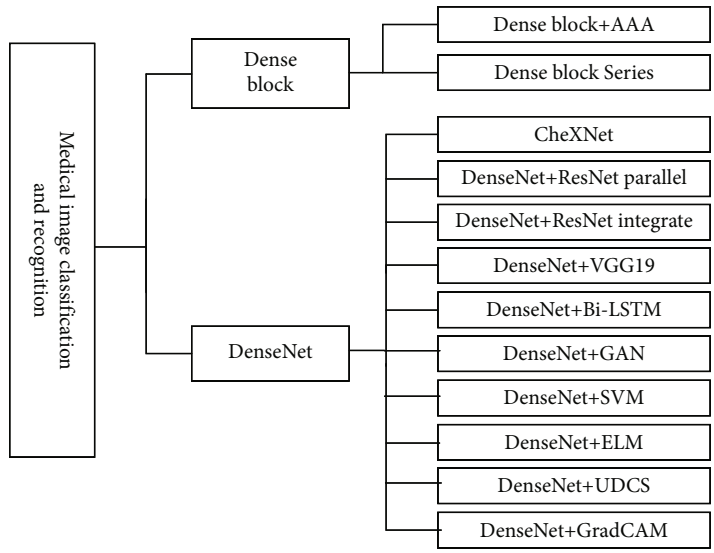


Figure 14: Application of DenseNet in medical classification and recognition.

segmentation, and localization detection; medical image classification commonly used application scenarios are disease diagnosis; medical image segmentation projects are mostly applied to segmentation of lesions and organs, such as brain segmentation, lung segmentation, heart segmentation, and liver segmentation. Therefore, this section summarizes applications of DenseNet in field of medical image analysis from these three aspects.

4.1. Application of DenseNet in Medical Pattern Recognition. Image pattern recognition is a popular research area in computer vision, which usually includes two steps: feature extraction and classifier. Medical image classification is further divided into image screening and lesion classification. Image screening uses detection images as model input to predict whether the disease is suffered or graded in severity. Lesion classification assists clinicians in diagnosing human lesions and grading the severity quantitatively, and it is important to identify tumors and nodules in human tissues and organs. This section summarizes applications of medical image pattern recognition from perspective of combining DenseNet with other methods; 12 improved methods are shown in Figure 14.

4.2. Combine Dense Block with Other Methods. First, Guo and Yuan [86] combined dense block and adaptive aggregation attention (AAA) module to classify wireless capsule endoscopy images (WCE); AAA captured global correlation to estimate abnormal regions and obtained 93.17% accuracy in quadruple crossover on CAD-CAP WCE dataset. Second, in pediatric radiology and forensic medicine, bone age assessment is one of clinical diagnostic techniques to evaluate skeletal development in children, Guo et al. [87] tandem six dense blocks to perform comprehensive evaluation of high- and low-quality X-ray images to achieve state-of-the-art performance under MAE metric.

4.3. DenseNet Variants. Rajpurkar et al. [88] proposed 121-layer convolutional neural network CheXNet based on DenseNet, applied to the largest publicly available chest X-ray dataset ChestX-ray14, which contains over 100,000 X-ray images of 14 diseases, and CheXNet exceeded average radiologist $F1$ metrics on a test set annotated by four practicing radiologists performance.

4.4. Combine DenseNet with Others. DenseNet integrates with other networks: first, chest X-ray images, as the most common screening technique for diagnosing chest diseases (atelectasis, cardiomegaly, pneumonia, etc.) suffer from complex background, many potential abnormalities, diverse interactions between abnormal patterns, etc. Traditional techniques focus on low-level features such as appearance, texture, and contrast, although deep networks can extract higher-level features, single network in intermediate layers also loses some unique details, and homogeneous networks learning complementary features from each other may overlap, Chen et al. [89] used DualCheXNet, a dual asymmetric network, to improve performance of multilabel chest disease classification, with parallel asymmetric subnetworks DenseNet and ResNet both using jump connections to create short paths, and enhanced gradient propagation from previous to following layers, residual connection to reuse early features, and dense connection to mine new features; advantages of different unique feature extraction methods complemented each other and were validated on ChestX-ray14 dataset, obtaining average AUC score of 0.823. Second, Srivastva et al. [90] proposed PlexNet integrated with pretrained DenseNet and ResNet for ECG biometrics, feature level fusion approach for tandem features, and fully connected layer based on similarity classification, using both migration learning and integrated learning to take advantage of model effectiveness and robustness demonstrated on PTB and CYBHI datasets. Third, Kedia et al. [91] combined DenseNet121 and VGG19 into CoVNet-19 two-level stacking

model, both models extracted chest X-ray features and minimized noise, SVM achieved three-level classification of COVID-19, pneumonia, and normal, obtaining 98.28% overall accuracy and 97.15% Mathews correlation coefficient and $F1$ scores of 99% in both binary classifications.

DenseNet combined with other network structures; fourth, Li et al. [92] combined DenseNet and bidirectional recurrent neural network (Bi-LSTM) to predict vivo RNA transcripts from in vitro data; Bi-LSTM learns RNA sequences, secondary structure information and pairwise probability features, which were reshaped into matrices and fed into DenseNet with prediction accuracy and scalability in with better performance. Fifth, Man et al. [93] screened mislabeled plaques in breast histopathology images based on generative adversarial network (GAN) unsupervised abnormality detection; DenseNet extracted multilevel features of plaques for benign and malignant classification, achieving excellent results on coarse-grained high-resolution images especially 40x and 100x images.

DenseNet is combined with other methods; sixth, Devnath et al. [94] used CheXNet to extract multilevel features of chest X-ray images, mixed SVM and CNN aggregation methods, transition layer low-level features, and deep-level features mapped to high-dimensional space after merging dichotomous classification. Seventh, Turkoglu [95] proposed a multicore extreme learning machine (ELM) method to extract deep-level features of chest CT with pretrained DenseNet201, and ELM classifier classified features with different activation methods to predict final presence of neocoronary pneumonia by majority voting. Eighth, Yuan et al. [96] addressed the problem of imbalance and discrepancy between small and large categories in wireless capsule endoscopy images and proposed DenseNet-UDCS polyp recognition model by calculating suitable features, unbalance discriminant loss (UD) function to improve model sensitivity to minority groups and category-sensitive loss (CS) constraint to learn features in hierarchically between different categories, so that intraclass differences minimized and interclass distance within batches maintained, with more advantages in model recognition accuracy, computational speed, and stability. Ninth, Ellis et al. [97] mixed weakly supervised, extrasupervised, and annotated classification of two-view chest radiographs; DenseNet-121 extracted features to generate 1D global feature vector and estimated scores; fully connected layer gradients were convolved with feature sets to estimate gradient class activation maps (Grad-CAM); attention mining branch learned GradCAM threshold operations using DenseNet-121, which masked important classification regions in input, increasing accuracy and sensitivity to negative prediction values, despite a decrease in false positives and decreases in overall accuracy.

4.5. Application of DenseNet in Medical Image Segmentation. Medical image segmentation is to identify contour of target region or internal voxel collections, which is a key task for clinical surgical images to guide tumor radiotherapy, medical image organ and lesion segmentation allows quantitative analysis of volume and shape-related clinical parameters. Effective automatic organ segmentation is very challenging,

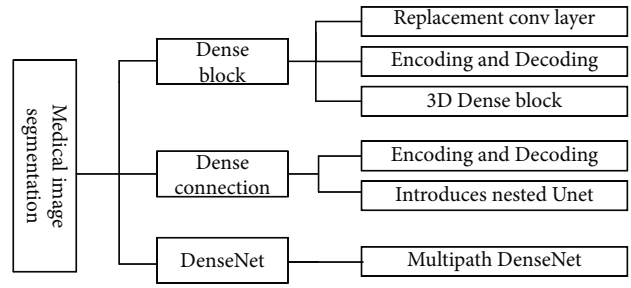


Figure 15: Application of DenseNet in medical image segmentation.

medical images have high complexity, lack of simple linear features, insignificant brightness differences between organs, connected to surrounding organs, and blurred boundaries, as well as interference factors such as image artifacts and noise, and segmentation results accuracy is also affected by partial volume effects, gray scale inhomogeneity, artifacts and gray scale proximity between different soft tissues. Traditional segmentation algorithms are difficult to process images accurately, and application of DenseNet makes medical image segmentation effect improved, summarizing the seven models shown in Figure 15.

4.5.1. Dense Block in Medical Image Segmentation Application Research. First, dense block replaces convolutional layer. Ding et al. [98] used dense block instead of codec convolutional layer to extract low-level visual features of multimodal brain tumors, and fused high-level semantic features to generate high-resolution features with fewer network parameters and fast segmentation, achieving good segmentation results. AIGhamdi et al. [99] introduced dense block in U-net for breast artery calcification detection; DU-net has dense connectivity, which helps to improve computational reusability and gradient mobility and improves accuracy and training difficulty. Second, codec introduces dense block. Vila et al. [100] performed semantic segmentation of carotid ultrasound plaques based on DenseNet, dense connectivity captured multiscale contextual information, and correlation coefficient of carotid intima-media thickness reached 0.81 in the experiment; robustness and generalization ability were validated on two datasets. Thirdly, 3D dense block was applied to segmentation networks. Bui et al. [101] proposed 3D-SkipDenseSeg, an accurate segmentation method for infant brain MRI, in which jump connections directly combine fine and coarse feature information from different levels of dense block to alleviate low accuracy problem caused by low-intensity contrast between MRI tissues, final model obtained the best dice similarity coefficient of 90.37, and segmentation accuracy and parameter efficiency are high.

(1) Dense Connection in Medical Image Segmentation Application Research. First, dense connection is introduced in codecs. Shi et al. [102] proposed dual dense context-aware network (DDCNet) to automatically segment hippocampus; dense connection is introduced between codecs to fuse high level and low level multiresolution features and contextual

information, which can enhance interlayer feature propagation and improve information flow between codecs. Second, nested U-net introduces dense connection; Li et al. [103] proposed ANU-Net combining deeply supervised codec structure and nested segmentation network, designed full resolution nested U-net between codecs, introduced dense connection between nested convolutional layers of same resolution to obtain full-resolution features at different semantic levels, and experimentally demonstrated that ANU-Net can suppress irrelevant background regions while increase target region weight.

(2) DenseNet in Medical Image Segmentation Application Research. Multipath DenseNet extracts features. Kitrungrotsakul et al. [104] proposed multipath DenseNet for hepatic vascular segmentation, binary classification of blocks in three planes, sagittal, coronal and transverse; IRCAD dataset experiments show that the model is currently the latest method accuracy.

4.6. Application of DenseNet in Medical Object Detection. Localization of anatomical structures and lesion regions in human organs is an important preprocessing step in clinical treatment planning and intervention, and localization of specific biomarkers or anatomical structures in medical images is directly related to treatment outcome. The key target of interest or lesion detection is to classify each pixel; DenseNet is applied to improve the effectiveness of medical image localization detection, with the following three main models.

First, Ghatwary et al. [105] detected esophageal abnormalities in endoscopic images based on faster R-CNN- and DenseNet-enhanced feature propagation between layers, combined with Gabor manual feature enhancement to detect texture details in stage, obtaining 92.1% and 91% accuracy on MICCAI2015 and Kvasir datasets. Second, Liu et al. [106] proposed fully convolutional dense pixel classifier (FC-DenseNet) depicting ribs and clavicles, aggregating contextual information with jump connections, dense blocks instead of upsampling paths in convolutional layer, and dense connections to enhance feature propagation and reuse, and FC-DenseNet automatically depicts chest radiographs from multiple public databases to generate accurate binary edge maps, verifying the effectiveness of the model for bone suppression. Third, Xiao et al. [107] designed Siam network based on DenseNet to track vessel wall displacement from ultrasound RF signals and target blocks of labeled data as convolutional kernels to calculate blocks in search region with similarity score maps, achieving first single-target displacement tracking in ultrasound pulse wave imaging, accurately predicting local mechanical properties of arteries, and tracking better than traditional single-target block matching better.

5. Summary and Outlook

In summary, DenseNet, which can learn deeper and more discriminative features from images, has been applied to several research areas of medical image analysis and has made breakthroughs; although DenseNet has been effectively validated in publicly available and relevant medical image databases [41], it still faces many challenges. First, the huge amount of medical data outside images [4] contains rich information, but it is difficult to be fully utilized, such as additional labels indicating lesion conditions, clinical diagnosis reports reflecting physicians' conclusions and descriptions, which can help deep learning models to improve diagnostic performance, and recurrent convolutional neural networks (RNN) can be utilized to incorporate medical other data into DenseNet.

Second, deep learning models rely on a large number of manually labeled datasets [108, 109], which usually require significant effort and time to collect, clean, and debug data, and changes and evolution of actual task requirements also lead to retagging of the datasets, data dependence of models and cost of dataset labeling are issues that need to be urgently addressed by researchers. With the limited labeled trainable data [110, 111] currently available for training, semisupervised and weakly supervised methods can be used to learn unlabeled, weak, and small portions of labeled data.

Finally, there are still many challenges to training a model with higher classification or prediction accuracy and better generalization ability. There are limitations in using multiple branches to achieving DenseNet broaden; DenseNet lightweight is not only studied from the direction of architecture design and model compression but also based on data considerations, such as dimensionality reduction of intermediate features; dense connections lead to redundancy of feature extraction in different layers; dense block has excellent performance, but transition layer structure of dimensionality reduction is too simple, and it is difficult for single sensory field to capture multilevel features of dense block; the existing attention mechanism is basically in form of "weighting," and appropriate attention mechanism should be designed according to data characteristics, such as residual shrinkage network [112]; "soft thresholding" is used to cope with strong noisy data.

Massive medical domain data presents challenges and opportunities for researchers in medical image analysis, by appropriate integration methods, different types of domain knowledge are used to assist deep learning to better accomplish corresponding tasks, and natural image pretraining weights are also introduced to improve network performance in medical domain. The massive data in medical image analysis will provide inexhaustible power as network architecture and algorithms improve; dense block and dense connection in DenseNet can be optimized for these new models, and DenseNet-base derivative models may also yield significant improvements in results for specific tasks.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (grant no. 62062003), Key Research and Development Project of Ningxia (grant no. 2020BEB04022), Natural Science Foundation of Ningxia (grant no. 2022AAC03149), and North Minzu University Research Project of Talent Introduction (grant no. 2020KYQD08).

References

[1] I. S. Krizhevsky and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceeding of Neural Information Processing Systems*, pp. 1097–1105, Curran Associates, Lake Tahoe, USA, 2012.

[2] G. Huang, Z. Liu, V. Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, https://arxiv.org/abs/1608.06993.

[3] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[4] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Medical Image Analysis*, vol. 69, 2021.

[5] J. He, T. Cao, F. Xu et al., "Artificial intelligence-based screening for diabetic retinopathy at community hospital," *Eye*, vol. 34, no. 3, pp. 572–576, 2020.

[6] D. Wang, C. Huang, S. Bao et al., "Study on the prognosis predictive model of COVID-19 patients based on CT radiomics," *Scientific Reports*, vol. 11, no. 1, 2021.

[7] R. Zhang, L. Zhao, T. Lou et al., "Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2149–2160, 2018.

[8] F. Li, M. Liu, and Alzheimer's Disease Neuroimaging Initiative, "Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 101–110, 2018.

[9] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, vol. 17, pp. 1175–1183, Honolulu, HI, USA, 2017.

[10] L. Ke, Y. Deng, W. Xia et al., "Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images," *Oral Oncology*, vol. 110, 2020.

[11] Z. Liu, Y. Zhang, W. Li, S. Li, Z. Zou, and B. Chen, "Multislice left ventricular ejection fraction prediction from cardiac MRIs without segmentation using shared SptDenNet," *Computerized Medical Imaging and Graphics*, vol. 86, 2020.

[12] D. M. H. Kishan and V. Janardhan, "Intracranial hemorrhage detection," *Materials Today: Proceedings*, vol. 43, no. 6, pp. 3706–3714, 2021.

[13] L. Li, Y. Yang, H. Liang, and B. Wu, "Transfer learning for establishment of recognition of COVID-19 on CT imaging using small-sized training datasets," *Knowledge-Based Systems*, vol. 218, article 106849, 2021.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[16] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," 2017, https://arxiv.org/abs/1707.01629.

[17] Y. Yang, Z. Zhong, T. Shen, and Z. Lin, "Convolutional Neural Networks with Alternately Updated Clique," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2413–2422, Salt Lake City, USA, 2018.

[18] L. J. Ba and R. Caruana, "Do Deep Nets Really Need to Be Deep?," 2013, https://arxiv.org/abs/1312.6184.

[19] B. Hanin and D. Rolnick, "Complexity of Linear Regions in Deep Networks," in *Proceedings of the 11th Asian Conference on Machine Learning (ACML)*, pp. 2596–2604, Nagoya, Japan, 2019.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, Las Vegas, NV, USA, 2016.

[21] H. Li, S. Zhuang, D. Li, J. Zhao, and Y. Ma, "Benign and malignant classification of mammogram images based on deep learning," *Biomedical Signal Processing and Control*, vol. 51, pp. 347–354, 2019.

[22] T. W. Arega and S. Bricq, "Automatic myocardial scar segmentation from multi-sequence cardiac MRI using fully convolutional densenet with inception and squeeze-excitation module," in *Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention*, Lima, Peru, 2020.

[23] B. Lodhi and J. Kang, "Multipath-DenseNet: a supervised ensemble architecture of densely connected convolutional networks," *Information Sciences*, vol. 482, pp. 63–72, 2019.

[24] B. Chen, T. Zhao, J. Liu, and L. Lin, "Multipath feature recalibration DenseNet for image classification," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 3, pp. 651–660, 2021.

[25] Y. Liu, P. Hao, P. Zhang, X. Xu, J. Wu, and W. Chen, "Dense convolutional binary-tree networks for lung nodule classification," *IEEE Access*, vol. 6, pp. 49080–49088, 2018.

[26] M. T. Nguyen, P. Siritanawan, and K. Kotani, "Saliency detection in human crowd images of different density levels using attention mechanism," *Signal Processing: Image Communication*, vol. 88, 2020.

[27] G. Huang, S. Liu, L. Maaten, and K. Q. Weinberger, "CondenseNet: An Efficient DenseNet Using Learned Group Convolutions," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2752–2761, Salt Lake City, USA, 2018.

[28] G. Li, M. Zhang, J. Li, F. Lv, and G. Tong, "Efficient densely connected convolutional neural networks," *Pattern Recognition*, vol. 109, article 107610, 2021.

[29] A. G. Howard, M. Zhu, B. Chen et al., "Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017, https://arxiv.org/abs/1704.04861.

[30] J. Chen, W. Wang, D. Zhang, A. Zeb, and Y. A. Nanehkaran, "Attention embedded lightweight network for maize disease recognition," *Plant Pathology*, vol. 70, no. 3, pp. 630–642, 2021.

[31] X. Yu, H. He, H. Ma, X. Qi, J. Lu, and Y. Zhao, "S-DenseNet: a DenseNet compression model based on convolution grouping strategy using skyline method," *IEEE Access*, vol. 7, pp. 183604–183613, 2019.

[32] Z. Huang, X. Zhu, M. Ding, and X. Zhang, "Medical image classification using a light-weighted hybrid neural network based on PCANet and DenseNet," *IEEE Access*, vol. 8, pp. 24697–24712, 2020.

[33] Y. Lee, J. W. Hwang, S. Lee, Y. Bae, and J. Park, "An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 752–760, Long Beach, USA, 2019.

[34] T. Li, W. Jiao, L. Wang, and G. Zhong, "Automatic DenseNet sparsification," *IEEE Access*, vol. 8, pp. 62561–62571, 2020.

[35] X. Zhu, Z. Li, X. Zhang, H. Li, Z. Xue, and L. Wang, "Generative adversarial image super-resolution through deep dense skip connections," *Computer Graphics Forum*, vol. 37, no. 7, pp. 289–300, 2018.

[36] Y. Wei and X. Liu, "Dangerous goods detection based on transfer learning in X-ray images," *Neural Computing and Applications*, vol. 32, no. 12, pp. 8711–8724, 2020.

[37] X. Wei, D. Wei, D. Suo, L. Jia, and Y. Li, "Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model," *IEEE Access*, vol. 8, pp. 61973–61988, 2020.

[38] B. Lei, Z. Xia, F. Jiang et al., "Skin lesion segmentation via generative adversarial networks with dual discriminators," *Medical Image Analysis*, vol. 64, article 101716, 2020.

[39] M. K. Hasan, L. Dahal, P. N. Samarakoon, F. I. Tushar, and R. Martí, "DSNet: automatic dermoscopic skin lesion segmentation," *Computers in Biology and Medicine*, vol. 120, 2020.

[40] H. T. Mustafa, M. Zareapoor, and J. Yang, "MLDNet: multi-level dense network for multi-focus image fusion," *Signal Processing: Image Communication*, vol. 85, 2020.

[41] X. Li, L. Shen, X. Xie et al., "Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection," *Artificial Intelligence in Medicine*, vol. 103, 2020.

[42] Z. Feng, J. Lai, X. Xie, and J. Zhu, "Image super-resolution via a densely connected recursive network," *Neurocomputing*, vol. 316, pp. 270–276, 2018.

[43] L. Yang, Y. Qin, and X. Zhang, "Lightweight densely connected residual network for human pose estimation," *Journal of Real-Time Image Processing*, vol. 18, no. 1, pp. 825–837, 2021.

[44] M. Wang, X. Yang, M. Anisetti, R. Zhang, M. K. Albertini, and K. Liu, "Image super-resolution via enhanced multi-scale residual network," *Journal of Parallel and Distributed Computing*, vol. 152, pp. 57–66, 2021.

[45] Y. Wu, Y. Qin, Z. Wang, X. Ma, and Z. Cao, "Densely pyramidal residual network for UAV-based railway images dehazing," *Neurocomputing*, vol. 371, pp. 124–136, 2020.

[46] Y. Long, H. Jia, Y. Zhong, Y. Jiang, and Y. Jia, "RXDNFuse: a aggregated residual dense network for infrared and visible image fusion," *Information Fusion*, vol. 69, pp. 128–141, 2021.

[47] B. Zhang, S. Liu, S. Zhou et al., "Cerebrovascular segmentation from TOF-MRA using model- and data-driven method via sparse labels," *Neurocomputing*, vol. 380, pp. 162–179, 2020.

[48] X. Wang, Y. Gu, X. Gao, and Z. Hui, "Dual residual attention module network for single image super resolution," *Neurocomputing*, vol. 364, pp. 269–279, 2019.

[49] W. Zeng, J. Peng, S. Wang, and Q. Liu, "A comparative study of CNN-based super-resolution methods in MRI reconstruction and its beyond," *Signal Processing: Image Communication*, vol. 81, 2020.

[50] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "DENSE-INception U-net for medical image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 192, 2020.

[51] S. Qamar, P. Ahmad, and L. Shen, "Dense encoder-decoder–based architecture for skin lesion segmentation," *Cognitive Computation*, vol. 13, no. 2, pp. 583–594, 2021.

[52] Z. Zhang, Z. Tang, Y. Wang et al., "Dense residual network: enhancing global dense feature flow for character recognition," *Neural Networks*, vol. 139, pp. 77–85, 2021.

[53] A. Adegun and S. Viriri, "FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images," *IEEE Access*, vol. 8, pp. 150377–150396, 2020.

[54] Y. Hong, B. Wei, Z. Han, X. Li, Y. Zheng, and S. Li, "MMCL-Net: spinal disease diagnosis in global mode using progressive multi-task joint learning," *Neurocomputing*, vol. 399, pp. 307–316, 2020.

[55] Z. Huang, Y. Zhao, X. Li et al., "Application of innovative image processing methods and AdaBound-SE-DenseNet to optimize the diagnosis performance of meningiomas and gliomas," *Biomedical Signal Processing and Control*, vol. 59, article 101926, 2020.

[56] S. Wang, Z. Liu, Y. Rong et al., "Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer," *Radiotherapy and Oncology*, vol. 132, pp. 171–177, 2019.

[57] A. Pemasiri, K. Nguyen, S. Sridharan, and C. Fookes, "Multi-modal semantic image segmentation," *Computer Vision and Image Understanding*, vol. 202, 2021.

[58] T. J. Jun, J. Kweon, Y. Kim, and D. Kim, "T-Net: nested encoder-decoder architecture for the main vessel segmentation in coronary angiography," *Neural Networks*, vol. 128, pp. 216–233, 2020.

[59] X. Yang, Y. Zhang, Y. Guo, and D. Zhou, "An image super-resolution deep learning network based on multi-level feature extraction module," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7063–7075, 2021.

[60] B. N. Mostefa, A. Moharmed, S. Rachida, and K. Rostom, "Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy," *Medical Image Analysis*, vol. 63, article 101692, 2020.

[61] Y. Zhou, Y. Zhang, X. Xie, and S. Kung, "Image super-resolution based on dense convolutional auto-encoder blocks," *Neurocomputing*, vol. 423, pp. 98–109, 2021.

[62] A. Signoroni, M. Savardi, S. Benini et al., "BS-Net: learning COVID-19 pneumonia severity on a large chest X-ray dataset," *Medical Image Analysis*, vol. 71, article 102046, 2021.

[63] K. Tseng, R. Zhang, C. Chen, and M. M. Hassan, "DNetUnet: a semi-supervised CNN of medical image segmentation for super-computing AI service," *The Journal of Supercomputing*, vol. 77, no. 4, pp. 3594–3615, 2021.

[64] J. Liu, M. Kocak, M. Supanich, and J. Deng, "Motion artifacts reduction in brain MRI by means of a deep residual network with densely connected multi-resolution blocks (DRN-DCMB)," *Magnetic Resonance Imaging*, vol. 71, pp. 69–79, 2020.

[65] M. Wang, H. Cai, J. Zhou, and M. Zhou, "Interlayer and intralayer scale aggregation for scale-invariant crowd counting," *Neurocomputing*, vol. 441, pp. 128–137, 2021.

[66] K. Sun, X. Wen, L. Yuan, and H. Xu, "Dense capsule networks with fewer parameters," *Soft Computing*, vol. 25, no. 10, pp. 6927–6945, 2021.

[67] D. Wang, G. Hu, and C. Lyu, "Multi-path connected network for medical image segmentation," *Journal of Visual Communication and Image Representation*, vol. 71, article 102852, 2020.

[68] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," 2017, https://arxiv.org/abs/1706.03762.

[69] J. Bastings and K. Filippova, "The Elephant in the Interpretability Room: Why Use Attention as Explanation when we Have Saliency Methods?," 2020, https://arxiv.org/abs/2010.05607.

[70] Z. Shi, T. Wang, Z. Huang, F. Xie, and G. Song, "A method for the automatic detection of myopia in Optos fundus images based on deep learning," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 37, no. 6, article e3460, 2021.

[71] Y. Yan, B. Ni, J. Liu, and X. Yang, "Multi-level attention model for person re-identification," *Pattern Recognition Letters*, vol. 127, pp. 156–164, 2019.

[72] S. Xue, D. G. Liu, and Y. Ma, "ATP-DenseNet: a hybrid deep learning-based gender identification of handwriting," *Neural Computing and Applications*, vol. 33, no. 10, pp. 4611–4622, 2021.

[73] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, and X. Long, "A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification," *Magnetic Resonance Imaging*, vol. 78, pp. 119–126, 2021.

[74] W. Bai, Y. Lu, Y. Huang, L. Zha, and J. Yang, "Densely convolutional attention network for image super-resolution," *Neurocomputing*, vol. 368, pp. 25–33, 2019.

[75] Q. Zhou, Z. Zhou, C. Chen et al., "Grading of hepatocellular carcinoma using 3D SE-DenseNet in dynamic enhanced MR images," *Computers in Biology and Medicine*, vol. 107, pp. 47–57, 2019.

[76] L. Wang, Y. Li, S. Peng, X. Tang, and B. Yin, "Multi-level feature fusion network for crowd counting," *IET Computer Vision*, vol. 15, no. 1, pp. 60–72, 2021.

[77] Z. Wei, H. Song, L. Chen, Q. Li, and G. Han, "Attention-based DenseUnet network with adversarial training for skin lesion segmentation," *IEEE Access*, vol. 7, pp. 136616–136629, 2019.

[78] R. Zhang, P. Xie, C. Wang, G. Liu, and S. Wan, "Classifying transportation mode and speed from trajectory data via deep multi- scale learning," *Computer Networks*, vol. 162, article 106861, 2019.

[79] D. Qin and X. Gu, "Single-image super-resolution with multilevel residual attention network," *Neural Computing and Applications*, vol. 32, no. 19, pp. 15615–15628, 2020.

[80] Q. Bi, K. Qin, H. Zhang, Z. Li, and K. Xu, "RADC-Net: a residual attention based convolution network for aerial scene classification," *Neurocomputing*, vol. 377, pp. 345–359, 2020.

[81] B. Liu, Y. Zhao, B. Yang, S. Zhao, R. Gu, and M. Gahegan, "A gastric cancer recognition algorithm on gastric pathological sections based on multistage attention-DenseNet," *Concurrency and Computation Practice and Experience*, vol. 33, no. 10, article e6188, 2021.

[82] R. Qin, Z. Wang, L. Jiang et al., "Fine-grained lung cancer classification from PET and CT images based on multidimensional attention mechanism," *Complexity*, vol. 2020, Article ID 6153657, 12 pages, 2020.

[83] Y. Ren, L. Yu, S. Tian, J. Cheng, Z. Guo, and Y. Zhang, "Serial attention network for skin lesion segmentation," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 799–810, 2021.

[84] Z. Deng, Z. Jiang, R. Lan, W. Huang, and X. Luo, "Image captioning using DenseNet network and adaptive attention," *Signal Processing: Image Communication*, vol. 85, 2020.

[85] X. Li, C. Li, M. M. Rahaman et al., "A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches," *Artificial Intelligence Review*, vol. 29, pp. 609–639, 2022.

[86] X. Guo and Y. Yuan, "Semi-supervised WCE image classification with adaptive aggregated attention," *Medical Image Analysis*, vol. 64, 2020.

[87] J. Guo, J. Zhu, H. Du, and B. Qiu, "A bone age assessment system for real-world X-ray images based on convolutional neural networks," *Computers & Electrical Engineering*, vol. 81, 2020.

[88] P. Rajpurkar, J. Irvin, K. Zhu et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," 2017, https://arxiv.org/abs/1711.05225.

[89] B. Chen, J. Li, X. Guo, and G. Lu, "DualCheXNet: dual asymmetric feature learning for thoracic disease classification in chest X-rays," *Biomedical Signal Processing and Control*, vol. 53, article 101554, 2019.

[90] R. Srivastva, A. Singh, and Y. N. Singh, "PlexNet: a fast and robust ECG biometric system for human recognition," *Information Sciences*, vol. 558, pp. 208–228, 2021.

[91] P. Kedia, B. Anjum, and R. Katarya, "CoVNet-19: a deep learning model for the detection and analysis of COVID-19 patients," *Applied Soft Computing*, vol. 104, article 107184, 2021.

[92] Z. Li, J. Zhu, X. Xu, and Y. Yao, "RDense: a protein-RNA binding prediction model based on bidirectional recurrent neural network and densely connected convolutional networks," *IEEE Access*, vol. 8, pp. 14588–14605, 2020.

[93] R. Man, P. Yang, and B. Xu, "Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks," *IEEE Access*, vol. 8, pp. 155362–155377, 2020.

[94] L. Devnath, S. Luo, P. Summons, and D. Wang, "Automated detection of pneumoconiosis with multilevel deep features learned from chest X-Ray radiographs," *Computers in Biology and Medicine*, vol. 129, 2021.

[95] M. Turkoglu, "COVID-19 detection system using chest CT images and multiple kernels-extreme learning machine based on deep neural network," *IRBM*, vol. 42, no. 4, pp. 207–214, 2021.

[96] Y. Yuan, W. Qin, B. Ibragimov et al., "Densely connected neural network with unbalanced discriminant and category sensitive constraints for polyp recognition," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 2, pp. 574–583, 2020.

[97] R. Ellis, E. Ellestad, B. Elicker, M. D. Hope, and D. Tosun, "Impact of hybrid supervision approaches on the performance of artificial intelligence for the classification of chest radiographs," *Computers in Biology and Medicine*, vol. 120, article 103699, 2020.

[98] Y. Ding, L. Gong, M. Zhang, C. Li, and Z. Qin, "A multi-path adaptive fusion network for multimodal brain tumor segmentation," *Neurocomputing*, vol. 412, pp. 19–30, 2020.

[99] M. AlGhamdi, M. Abdel-Mottaleb, and F. Collado-Mesa, "DU-Net: convolutional network for the detection of arterial calcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3240–3249, 2020.

[100] M. D. M. Vila, B. Remeseiro, M. Grau et al., "Semantic segmentation with DenseNets for carotid artery ultrasound plaque segmentation and CIMT estimation," *Artificial Intelligence in Medicine*, vol. 103, 2020.

[101] T. D. Bui, J. Shin, and T. Moon, "Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation," *Biomedical Signal Processing and Control*, vol. 54, article 101613, 2019.

[102] J. Shi, R. Zhang, L. Guo et al., "Dual dense context-aware network for hippocampal segmentation," *Biomedical Signal Processing and Control*, vol. 61, article 102038, 2020.

[103] C. Li, Y. Tan, W. Chen et al., "ANU-Net: attention-based nested U-Net to exploit full resolution features for medical image segmentation," *Computers & Graphics*, vol. 90, pp. 11–20, 2020.

[104] T. Kitrungrotsakul, X. Han, Y. Iwamoto et al., "VesselNet: a deep convolutional neural network with multi pathways for robust hepatic vessel segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 74–83, 2019.

[105] N. Ghatwary, X. Ye, and M. Zolgharni, "Esophageal abnormality detection using DenseNet based faster R-CNN with Gabor features," *IEEE Access*, vol. 7, pp. 84374–84385, 2019.

[106] Y. Liu, X. Zhang, G. Cai et al., "Automatic delineation of ribs and clavicles in chest radiographs using fully convolutional DenseNets," *Computer Methods and Programs in Biomedicine*, vol. 180, article 105014, 2019.

[107] C. Xiao, Z. Li, J. Lu et al., "A new deep learning method for displacement tracking from ultrasound RF signals of vascular walls," *Computerized Medical Imaging and Graphics*, vol. 87, article 101819, 2021.

[108] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense modeling," *Applied Soft Computing*, vol. 75, pp. 323–332, 2019.

[109] M. M. Rahaman, C. Li, Y. Yao, and X. Zhao, "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[110] Y. Li, C. Li, X. Li et al., "A comprehensive review of Markov random field and conditional random field approaches in pathology image analysis," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 609–639, 2022.

[111] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, article 106885, 2021.

[112] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4681–4690, 2020.

Hindawi
BioMed Research International
Volume 2024, Article ID 9878053, 1 page
https://doi.org/10.1155/2024/9878053

Retraction

Retracted: Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Lv, N. Ta, T. Chen, J. Zhao, and H. Wei, "Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation," *BioMed Research International*, vol. 2022, Article ID 7020804, 8 pages, 2022.

Hindawi
BioMed Research International
Volume 2022, Article ID 7020804, 8 pages
https://doi.org/10.1155/2022/7020804

RETRACTED

Research Article

Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation

Xinyu Lv,[1] Na Ta,[1] Tao Chen,[1] Jing Zhao,[2] and Haicheng Wei[1]

[1]*School of Electrical and Information Engineering, North Minzu University, Yinchuan 750021, China*
[2]*School of Information Engineering, Ningxia University, Yinchuan 750021, China*

Correspondence should be addressed to Haicheng Wei; wei_hc@nun.edu.cn

Received 24 January 2022; Accepted 31 March 2022; Published 14 April 2022

Academic Editor: Chen Li

Copyright © 2022 Xinyu Lv et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A gait feature analysis method based on AlphaPose human pose estimation fused with sample entropy is proposed to address complicated, high-cost, and time-consuming postoperative rehabilitation of patients with joint diseases. First, TensorRT was used to optimize the inference of AlphaPose, which consists of the target detection algorithm YOLOv3 and the pose estimation algorithm. It can speed up latency and throughput by about 2.5 times while maintaining the algorithm's accuracy. Second, the optimized human posture estimation algorithm AlphaPose_trt was used to process gait videos of healthy people and patients with knee arthritis. The joint point motion trajectories of the two groups were extracted, and the sample entropy algorithm quantified the joint trajectory signals for feature analysis. The experimental results showed significant differences in the entropy of the heel and ankle joint motion signals between healthy people and arthritic patients ($p < 0.01$), which can be used to identify patients with knee arthritis. This technique can assist doctors in determining needed postoperative joint surgery rehabilitation.

1. Introduction

With increases in the aging population and the accelerating pace of life, the number of people suffering from joint diseases continues to increase [1]. Orthopedic surgeons often need to use professional equipment to diagnose patients' disease conditions in diagnosis and treatment [2], but these analytical methods require specialized skills and can be painful and financially burdensome for patients [3]. In addition, it is often challenging to distinguish joint disease from gait instability, making it more difficult for doctors to judge the patient's skeletal motion state [4]. Therefore, gait analysis can determine whether there is an abnormality in a particular joint of a patient [5]. An objective determination of the patient's gait characteristics can guide clinical rehabilitation treatment [6].

Yang et al. [7] compared the gait characteristics of healthy young and aged people for dual tasks using a three-dimensional gait analysis system, which can be used as a reference for preventing falls in the elderly. Cuadrado et al. [8] proposed an extended Kalman filter (EKF) to analyze the gait of healthy subjects using optical markers and an inertial measurement unit (IMU). The results show a good correlation between the parameters obtained by the two methods. Seifert et al. [9] proposed a gait classification method based on physical features, subspace features, and harmonic modeling that correctly identified gait categories. However, the above methods have some disadvantages, such as complicated analysis processes, high cost, and long experimental cycles.

In recent years, with the maturity of deep learning technology, machine learning has gradually been applied in various areas of the medical field, such as precise cell classification [10] and image analysis for pathology [11, 12]. In addition, human pose estimation has broad application prospects in computer vision, pattern recognition, video/image sequence processing, and other technologies [13]. Cao et al. [14] developed the part affinity field (PAF) nonparametric representation method to learn how to associate body parts with individuals in an image to detect real-time multiperson two-dimensional poses. Xiu et al. [15] proposed an efficient tracker for multiperson joint pose estimation in complex

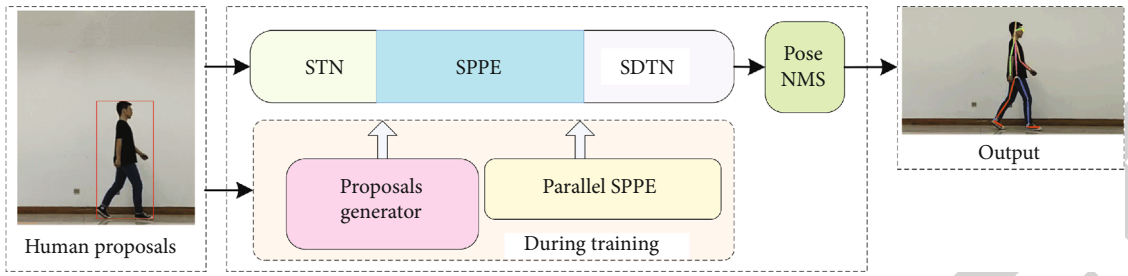


Figure 1: AlphaPose algorithm.

unconstrained video. With the advent of times-series pose estimation algorithms, such as AlphaPose [16], human pose recognition is more convenient and faster.

Therefore, a human gait analysis algorithm based on the sample entropy fusion AlphaPose algorithm is proposed in this paper. TensorRT [17] was used to optimize the AlphaPose reasoning. Combined with sample entropy quantization, the motion trajectory signal of joint points is extracted. Statistical analysis found significant differences in the sample entropies of heel and ankle trajectory signals between patients and healthy people. This method can help doctors judge the rehabilitation of patients after operations.

1.1. Gait Analysis Method Based on AlphaPose Fusion Sample Entropy

1.1.1. AlphaPose Human Pose Estimation. AlphaPose adopts a top-down attitude detection strategy. This method first detects the human body and then recognizes the human posture. A flowchart of the AlphaPose algorithm is shown in Figure 1.

AlphaPose consists mainly of the target detection algorithm YOLOv3 [18] and the pose estimation algorithm. First, the algorithm uses a target detection model to detect the person. After acquiring human proposals, the space transformation network (STN) adaptively transforms the input image into various spatial transformations. The spatially transformed images are then input to the single-person pose estimation (SPPE) network. The estimated human posture is then remapped to the original image coordinates using the space detransformation network (SDTN). In combination with parametric pose nonmaximum suppression (PPNMS), the pose similarity is calculated by defining the pose distance to eliminate redundant detection frames.

1.1.2. Sample Entropy Algorithm. Sample entropy is widely used in gait analysis. For example, sample entropy can quantify the components of plantar pressure and torque for different activities, from sitting to walking [19]. In this paper, sample entropy is used to reflect the signal characteristics of different people's joints when they walk.

Sample entropy is a detection method to measure the complexity of a time-series signal. The higher the complexity of the signal, the larger its sample entropy. Sequences with higher self-similarity are simpler and have lower sample entropy. However, the calculation of sample entropy does not depend on the data length. The algorithmic steps for computing sample entropy are as follows:

(1) The data of n sample points are used to compose the time series X

(2) Divide the data of n sample points to form a time series X. Divide X into $n-m+1$ arrays $X_m(i)$ of dimension m. $X_m(i)$ is shown in the following equation:

$$X_m(i) = [X(i), X(i+1), \cdots, (x+m-1)], \quad 1 \le i \le n-m+1. \tag{1}$$

When $m=2$, $X_2(i) = [[1,2],[2,3],\cdots,[n-1,n]]$, $1 \le i \le n-m+1$. When $m=3$ $X_3(i) = [[1,2,3],[2,3,4],\cdots,[n-2,n-1,n]]$, $1 \le i \le n-m+1$. An array $X_j(i)$ of dimension j is compared to another vector of the same dimension. If the differences in the absolute values of their corresponding points are all less than the threshold r, then the two arrays are determined to be similar. If the differences in the absolute values of their corresponding points are all greater than the threshold r, they are not similar. Typically, $r = (0.1 - 0.25) \times \mathrm{SD}$, where SD is the standard deviation of the time series composed of sample point data

(3) The average probability that all vectors in the $n-m+1$ arrays $X_m(i)$ of dimension m are similar is as shown in the following equation:

$$\varphi_m = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} X_m(i) \tag{2}$$

(4) Calculate the average probability: the dimension is increased to $m+1$, and the above steps are repeated to obtain the sample entropy. It is shown in the following equation:

$$SE = -\ln \frac{\varphi_m}{\varphi_{m+1}} \tag{3}$$

1.1.3. Ensemble Empirical Mode Decomposition (EEMD) Algorithm. The EEMD [20] has good adaptability in processing nonstationary nonlinear signals, and its algorithmic decomposition steps are as follows:

RETRACTED

(1) Suppose the gait signal collected by the human posture estimation algorithm is $x(t)$. The sequence of white noise added for i times is $n_i(t)$. After adding white noise, the gait signal is shown in the following equation:

$$x_i(t) = x(t) + n_i(t). \quad (4)$$

$x_i(t)$ is the gait signal after adding white noise for the i^{th} time

(2) The empirical mode decomposition (EMD) [21] decomposition of the signal x_i after adding white noise is performed to obtain the intrinsic mode function (IMF) at different scales. It is shown in the following equation:

$$x_i(t) = \sum_{i=1}^{n} C_{i,n}(t) + r_{i,n}(t), \quad (5)$$

where n is the number of intrinsic mode functions obtained after the decomposition of the signal $x_i(t)$, $r_{i,n}(t)$ is the residual component, and $C(t)$ is the corresponding IMF

(3) The process of applying (4) and (5) is repeated for each addition of white noise. The j^{th} IMF component of $x_i(t)$ decomposed by EMD is obtained. It is shown in the following equation:

$$x_{i,j}(t) = \sum_{j=1}^{n} C_{j,n}(t) + r_{j,n}(t) \quad (6)$$

(4) The final EEMD result is obtained by averaging all the obtained IMF components as in the following equation:

$$C_n(t) = \frac{1}{N}\sum_{k=1}^{N} C_{j,n}(t) \quad (7)$$

1.2. Experiment and Analysis

1.2.1. Experimental Procedure, Platform, and Subjects. The experimental procedure is shown in Figure 2. First, the human posture estimation algorithm extracts the trajectory signal of human walking joint points in the video, and this signal is normalized. The sample entropy is then calculated for the joint trajectory signal. Finally, an independent samples t-test is used to distinguish between healthy subjects and patients.

The experimental platform uses AlphaPose to extract the joint motion trajectories from the captured gait videos. The joint trajectory signal is uploaded to a server equipped with an Nvidia model 2080Ti GPU (graphics processing unit) for feature extraction. Finally, statistical analysis is used to distinguish between the gaits of patients with bone and joint diseases and healthy people.

An orthopedist provided the experimental data, and each test subject was informed and consented to the experiment. Twelve patients with knee arthritis were selected as the study group and 12 healthy individuals as the control group. There was no statistically significant difference between the study and control groups in general characteristics, such as gender, age, or body mass index.

1.2.2. Model Acceleration of Human Pose Estimation Algorithm Based on TensorRT. TensorRT is a neural network inference engine that maximizes inference throughput and efficiency. The programmability of CUDA (Compute Unified Device Architecture) enables TensorRT to address the increasingly diverse and complex trends of deep neural networks. It is possible for TensorRT to automatically optimize the trained neural network to ensure the algorithm's accuracy and increase its speed. The accelerated reasoning process of the attitude estimation algorithm model based on TensorRT is shown in Figure 3.

In the TensorRT inference acceleration, the deep learning algorithm trained on the computer side must be transformed first. Then, the convolutional neural network of the improved model is optimized using the TensorRT analytic model, combining parameters such as precision and target deployment GPU. The optimized engine can be serialized to memory and then turned into an engine to accelerate the reasoning speed by deserialization to obtain the final prediction results. To facilitate the collection and analysis of gait videos, TensorRT is used in this paper to accelerate the processing of the human pose estimation network AlphaPose. The verification set is from MS COCO val2017 [22], and the results are shown in Table 1.

The mean average precision (mAP) of the AlphaPose_trt human pose estimation algorithm model after acceleration through TensorRT inference is 71.74%, which is unchanged compared with the mAP before acceleration. This result shows that the human pose estimation algorithm model's inferred acceleration does not affect the joint points' detection accuracy. AlphaPose_trt provides a significant improvement in inference speed and ensures human joint extraction accuracy during gait analysis. It is easy to deploy in embedded devices.

AlphaPose mainly consists of the target detection algorithm YOLOv3 and the pose estimation algorithm. Table 2 records the inference time and the throughput of the YOLOv3 model for different batch sizes. Table 2 shows that the batch size affects the YOLOv3 target detection model results, such as latency and throughput, which increase with increasing batch size. The inference speed of YOLOv3 based on TensorRT is faster than the original target detection model using YOLOv3 alone. Table 3 records the inference time and the throughput of the pose estimation model for human joint point extraction for different batch sizes.

The algorithm model AlphaPose_trt optimized by TensorRT inference has lower latency, higher throughput, and faster inference speed than the original model. The results of the AlphaPose inference optimization are shown in Table 4.

When the input batch size is 1, the latency metric of the TensorRT-optimized AlphaPose_trt inference drops by

RETRACTED

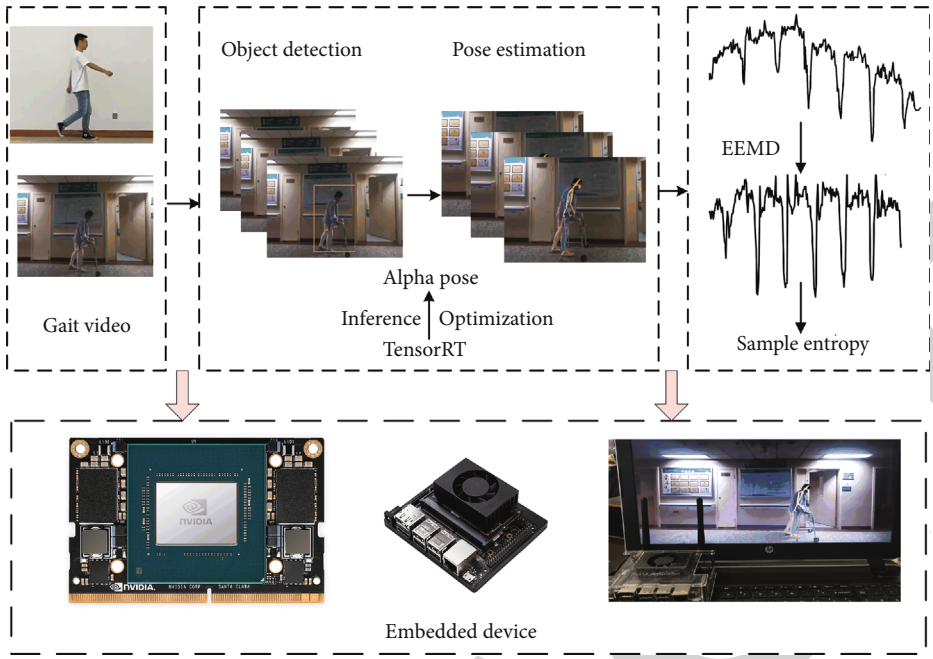


Figure 2: The process of gait analysis method.

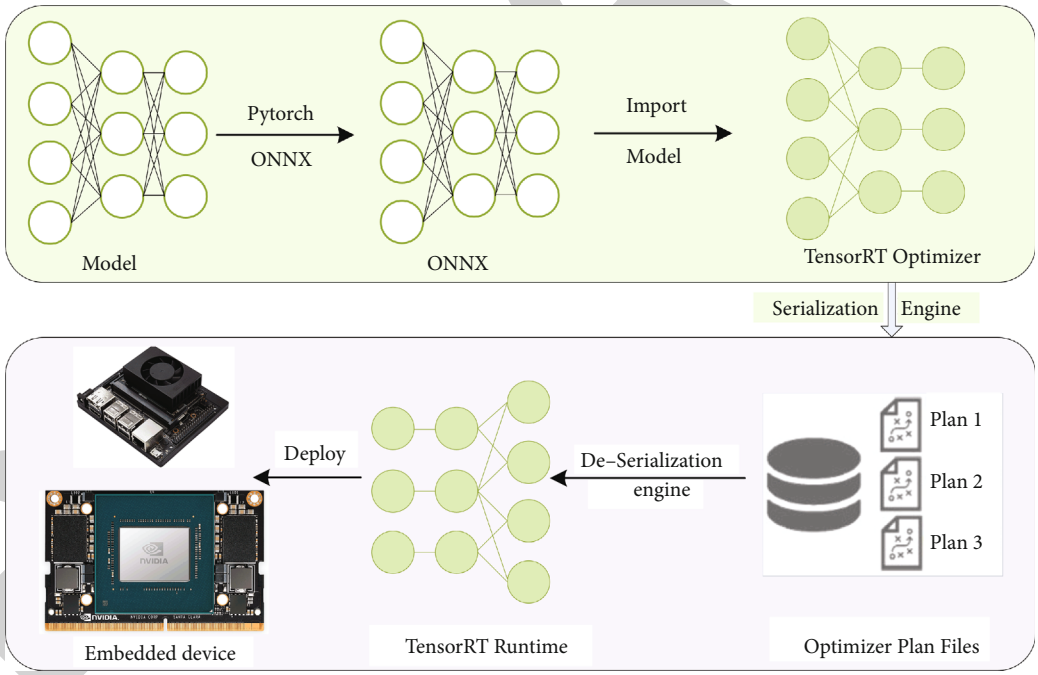


Figure 3: TensorRT inference optimization process.

Table 1: Comparison of accuracy of AlphaPose before and after inference acceleration.

Method	mAP (%)
AlphaPose	71.74
AlphaPose_trt	71.74

18.49 ms, while the throughput metric increases by 37.86. When the input batch size is 2, the latency indicator of AlphaPose_trt decreases by 28.18 ms, and the throughput increases by 63.86 compared to the original AlphaPose. When the input batch size is 4, the latency indicator of AlphaPose_trt drops by 37.70 ms, and the throughput increases by 97.19 compared to the original AlphaPose. These results show that AlphaPose_trt can achieve faster inference speeds without loss of accuracy.

Table 2: Comparison of the results of YOLOv3 inference optimization.

Mode	Batch size	Latency (ms)	Throughput
YOLOv3	1	24.49	40.83
	2	37.98	52.66
	4	53.48	74.79
YOLOv3_trt	1	13.71	72.94
	2	17.59	113.70
	4	22.96	174.22

Table 3: Comparison of the results of inference optimization by pose estimation.

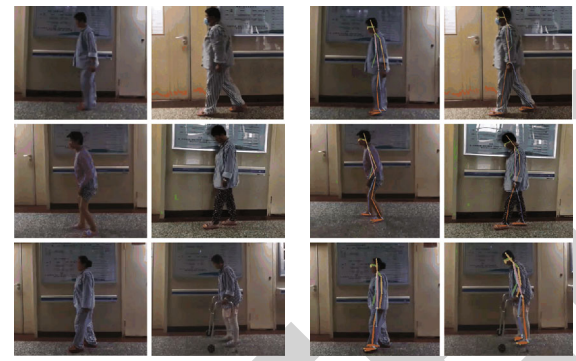
Mode	Batch size	Latency (ms)	Throughput
Pose estimation	1	8.71	114.81
	2	8.99	222.46
	4	9.08	440.53
Pose estimation_trt	1	1.00	1000.00
	2	1.20	1666.67
	4	1.86	2150.54

Table 4: Comparison of results of AlphaPose inference optimization.

Mode	Batch size	Latency (ms)	Throughput
AlphaPose	1	33.20	30.12
	2	46.97	42.58
	4	62.52	63.97
AlphaPose_trt	1	14.71	67.98
	2	18.79	106.44
	4	24.82	161.16

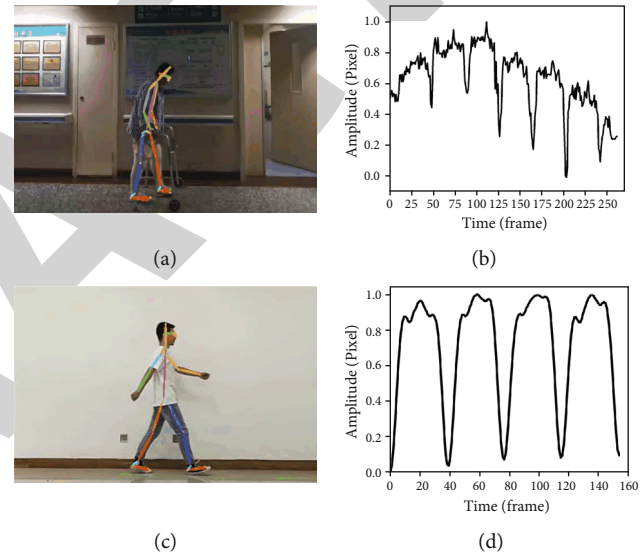
The optimization effect of TensorRT is related to the GPU hardware performance. Batch size refers to the number of samples processed by the GPU during inference. Table 4 shows that as the batch size increases, the average latency of the processed samples decreases, and the inference time shortens. In theory, better performance hardware devices such as GPUs permit a larger batch size parameter to be set. At the same time, GPU utilization is improved when the algorithm performs inference, and the inference speed is accelerated. However, for hardware-limited devices such as the GPU, batch sizes that are too large can cause the display memory to be exceeded when the algorithm processes longer gait videos. This factor prevents the algorithm from running smoothly. After the experiments, it was found that when the batch size is four, the algorithm is the most stable for gait video inference and can significantly accelerate.

1.2.3. Joint Point Extraction Based on Human Posture Estimation. The human gait video is used to extract the human skeleton and joint points through AlphaPose. The results are shown in Figure 4.



(a) (b)

Figure 4: Graph of human posture estimation results. (a) The captured human gait video. (b) The detection result of AlphaPose_trt.



(a) (b) (c) (d)

Figure 5: Signals of heel joint point motion trajectories in both groups. (a) Patient pose estimation. (b) Patient's heel joint trajectory signal. (c) Normal pose estimation. (d) Normal heel joint trajectory signal.

Figure 4(a) shows the gait video of some patients in the study group. Some patients with severe disease cannot walk upright and must use a walker, but walkers can easily obscure the body's joints, resulting in missed joint point detection. The complexity of the surrounding environment can also cause adverse effects. AlphaPose_trt can quickly and accurately detect human joints in the video stream, avoiding the influence of a complex environment and ensuring feature extraction accuracy for human joint trajectories, as shown in Figure 4(b). The trajectory of joint points is obtained, transformed into a one-dimensional signal, and normalized. The results are shown in Figure 5, taking the heel joint as an example.

During the video recording process, there may be complex external environmental effects such as slight lens shifts or shaking, resulting in regular upward or downward drifting of the extracted joint trajectory signal. EEMD is used

RETRACTED

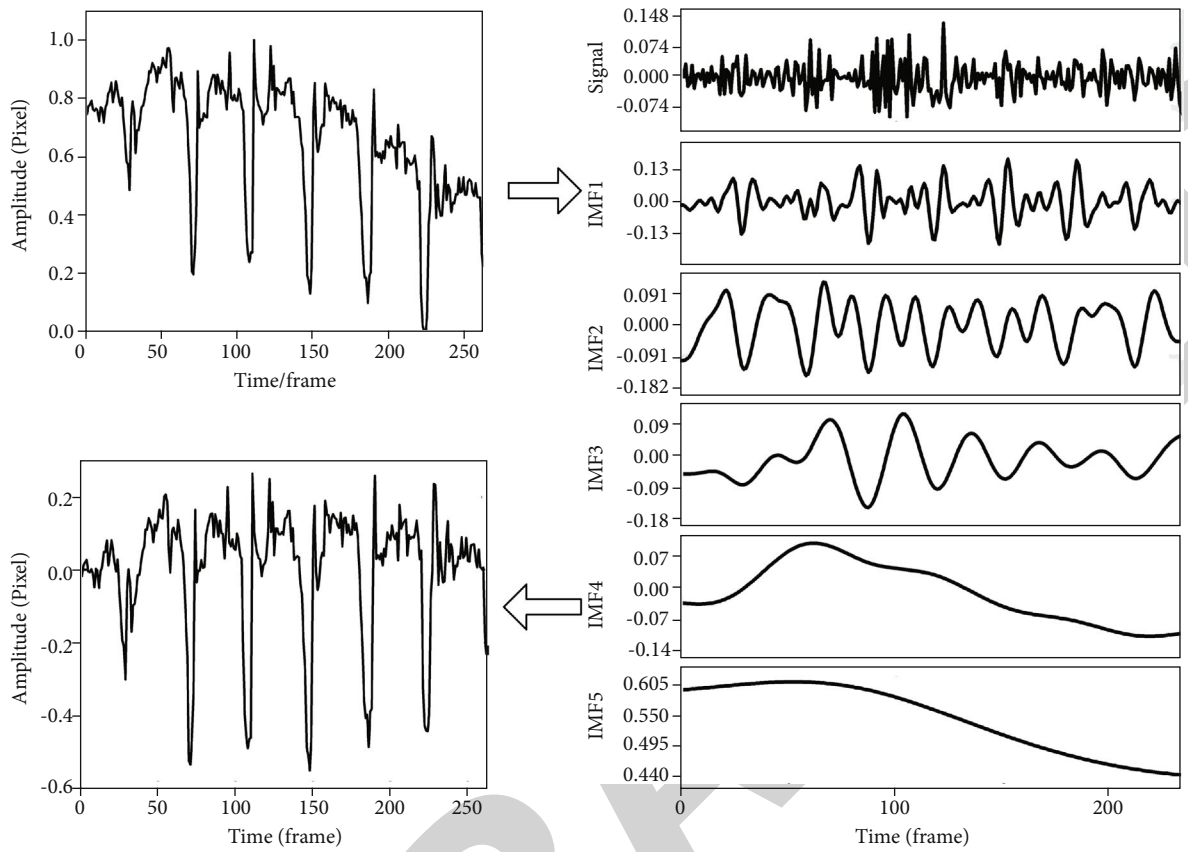


Figure 6: The debaseline drift of heel joint point trajectory signal.

Table 5: Results of the analysis of difference between the two populations.

Sample entropy	Study group	Control group
SE_{ankle}	1.07±0.26	1.49±0.27**
SE_{heel}	0.76±0.20	1.37±0.27**

to deconstruct the original signal to obtain multiple IMF components and residuals. Each IMF component is added to obtain the joint point trajectory signal with baseline drift removed for sample entropy calculation. The debaseline drift process is shown in Figure 6.

1.2.4. Calculation of Sample Entropy of Joint Trajectory Signal. The extracted joint trajectory signals are characterized using sample entropy. The sample entropy is used to distinguish the joint point trajectory signal degree of confusion between the two groups and quantify the joint point trajectory signal. The study group and the control group are accurately distinguished. This method can help doctors judge the rehabilitation status of patients.

SPSS (Statistical Package for the Social Sciences) was used for the statistical analysis. An independent samples t -test was used to determine significant differences between the healthy and patient populations. A value of $p < 0.05$ was considered to be statistically significant. There was a difference in entropy values between the study and control groups. The entropy values of the study and control groups can be considered statistically significantly different when $p < 0.01$. The results are shown in Table 5.

In Table 5, ** indicates a significant difference ($p < 0.01$). SE_{ankle} is the ankle joint point signal sample entropy. The range of sample entropy values is 1.07±0.26 for healthy subjects and 1.49±0.27 for patients. Therefore, significant differences exist between the two groups. SE_{heel} is the sample entropy of the signal at the heel joint point, where the range is 0.76±0.20 for healthy individuals and 1.37±0.27 for patients. There were significant differences in SE_{ankle} and SE_{heel} between the two groups ($p < 0.01$).

1.2.5. Analysis of Experimental Results. The patients were all suffering from joint diseases of the lower extremities, so the experiments used mainly these joints because patients' gaits would be affected by related pain or other factors. SE_{ankle} and SE_{heel} in the study and control groups were significantly different ($p < 0.01$). The main reason for this difference is that patients walk carefully to avoid accidents such as falls because of the pain in the affected area and poor balance. This behavior led to a more complex gait while walking, with more significant curve fluctuations and correspondingly larger sample entropy values. Therefore, the sample entropy values of the patient's heel and ankle signals were higher than normal.

RETRACTED

2. Conclusion

Patients rehabilitating from joint surgery need professional equipment and rehabilitation judgments by doctors. The process is complicated and tedious, and the cost is high. Therefore, gait analysis through video collection has important practical significance. TensorRT has been used to optimize the inference of AlphaPose, reducing the runtime latency of the algorithm and improving its throughput. The experimental results show that TensorRT can increase latency and throughput by about 2.5 times, facilitating subsequent algorithms in embedded devices.

In this paper, the sample entropy algorithm is combined to analyze the gait of two groups of people based on human pose estimation. The algorithm quantifies the joint point trajectory signals of the study and control groups and judges whether the entropy values of the two groups' joint point trajectory signals are statistically different. The results showed significant differences in the trajectory signal sample entropies of the heel joint and ankle joint between the study and control groups ($p < 0.01$), thus distinguishing the two groups of people. Applying this method in rehabilitation judgment of joint diseases is expected to have high clinical value.

Data Availability

Data processing was supported by Ningxia Technology Innovative Team of advanced intelligent perception and control.

Conflicts of Interest

The authors declare that there is no conflict of interest that exists in the submission of this manuscript.

Acknowledgments

This research was funded by the Natural Science Foundation of Ningxia (No. 2022AAC03006, No. 2022AAC03244), National Natural Science Foundation of China (No. 61861001), Ningxia Technology Innovative Team of Advanced Intelligent Perception and Control, and Ningxia Postgraduate Education Reform Project (No. 2021-34).

References

[1] M. Zhang, J. L. Theleman, K. A. Lygrisse, and J. Wang, "Epigenetic mechanisms underlying the aging of articular cartilage and osteoarthritis," *Gerontology*, vol. 65, no. 4, pp. 387–396, 2019.

[2] J. Popko, M. Kwiatkowski, and M. Gałczyk, "Scoliosis: review of diagnosis and treatment," *Polish Journal of Applied Sciences*, vol. 65, no. 4, pp. 31–35, 2018.

[3] S. Song, R. Shi, L. Lin et al., "A study on gait characteristics in patients with lumbar disc herniation," *Chinese Journal of Rehabilitation Medicine*, vol. 35, no. 3, pp. 306–312, 2020.

[4] M. Yin, Q. Li, W. Wang, P. Yao, and Z. Qi, "6-DOF gait changes of knee osteoarthritis in different periods," *Chinese Journal Of Rehabilitation Medicine*, vol. 33, no. 33, pp. 1341–1343, 2018.

[5] X. Lv, S. Wang, T. Chen et al., "Human gait analysis method based on sample entropy fusion AlphaPose algorithm," in *2021 33rd Chinese Control and Decision Conference (CCDC)*, pp. 1543–1547, Yunnan, China, May 2021.

[6] H. Liu, M. Hou, W. Huang, R. Lin, and Z. Jiang, "Current status of research on three-dimensional gait analysis for chronic nonspecific low back pain," *Theory and Practice of Rehabilitation in China*, vol. 25, no. 8, pp. 882–885, 2019.

[7] F. Yang, X. Wang, M. Hou et al., "Three dimensional gait analysis and comparison of gait characteristics between young and old people under dual task," *Chinese Journal Of Tissue Engineering*, vol. 25, no. 3, pp. 344–349, 2021.

[8] J. Cuadrado, F. Michaud, U. Lugrís, and M. P. Soto, "Using accelerometer data to tune the parameters of an extended Kalman filter for optical motion capture: preliminary application to gait analysis," *Sensors*, vol. 21, no. 2, article 427, 2021.

[9] A. K. Seifert, M. G. Amin, and A. M. Zoubir, "Toward unobtrusive in-home gait analysis based on radar micro-Doppler signatures," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2629–2640, 2019.

[10] M. M. Rahaman, C. Li, Y. Yao et al., "DeepCervix: a deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," *Computers in Biology and Medicine*, vol. 2021, pp. 1–12, 2021.

[11] S. Ai, C. Li, X. Li et al., "A state-of-the-art review for gastric histopathology image analysis approaches and future development," *BioMed Research International*, vol. 2021, 19 pages, 2021.

[12] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.

[13] L. Xu, J. Chen, F. Wang, Y. Chen, W. Yang, and C. Yang, "Machine-learning-based children's pathological gait classification with low-cost gait-recognition system," *Biomedical Engineering Online*, vol. 20, no. 1, pp. 1–19, 2021.

[14] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[15] Y. Xiu, J. Li, H. Wang, Y. Fang, and C. Lu, "Pose flow: efficient online pose tracking," 2018, http://arxiv.org/abs/1802.00977.

[16] H. Fang, S. Xie, Y. Tai, and C. Lu, "RMPE: regional multi-person pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2334–2343, Venice, Italy, Oct 2017.

[17] H. B. Yoo, M. S. Park, and S. H. Kim, "Real time face detection method using TensorRT and SSD," *KIPS Transactions on Software and Data Engineering*, vol. 9, no. 10, pp. 323–328, 2020.

[18] S. Wang, J. Zhao, N. Ta, X. Zhao, M. Xiao, and H. Wei, "A real-time deep learning forest fire monitoring algorithm based on an improved pruned+ KD model," *Journal of Real-Time Image Processing*, vol. 18, no. 6, pp. 2319–2329, 2020.

[19] Y. Ma, H. Li, Y. Liang, G. Zhao, and X. Gao, "An empirical study of the sample entropy properties of body balance and sit-up-walk ground reaction force in the elderly," *Integrated Technology*, vol. 5, no. 6, pp. 52–61, 2016.

RETRACTED

Hindawi
BioMed Research International
Volume 2022, Article ID 9623173, 9 pages
https://doi.org/10.1155/2022/9623173

Research Article

Habitat Analysis of Breast Cancer-Enhanced MRI Reflects BRCA1 Mutation Determined by Immunohistochemistry

Tianming Du and Haidong Zhao

Department of Breast Surgery, Second Affiliated Hospital of Dalian Medical University, 467, Zhongshan Road, Shahekou District, Dalian, Liaoning 116023, China

Correspondence should be addressed to Haidong Zhao; z.hddl@hotmail.com

Received 26 February 2022; Accepted 15 March 2022; Published 30 March 2022

Academic Editor: Chen Li

Copyright © 2022 Tianming Du and Haidong Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To use habitat analysis (also termed habitat imaging) for classifying untreated breast cancer-enhanced magnetic resonance imaging (MRI) in women. Moreover, we intended to obtain clustering parameters to predict the BReast CAncer gene 1 (BRCA1) gene mutation and to determine the use of MRI as a noninvasive examination tool. *Methods*. We obtained enhanced MRI data of patients with breast cancer before treatment and selected some sequences as the source of habitat imaging. We used the *k*-means clustering to classify these images. According to the formed subregions, we calculated several parameters to evaluate the clustering. We used immunohistochemistry to detect BRCA1 mutations. Moreover, we separately determined the ability of these parameters through independent modeling or multiple parameter joint modeling to predict these mutations. *Results*. Of all extracted values, separation (SP) demonstrated the best prediction performance for a single parameter (area under the receiver operating characteristic curve (AUC), 0.647; 95% confidence interval (CI), 0.557–0.731). Simultaneously, models based on the Calinski-Harabasz Index and sum of square error performed better in the training (AUC, 0.903; 95% CI, 0.831–0.96) and verification (AUC, 0.845; 95% CI, 0.723–0.942) sets for multiparameter joint modeling. *Conclusion*. Based on the enhanced MRI of breast tumors and the subregions generated according to the habitat imaging theory, the parameters extracted to describe the clustering effect could reflect the BRCA1 status. Differences between clusters, including the general differences of cluster centers and clusters and the similarity of samples within clusters, were the embodiment of this mutation. We propose an algorithm to predict the BRCA1 mutation of a patient according to the enhanced MRI of the breast tumor.

1. Introduction

In 2020, the number of new breast cancer cases worldwide was 2.26 million, and breast cancer has replaced lung cancer as the most common cancer type. Moreover, breast cancer is the most common cancer leading to death in women, both in developing and developed countries [1]. Like all cancers, it is considered a heterogeneous type, and the biological behavior, treatment response, and clinical outcomes are closely related to the molecular typing of breast cancer determined by the estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and KI-67. Breast cancer is categorized by the following molecular classification: Luminal A, Luminal B, HER2 overexpression, and triple negative. The risk factors of breast cancer with different molecular classifications demonstrate considerable differences. For example, triple-negative breast cancer often displays a younger age of onset, compared with the most common Luminal A breast cancer. Of several risk factors, genetic susceptibility is the most important factor and is often associated with race [2]. It can become explicit with the environmental background and heredity. According to statistics, 20% to 40% of hereditary breast cancers can be attributed to harmful mutations in BReast CAncer gene 1 (BRCA1) and BReast CAncer gene 2 (BRCA2) [3]. In addition, 90% of BRCA1-associated breast cancers reveal negative ER expression [4]. Meanwhile, PR expression in BRCA1-associated tumors is lower than that in sporadic

tumors [5], thus making breast tumors in patients with BRCA1 mutations more likely to be classified as triple-negative breast cancer, characterized by the loss of ER and PR expression and HER2 overexpression. Patients with this molecular subtype are considered to have poor prognosis. Triple-negative breast cancer has a shorter survival than other subtypes, with a 40% mortality within 5 years of diagnosis. Furthermore, it is highly invasive biologically, with distant metastasis occurring in approximately 46% of patients with triple-negative breast cancer. Following metastasis, the median survival time is only 13.3 months, and the postoperative recurrence rate is 25% even after surgery. In addition to the above-mentioned clinical manifestations [6, 7], BRCA1 mutations are pathologically associated with basal-like phenotypes in triple-negative breast cancer [8, 9].

The BRCA gene is involved in double-stranded DNA damage repair through homologous recombination; therefore, women with this mutation lack the tumor suppressor protein to repair damaged DNA [10]. BRCA proteins are essential for chromosomal stability, and their primary function is to protect the genome from damage. A recent study further suggests that BRCA transcription regulates numerous genes involved in DNA repair, cell cycle, and apoptosis. These functions are mediated by several cellular proteins that interact with BRCA, besides an association with phosphorylation events. Women with BRCA mutations have a higher risk of breast cancer at a younger age than those without [11]. Furthermore, this mutation is associated with an increased risk of ovarian cancer [12].

Patients with BRCA-associated breast cancer have a higher risk of ipsilateral and contralateral breast cancer recurrence with worse prognosis, compared with sporadic breast cancer. Despite numerous distinct clinical particularities of BRCA-associated breast cancer, current treatment recommendations are substantially similar to those for sporadic breast cancer, including surgery, radiation, and chemotherapy. However, BRCA1/2-associated breast cancer is accompanied by a defect of homologous recombination repair function, which may be more sensitive to DNA-damaging drugs, such as platinum drugs or poly-ADP ribose polymerase inhibitors. With increased tumor specificity, chemotherapy regimens may be used more for patients with BRCA mutations [13].

Immunohistochemistry (IHC) is an inexpensive and effective technique easily available to the majority of pathologists. It has been considerably used in several tumors. Researchers have made substantial progress in detecting BRCA1 mutations in digestive system cancer to predict its prognosis in response to chemotherapy [14]. Some studies have also suggested the possibility of using IHC in breast tumors [15].

Enhanced magnetic resonance imaging (MRI) is a sensitive breast imaging detection method [16]. Compared with mammary gland molybdenum target image and ultrasound, enhanced MRI has higher resolution and can observe the tissue perfusion status. Clinicians can use different functional MRI sequences to measure spatial differences in the cell density, tissue structure, perfusion, and metabolism. For people aged 25 years to 29 years with BRCA mutations, the NCCN guidelines recommend annual breast MRI screening or annual mammography [17]. Currently, there are some MRI studies on BRCA mutations in patients with breast cancer. The primary research method is based on first-order features and features extracted from gray-level cooccurrence matrix, gray-level run-length matrix, and gray-level size zone matrix. Those studies combined the features with the clinical information of patients to establish model [18, 19]. However, these methods predominantly explore the BRCA mutation imaging on the entire tumor but do not consider the tumor heterogeneity reflected by imaging.

Habitat analysis, also termed habitat imaging, is an imaging technology aimed at capturing subtle differences among tumors. It uses several algorithms to segment the tumor and its surrounding environment, aiming to obtain subregions which reflect the heterogeneity of the tumor [20]. In other words, this image-processing technique focuses on similar subregions within the tumor formed by the pressure of survival. In this study, we used the cluster parameters extracted from the habitat imaging of patients with enhanced MRI to predict BRCA1.

2. Materials and Methods

2.1. Study Participants. The radiology database of the Second Affiliated Hospital of Dalian Medical University was reviewed. We identified 187 patients who underwent MRI from March 2018 to June 2021. Fifty-one and 11 patients were excluded owing to an incomplete image sequence and the absence of distinct BRCA1 results. Therefore, a total of 125 patients were included in this study. The inclusion criteria were as follows: (i) no breast diseases before imaging examination and (ii) did not receive treatment for breast cancer or may artificially change breast imaging. The patients underwent MRI examinations before the surgery and were diagnosed as grades 3, 4, and 5 according to the Breast Imaging Reporting and Data System. Surgery or biopsy was performed within 1 week to confirm primary breast cancer diagnosis and BRCA1 mutation. We excluded specific breast malignancies, such as inflammatory breast cancer, Paget's disease, and breast cancer because of metastasis. Moreover, men and pregnant women were excluded. There was no specific information about patients in the study, so the study did not involve ethical issues.

2.2. Enhanced MRI. We used the American GE1.5 T Signa HDxt MRI scanner, and the receiving coil was a special one for the surface of the breast. In the prone position, the bilateral breast naturally hung in the concave hole of the coil. The scan sequence and parameters were as follows: diffusion-weighted imaging, b – value = 800 s/mm, repetition time (TR) = 5,600 ms, echo time (TE) = 74.4 ms, matrix 130×128, field of view (FOV) = 33 cm × 33 cm, and layer thickness 5 mm. All patients underwent dynamic contrast-enhanced MRI following DWI sequence scanning. Gadolinium diamine was used as the contrast agent. The injection volume was 0.2 mmol/kg, and the flow rate was 2 ml/s to 3 ml/s. Following injection, 20 ml normal saline was used to flush the tube. We performed continuous nonstop

scanning. T1-weighted image plain scanning was initially performed. Following gadolinium injection, we continuously scanned nine phases, each phase of 47 s. A total of 10 phases were scanned. The scanning time was 7 min 6 s, and the turning angle was 15°. Other scanning parameters were as follows: TR = 5.1 ms, TE = 2.5 ms to 12 ms, matrix 320 × 384, FOV = 30.2 cm × 30.2 cm, and layer thickness 5 mm. Two radiologists with >10 years of experience in breast imaging diagnosis independently interpreted the MRI results. After discussing the images, they reached a diagnostic consensus. All data have been transferred to the GE workstation (Advantage Windows 4.5, General Electric, Madison, WI, USA).

2.3. Habitat Generation. The Pydicom Library of Python was used to read the spatial location of the image. First, we used the nearest neighbor interpolation to process the images into pictures with equal pixel spacing, followed by image registration according to the spatial location. The region of interest (ROI) was manually obtained on a high-signal DWI, and there was no necrosis or cystic component under ideal circumstances. In the case of no satisfactory image (usually because of low resolution), the ROIs were drawn by referencing T1 images. According to the 3-D MRI images, we developed a segmentation algorithm with wide adaptability. First, we applied the Markov random field segmentation method to voxel data combined with the interaction between voxels, considering 27 individual elements around each voxel ($3 \times 3 \times 3$). Images of the DWI sequence, T1 sequence, OutPhase T2 sequence, InPhase T2 sequence, WATER T2 sequence, and FAT T2 sequence were standardized according to the gray value in ROIs to form new images to be clustered [21]. For each voxel in the ROI, the gray value in each image was used as a six-dimensional coordinate and the proportion of each weight was similar. The K clustering k-means algorithm was used for image segmentation (Figure 1) [22] in the following steps: (1) we selected K points as the initial cluster center; (2) for each sample in the dataset, we calculated its Euclidean distance to the K cluster centers and divided it into the class corresponding to the center with the smallest distance; (3) for each category, we recalculated the cluster center (i.e., the centroid of all samples belonging to the category); and (4) we repeated steps 2 and 3 until reaching a termination condition, such as iteration times and minimum error change. The final clustering results were evaluated using the maximum expectation. For the processed 3-D tumor ROI, we divided each tumor region into three spatially limited subregions with different characteristics using the aforementioned algorithm [23].

2.4. Immunohistochemistry. Pathological specimens were obtained from the enrolled patients. No data were excluded owing to missing values or ambiguity. All specimens were fixed with 4% neutral formaldehyde, embedded in paraffin, and continuously sectioned at a distance of 4 μm. Following IHC staining, the specimens were observed and photographed under a microscope. We used the immunohistochemical SP method to detect the expression of susceptible genes. Specific steps were performed according to the standard instructions, and professional pathologists interpreted the films. We determined the comprehensive staining intensity and the percentage of positive cells. The final results were divided into the following categories: <5% visible staining: -; 5% to 25% visible staining: +; 26% to 50% visible staining: ++; and >50% visible staining: +++, where - was defined negative and +, ++, and +++ were defined positive.

2.5. Clustering Features. The clustering index was extracted from the clustering method to the following five items:

(1) Inertia, within cluster sum of square error

For each cluster, we calculated the distance between the samples in the cluster and the cluster center and added the values to describe the similarity between each cluster.

(2) Calinski-Harabasz Index

The Calinski-Harabasz Index is defined as the ratio of discrete between groups to that within groups. The larger the value of the index, the smaller the distance within the cluster. The greater the distance between clusters, the better the clustering effect. The distance within the cluster is represented by that between the sample point within the cluster and the center point of the cluster.

(3) Silhouette coefficient

The silhouette coefficient was used to describe the cohesion and separation of clusters and defined the average distance between the point and other points within the cluster. The final silhouette coefficient is the average of silhouette coefficient of each sample. Its value ranges from -1 to 1, and the value is approximately 1 for a considerably greater separation degree between clusters than the cohesion. The closer the value to 1, the better the clustering effect.

(4) Separation

The separation value was obtained by calculating the arithmetic average of the sum of distances between each two cluster center points. In contrast to compactness, the indicator only considered the distance between different clusters. The larger the number, the better the clustering effect.

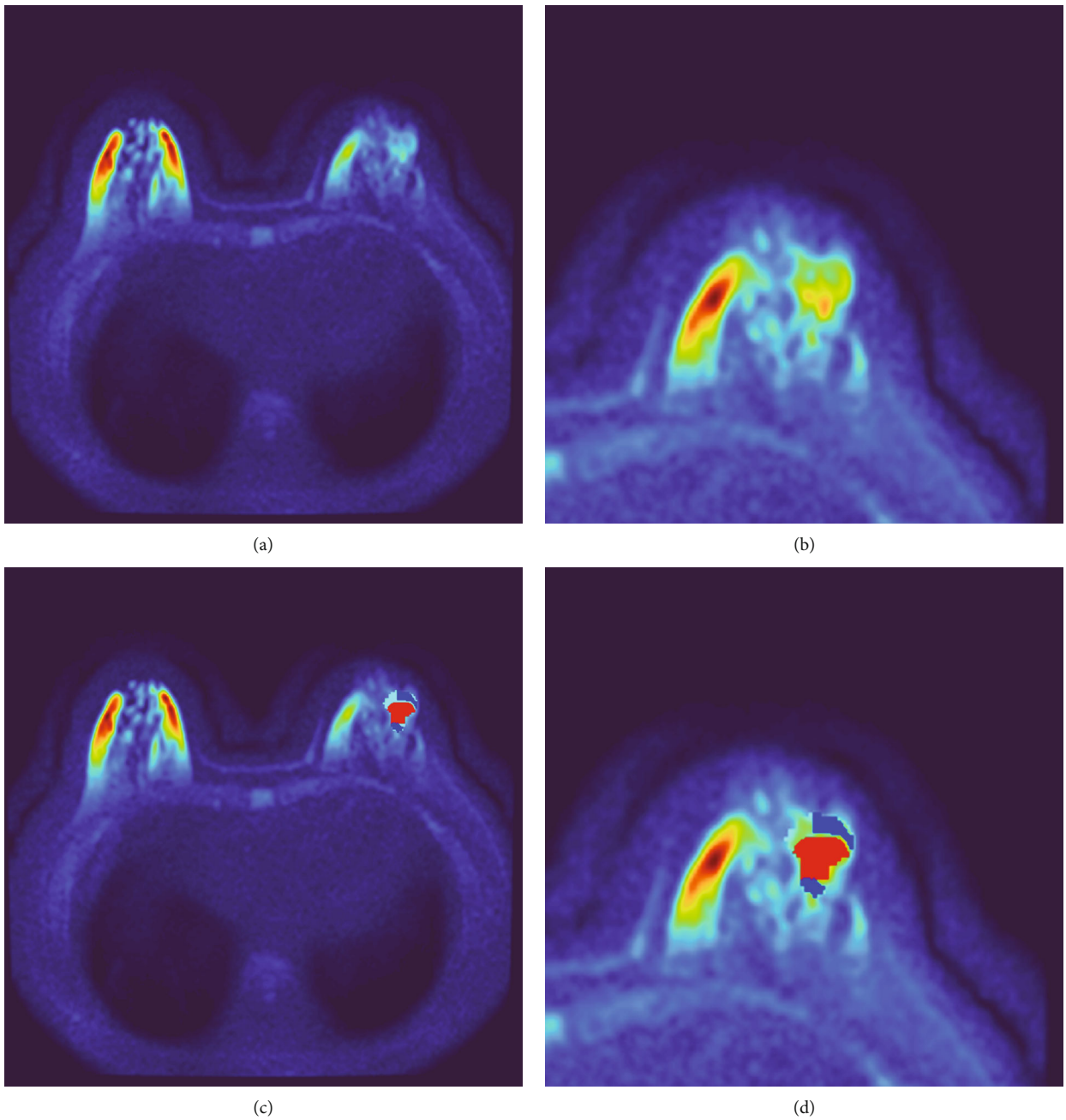
(5) Davies-Bouldin Index

The greater the distance between clusters and the smaller the distance within the cluster, the smaller the value is. The smaller the value of the Davies-Bouldin Index, the better the clustering performance.

3. Results

3.1. Clinical Features. The patient's clinical information is as follows (Table 1).

3.2. Features and Prediction Model. SP demonstrated excellent predictive power for the clustering features (area under



(a) (b) (c) (d)

Figure 1: An example of habitat imaging. (a) A pseudocolor map of a 45-year-old patient with breast cancer using diffusion-weighted imaging. Following image diagnosis, the breast tumor (considered malignancy) has been confirmed by postoperative pathology. (b) The picture only depicts the segment containing the tumor. (c) The graph depicts three different subregions (three colors) following habitat analysis. (d) The picture only depicts the segment containing the tumor.

the receiver operating characteristic curve (AUC), 0.647; 95% confidence interval (CI), 0.557–0.731) (Figure 2).

Then, we take the AdaBoost algorithm for modeling, the weak classifier is Gaussian process, and we select two characteristics as the modeling parameters, respectively, Calinski-Harabasz Index and Inertia. The prediction model also displayed good diagnostic efficiency (training group: AUC, 0.903; 95% CI, 0.831–0.96; accuracy, 0.897; specificity, 0.72; and sensitivity, 0.968; test group: AUC, 0.845; 95% CI, 0.723–0.942; accuracy, 0.789; specificity, 0.636; and sensitivity, 0.852) (Figure 3).

We plotted the calibration curve for the prediction model (Figure 4). Our prediction model displayed stability and considerable prediction ability.

4. Discussion

In this study, we classified the breast cancer image into three subregions by using the gray distribution of multiple sequences of images of patients. After this, we use various parameters to identify BRCA1 mutations, which can assist pathological diagnosis. For example, despite a large staining

Table 1: Clinical characteristics of patients with breast cancer.

Characteristics	Training group ($n = 100$)	Test group ($n = 25$)
Age (years)	55.2	56.1
ER		
Positive	80	19
Negative	20	6
PR		
Positive	67	18
Negative	23	7
HER2		
Positive	31	5
Negative	69	20
Stage I	69	20
Stage II	23	4
Stage III	5	1
Stage IV	3	0
Lymphonodus		
Positive	30	8
Negative	70	17
BRCA1		
Positive	69	20
Negative	31	5

ER: estrogen receptor; PR: progesterone receptor; HERK2: human epidermal growth factor receptor 2; BRCA1: BReast CAncer gene 1.

area that can be recognized by naked eyes, the degree of staining is shallow and difficult to count by equipment. In this case, it is difficult for pathologists to diagnose BRCA1 by only pathological slices, and this method can provide an auxiliary diagnostic basis.

BRCA1 mutations affect the incidence, progression, diagnosis, treatment, and prognosis of breast cancer throughout the course of the disease [24]. Simultaneously, BRCA mutation-associated breast cancers are more likely to develop lymph node metastases. In addition to macroscopic changes in clinical manifestations, such as morbidity and prognosis in terms of tumor heterogeneity, tumors with BRCA1 mutation usually display invasive ductal carcinoma pathologically and have a higher incidence of myeloid carcinoma and atypical myeloid carcinoma [25]. Despite their role in the auxiliary diagnosis of breast cancer, it is relatively difficult to make a preliminary judgment only by imaging. In ultrasound examination, BRCA1 and BRCA2-mutated breast cancer demonstrates an irregular shape and blurred boundary hypoechoic masses. However, unlike BRCA2, 29.5% of BRCA1 mutations display posterior echo enhancement, usually in benign breast lesions, such as fibroadenomas or cysts. The aforementioned feature may lead to breast cancer with BRCA1 mutation being misdiagnosed as a benign tumor to some extent [26]. Invasive breast cancers with BRCA1 mutations do not reveal more calcification on breast cancer-specific mammography. In terms of the medical information obtained, MRI is a significantly better imaging method than ultrasound and molybdenum target. However, it partially addresses the problem of BRCA1 displaying benign tumor characteristics in ultrasound. This is because a radiologist can obtain additional information in MRI, including internal enhancement and hemodynamics. However, in this study, we identified some invasive breast cancers with smooth margins and other manifestations similar to benign masses. Nonetheless, BRCA1 breast cancers have obvious edge enhancement, which enables excluding fibroadenoma diagnosis. This marginal enhancement results from peripheral vascular proliferation and central necrosis. In summary, it is difficult to make qualitative diagnosis by only relying on the experience of the imaging physicians and basic influencing parameters. This warrants the development of a noninvasive, qualitative imaging diagnostic model. Radiology research at present is more inclined to generate a ROI first and then extract texture information from it, which considers the tumor as a whole. Finally, the researchers explore the relationship between texture and patients' clinical information by machine learning and other methods. Some previous experiments were based on the theory of habitat imaging, but only on animal models [18]. Some study based on the theory of habitat imaging, but after researchers segmented images, they chose to extract the texture information of each subregion, and then established a neural network model by those texture information. This method emphasizes more the differences between subregions than the connections. Undoubtedly, characterizing the entire tumor as an indivisible whole is not the most sensitive method for assessing the intratumor heterogeneity. Habitat analysis lends its name to the definition of an ecological habitat surrounding a population of species. This emerging approach explicitly divides tumors containing similar characteristics into subregions. Such differences are not only from macroscopic features, such as necrosis caused by uninhibited growth, but also from significant genetic heterogeneity between and within tumors [27]. This is usually attributed to random mutations that drive cloning and evolution [28]. Nonetheless, tumor cells can promote independent growth owing to local angiogenesis, which will lead to chaos in the distribution of blood vessels in the tumor and cause tumor blood flow within random and cyclical changes. Eventually, it will cause the uneven distribution of oxygen, glucose, hydrogen ions, and serum growth factors. Differentially acting tumor cells lead to the specific evolution of tumor population to adapt to this environment [29, 30]. Based on this theory, the imaging results of tumors can predict the associated genetic factors as well as the characteristics of tumor populations generated under specific circumstances [31]. This in turn allows habitat imaging to strongly explain the different outcomes in patients with similar molecular subtype, following an identical therapeutic strategy [17]. Moreover, habitat imaging plays an important role in explaining the causes of drug resistance during chemotherapy in a particular patient. For example, changes in the composition of tumor subareas often attribute to the failure of patients in achieving pathological complete response during neoadjuvant chemotherapy. Meanwhile, the development of drug resistance in a specific subregion leads to poor therapeutic response in therapy. This phenomenon has expanded the use of habitat imaging from screening

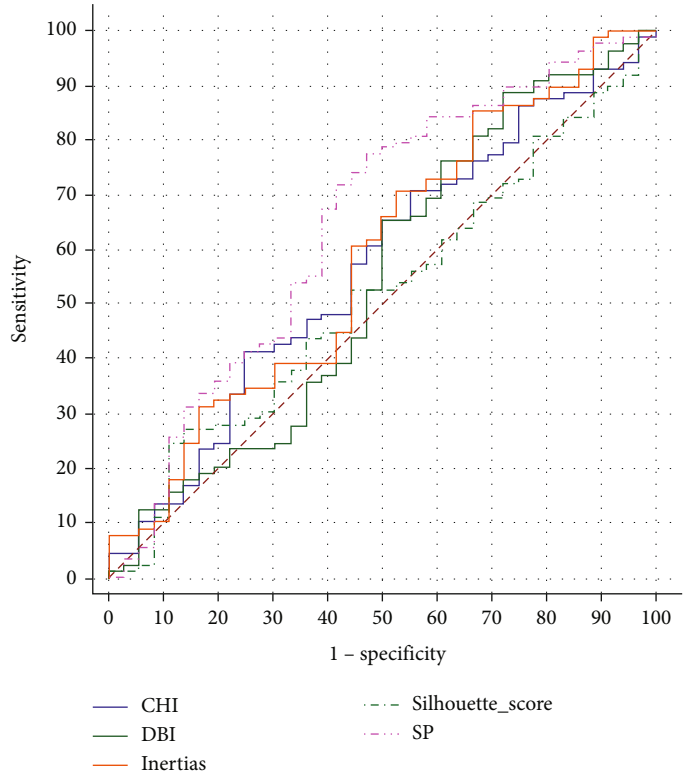


Figure 2: Receiver operating characteristic for clustering features.

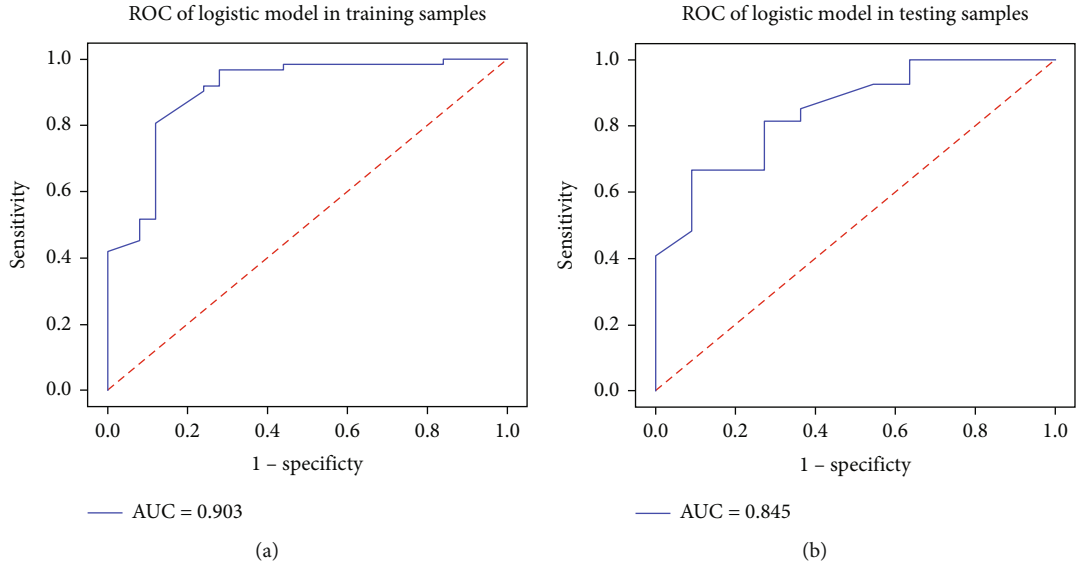


Figure 3: Receiver operating characteristic of the prediction model: (a) the performance of training set; (b) the performance of the validation set.

people at a high risk of cancer to assessing the treatment and patient outcomes throughout the disease. Currently, habitat imaging has made good progress in neural tumors [32].

Different from traditional imaging omics studies, we classified the tumors into subregions and analyzed their relationship to explain spatial heterogeneity. We consider both the differences between each cluster and the cluster shape. In this study, we extracted five parameters of a term used to describe the heterogeneity of cancer. These parameters were used for the simultaneous differentiation between each tumor subregion, thus considering the similarity between each cluster sample. In other words, our parameters can

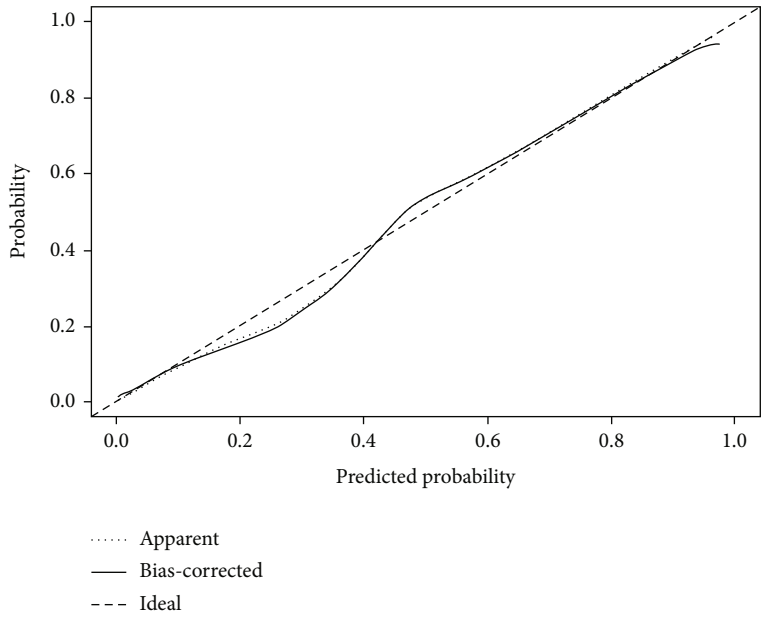


Figure 4: Calibration curve of the prediction model.

describe each cluster more effectively. Moreover, it described the degree of heterogeneity within the tumor from various angles in detail, thus providing a description of the tumor. We focused not only on the absolute characteristics of the tumor but also on the relative relationship between the subregions in case of individual differences. Some habitat imaging studies use images of tumors as samples separately or simply select the largest cross-section of the image as a sample. The aforementioned method can achieve certain advantages in sample representativeness. Nonetheless, tumors do not have a fixed shape and growth direction like some tissues or organs. In other words, tumors do not have a homogeneous and stable physiological and anatomical structure. Partial cardiac tumor necrosis and uneven texture distribution (i.e., tumor heterogeneity) are the major pathological changes. Moreover, necrosis can predict the effect of tumor therapy [33]. Thus, cross-sectional images do not represent the structural characteristics of the entire tumor. The selection of the largest cross-section will result in imbalance of the segmentation. However, it will introduce an error for the experimental method that principally considers the relative relationship of the subregion. In fact, all five of the parameters we showed can be affected by this. Therefore, the 3-D image of the whole tumor needs to be used as the segmentation object. In summary, the proportion of each subregion of the entire tumor, cluster shape, and cluster center affected the modeling, thereby indicating the need for considering the tumor as a whole.

Despite considering several aspects, there were several shortcomings in this study. First, the use of a specific imaging protocol for enhanced MRI, not common in clinical practice, was a major limitation. Despite attempting to standardize the imaging and selecting fewer additional multicenter data and the validation set, technical factors, such as the field strength, repetition time, echo time, and flip angle, may have affected the results. This warrants greater and larger multicenter studies to confirm and validate our findings. Another limitation was in terms of image registration. One benefit of the imaging protocol was that the image sequence could be easily registered, for each sequence of patient images was generated at the same time. There is still a lot of work to be done to roll out the technology. Manual recalibration following the rough registration of the algorithm is a feasible method to ensure the consistency of the voxel represented by similar coordinates. In addition, we excluded a significant number of patients, which resulted in the inclusion of fewer patients, particularly owing to the absence of prospective cases. In the image processing, we only adopted a filter to remove the image noise, which was justified from the results. By contrast, the oversegmented technology [34] has been used by researchers to initially divide the tumor into smaller blocks and subsequent clusters, which may be more helpful in describing the real growth of the tumor. However, despite not adopting the oversegmented technology, we considered the distribution of local tissue abnormalities during image processing. Thus, we used the mean value of peripheral voxels within the range of $3 * 3 * 3$ to replace each voxel, which may help us offset the impact of noise. A reasonable theory is that clusters formed by our approach were more representative, provided greater space continuity. The oversegmented technology can be more detailed and generate an accurate depiction of the tumor tissue heterogeneity distribution. However, the application of the oversegmented technology may result in two situations. Fewer number of clusters will deteriorate the continuity of the segmented subregions. By contrast, increasing the number of clusters will lead to the poor representation of some subregions, thus posing challenges to the complexity of subsequent modeling. Because of the above-mentioned technical difficulties, obtaining satisfactory training results necessitates the use of more powerful modeling methods, such as neural network. By contrast, the model should be trained

with a larger amount of data to avoid overfitting problems. The selection of the neural network and the inclusion of excessive data should be avoided in this experiment. In addition, the weights of all sequences in the clustering were equal and fixed during the final use of image clustering. We did not consider the redundancy of clinical information displayed by different sequences. The 3-D image segmentation method resulted in limited amount of data. Therefore, we adopted the efficient clustering method, which eventually led to overfitting. Hence, we used the image data to the maximum possible to provide various parameters for clustering. This may have led to a high proportion of information describing some aspect of the clustering effect, i.e., the relationship between subregions, thereby affecting the accuracy of clustering. In the final modeling stage, the adoption of greater parameters could not improve the final effect of the model. It also shows that the redundant information interfered with the model.

5. Conclusions

In summary, a model is established based on clustering parameters extracted from subregions formed by habitat imaging. The model can predict BRCA1 mutations detected by immunohistochemistry satisfactorily. And it provided a reference for screening the high-risk population, surgery, drug therapy, and prognosis.

Data Availability

The MRI and clinic Information data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, pp. 209–249, 2021.

[2] D. J. Toft and V. L. Cryns, "Minireview: basal-like breast cancer: from molecular profiles to targeted therapies," *Molecular Endocrinology*, vol. 25, no. 2, pp. 199–211, 2011.

[3] J. D. Fackenthal and O. I. Olopade, "Breast cancer risk associated with *BRCA1* and _BRCA2_ in diverse populations," *Nature Reviews Cancer*, vol. 7, no. 12, pp. 937–948, 2007.

[4] S. R. Lakhani, M. J. Van De Vijver, J. Jacquemier et al., "The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations inBRCA1andBRCA2," *Journal of Clinical Oncology*, vol. 20, no. 9, pp. 2310–2318, 2002.

[5] G. T. Lang, J. X. Shi, X. Hu et al., "The spectrum of BRCA mutations and characteristics of BRCA-associated breast cancers in China: screening of 2,991 patients and 1,043 controls by next-generation sequencing," *International journal of cancer*, vol. 141, no. 1, pp. 129–142, 2017.

[6] B. G. Haffty, Q. Yang, M. Reiss et al., "Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer," *Journal of Clinical Oncology*, vol. 24, no. 36, pp. 5652–5657, 2006.

[7] L. Steward, L. Conant, F. Gao, and J. A. Margenthaler, "Predictive factors and patterns of recurrence in patients with triple negative breast cancer," *Annals of Surgical Oncology*, vol. 21, no. 7, pp. 2165–2171, 2014.

[8] B. D. Lehmann, J. A. Bauer, X. Chen et al., "Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies," *The Journal of Clinical Investigation*, vol. 121, no. 7, pp. 2750–2767, 2011.

[9] W. D. Foulkes, J. S. Brunet, I. M. Stefansson et al., "The prognostic implication of the basal-like (cyclin E^{high}/$p27^{low}$/$p53^{+}$/glomeruloid-microvascular-proliferation^{+}) phenotype ofBRCA1-related breast cancer," *Cancer Research*, vol. 64, no. 3, pp. 830–835, 2004.

[10] A. R. Venkitaraman, "Cancer susceptibility and the functions of BRCA1 and BRCA2," *Cell*, vol. 108, no. 2, pp. 171–182, 2002.

[11] S. Winters, C. Martin, D. Murphy, and N. K. Shokar, "Breast cancer epidemiology, prevention, and screening," *Progress in Molecular Biology and Translational Science*, vol. 151, pp. 1–32, 2017.

[12] A. Antoniou, P. D. P. Pharoah, S. Narod et al., "Average risks of breast and ovarian cancer associated with _BRCA1_ or _BRCA2_ mutations detected in case series unselected for family history: a combined analysis of 22 studies," *The American Journal of Human Genetics*, vol. 72, no. 5, pp. 1117–1130, 2003.

[13] L. Tobalina, J. Armenia, E. Irving, M. J. O'Connor, and J. V. Forment, "A meta-analysis of reversion mutations in _BRCA_ genes identifies signatures of DNA end-joining repair mechanisms driving therapy resistance," *Annals of Oncology*, vol. 32, no. 1, pp. 103–112, 2021.

[14] G.-H. Wang, C.-M. Zhao, Y. Huang, W. Wang, S. Zhang, and X. Wang, "BRCA1 and BRCA2 expression patterns and prognostic significance in digestive system cancers," *Human Pathology*, vol. 71, pp. 135–144, 2018.

[15] F. H. Vaz, P. M. Machado, R. D. Brandao et al., "Familial breast/ovarian cancer andBRCA1/2genetic screening: the role of immunohistochemistry as an additional method in the selection of patients," *Journal of Histochemistry & Cytochemistry*, vol. 55, no. 11, pp. 1105–1113, 2007.

[16] C. C. Riedl, N. Luft, C. Bernhart et al., "Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density," *Journal of Clinical Oncology*, vol. 33, no. 10, pp. 1128–1135, 2015.

[17] M. B. Daly, T. Pal, M. P. Berry et al., "Genetic/familial high-risk assessment: breast, ovarian, and pancreatic, version 2.2021, NCCN clinical practice guidelines in oncology," *Journal of the National Comprehensive Cancer Network*, vol. 19, no. 1, pp. 77–102, 2021.

[18] B. V. Jardim-Perassi, S. Huang, W. Dominguez-Viqueira et al., "Multiparametric MRI and coregistered histology identify tumor habitats in breast cancer mouse models," *Cancer Research*, vol. 79, no. 15, pp. 3952–3964, 2019.

[19] R. Lo Gullo, I. Daimiel, C. Rossi Saccarelli et al., "Improved characterization of sub-centimeter enhancing breast masses

on MRI with radiomics and machine learning in BRCA mutation carriers," *European radiology*, vol. 30, no. 12, pp. 6721–6731, 2020.

[20] R. J. Gillies and Y. Balagurunathan, "Perfusion MR imaging of breast cancer: insights using "habitat imaging"," *Radiology*, vol. 288, no. 1, pp. 36-37, 2018.

[21] J. P. Monaco and A. Madabhushi, "Class-specific weighting for Markov random field estimation: application to medical image segmentation," *Medical Image Analysis*, vol. 16, no. 8, pp. 1477–1489, 2012.

[22] M. Dias, B. Rocha, J. F. Teixeira, and H. P. Oliveira, "Automatic sternum segmentation in thoracic MRI," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1018–1021, Berlin, Germany, 2019.

[23] E.-K. Kim, S. Y. Park, and S.-W. Kim, "Clinicopathological characteristics of BRCA-associated breast cancer in Asian patients," *Journal of Pathology and Translational Medicine*, vol. 54, no. 4, pp. 265–275, 2020.

[24] R. D. Kennedy, J. E. Quinn, P. G. Johnston, and D. Paul Harkin, "*BRCA1* : mechanisms of inactivation and implications for management of patients," *The Lancet*, vol. 360, no. 9338, pp. 1007–1014, 2002.

[25] P. O. Chappuis, V. Nethercot, and W. D. Foulkes, "*Clinicopathological characteristics of BRCA1- and BRCA2-related breast cancer*," in *In Seminars in surgical oncology, vol. 18, no. 4, pp. 287-295*, John Wiley & Sons, Inc., New York, 2000.

[26] S. M. Ha, E. Y. Chae, J. H. Cha, H. H. Kim, H. J. Shin, and W. J. Choi, "Association of BRCA mutation types, imaging features, and pathologic findings in patients with breast cancer with BRCA1 and BRCA2 mutations," *American Journal of Roentgenology*, vol. 209, no. 4, pp. 920–928, 2017.

[27] J. M. Price, S. P. Robinson, and D. M. Koh, "Imaging hypoxia in tumours with advanced MRI," *The Quarterly Journal of Nuclear Medicine and Molecular Imaging: Official Publication of the Italian Association of Nuclear Medicine (AIMN)[and] the International Association of Radiopharmacology (IAR),[and] Section of the Society of... 2013*, vol. 57, no. 3, pp. 257–270, 2013.

[28] R. A. Gatenby and R. J. Gillies, "A microenvironmental model of carcinogenesis," *Nature Reviews Cancer*, vol. 8, no. 1, pp. 56–61, 2008.

[29] H. Kimura, R. D. Braun, E. T. Ong et al., "Fluctuations in red cell flux in tumor microvessels can lead to transient hypoxia and reoxygenation in tumor parenchyma," *Cancer Research*, vol. 56, no. 23, pp. 5522–5528, 1996.

[30] P. Carmeliet and R. K. Jain, "Molecular mechanisms and clinical applications of angiogenesis," *Nature*, vol. 473, no. 7347, pp. 298–307, 2011.

[31] S. J. Hectors, M. Wagner, O. Bane et al., "Quantification of hepatocellular carcinoma heterogeneity with multiparametric magnetic resonance imaging," *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017.

[32] M. Zhou, L. Hall, D. Goldgof et al., "Radiologically defined ecological dynamics and clinical outcomes in glioblastoma multiforme: preliminary results," *Translational Oncology*, vol. 7, no. 1, pp. 5–13, 2014.

[33] A. H. Abdelhafez, B. C. Musall, B. E. Adrada et al., "Tumor necrosis by pretreatment breast MRI: association with neoadjuvant systemic therapy (NAST) response in triple-negative breast cancer (TNBC)," *Breast Cancer Research and Treatment*, vol. 185, no. 1, pp. 1–12, 2021.

[34] M. T. Bennai, Z. Guessoum, S. Mazouzi, S. Cormier, and M. Mezghiche, "A stochastic multi-agent approach for medical-image segmentation: application to tumor segmentation in brain MR images," *Artificial Intelligence in Medicine*, vol. 110, article 101980, 2020.