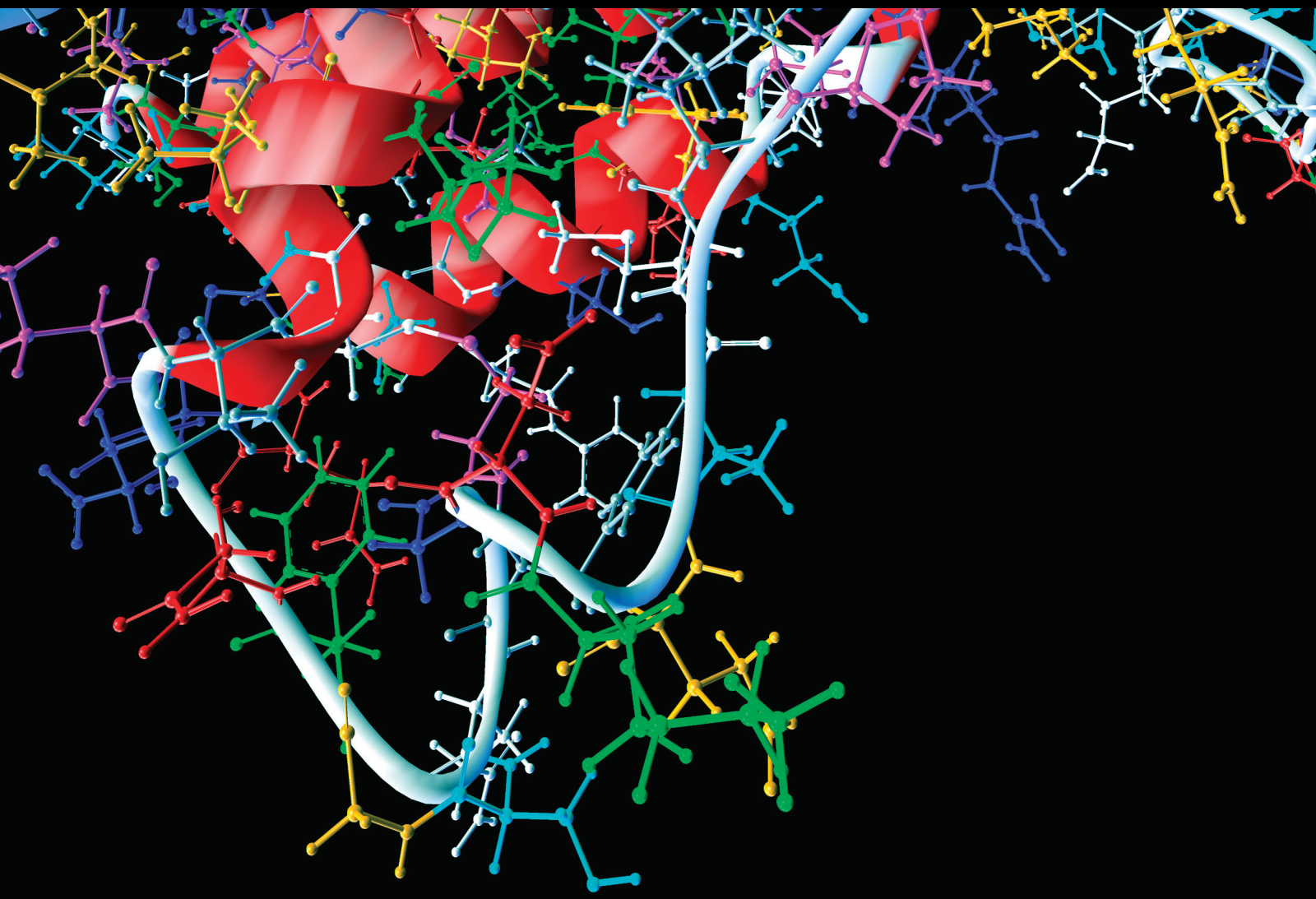
Computational and Mathematical Methods in Medicine

Machine Learning and Artificial Intelligence Methods in Computer Vision and Visualization for Healthcare

Lead Guest Editor: Pan Zheng
Guest Editors: Tao Song and Iman Liao



Machine Learning and Artificial Intelligence Methods in Computer Vision and Visualization for Healthcare

Computational and Mathematical Methods in Medicine

Machine Learning and Artificial Intelligence Methods in Computer Vision and Visualization for Healthcare

Lead Guest Editor: Pan Zheng
Guest Editors: Tao Song and Iman Liao

Contents

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 3812509, 7 pages
https://doi.org/10.1155/2022/3812509

Research Article

Accurate 3D Reconstruction of White Matter Hyperintensities Based on Attention-Unet

Xun Wang,[1] **Lisheng Wang**,[1] **Jianjun Yang**,[2] **and Xiaoya Feng**[3]

[1]*College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, 266580 Shandong, China*
[2]*Department of General Practice, Shandong Provincial Third Hospital, Shandong University, Jinan, 250031 Shandong, China*
[3]*Department of Neurology, Shandong Provincial Third Hospital, Shandong University, Jinan, 250031 Shandong, China*

Correspondence should be addressed to Jianjun Yang; qlbsh1@163.com and Xiaoya Feng; 582610913@qq.com

Received 17 September 2021; Revised 22 February 2022; Accepted 7 March 2022; Published 23 March 2022

Academic Editor: Huiling Chen

White matter hyperintensities (WMH), also known as white matter osteoporosis, have been clinically proven to be associated with cognitive decline, the risk of cerebral infarction, and dementia. The existing computer automatic measurement technology for the segmentation of patients' WMH does not have a good visualization and quantitative analysis. In this work, the author proposed a new WMH quantitative analysis and 3D reconstruction method for 3D reconstruction of high signal in white matter. At first, the author using ResUnet achieves the high signal segmentation of white matter and adds the attention mechanism into ResUnet to achieve more accurate segmentation. Afterwards, this paper used surface rendering to reconstruct the accurate segmentation results in 3D. Data experiments are conducted on the dataset collected from Shandong Province Third Hospital. After training, the Attention-Unet proposed in this paper is superior to other segmentation models in the segmentation of high signal in white matter and Dice coefficient and MPA reached 92.52% and 92.43%, respectively, thus achieving accurate 3D reconstruction and providing a new idea for quantitative analysis and 3D reconstruction of WMH.

1. Introduction

The decline of cognitive function, the risk of cerebral infarction, and dementia seriously affect the health of modern people. At present, there are many methods to predict cognitive decline and cerebral infarction dementia. Qiao and others believe that MMSE prediction plays an important role in the early detection of Alzheimer's disease. They use the convolutional neural network to predict MMSE more effectively [1]. Solovyev and others believe that the decline of cognitive ability is related to Alzheimer's disease and capillary stagnation and the convolutional neural network should be used to detect capillaries, which achieves good results [2]. In recent years, the quantitative analysis of white matter hyperintensity (WMH) has attracted extensive attention. White matter hyperintensities (WMH), also known as white matter osteoporosis, are characterized by high signal in T2-weighted magnetic resonance imaging (MRI) FLAIR or T1W sequence images and fluid attenuation inversion recovery sequences. The pathological changes are mainly non-characteristic white matter injury, manifested as membrane discontinuity, glial cell proliferation, white matter fiber looseness, pale or swelling of myelin sheath, diffuse axonal injury, and vacuole formation [3, 4]. It was confirmed by the clinic that WMH is associated with cognitive decline, risk of cerebral infarction, and dementia and with gait disorders, balance disorders, and urinary incontinence. Except that, studies have shown that more than 90% of people over the age of 60 have white matter hyperintensities.

The existing WMH quantitative analysis methods at home and abroad are semiautomatic measurement, and doctors need to supervise the whole process. In addition, the existing methods cannot directly display the reconstruction effect and cannot well assist doctors in quantitative analysis. Hernández et al. [5] proposed and evaluated the indicators of white matter damage and realized the quantitative assessment of WMH. Muhammad et al. [6] proposed a convolutional neural network to segment conventional WMH and achieved good results. Dadar et al. [7] used the machine learning method to achieve the segmentation of WMH and

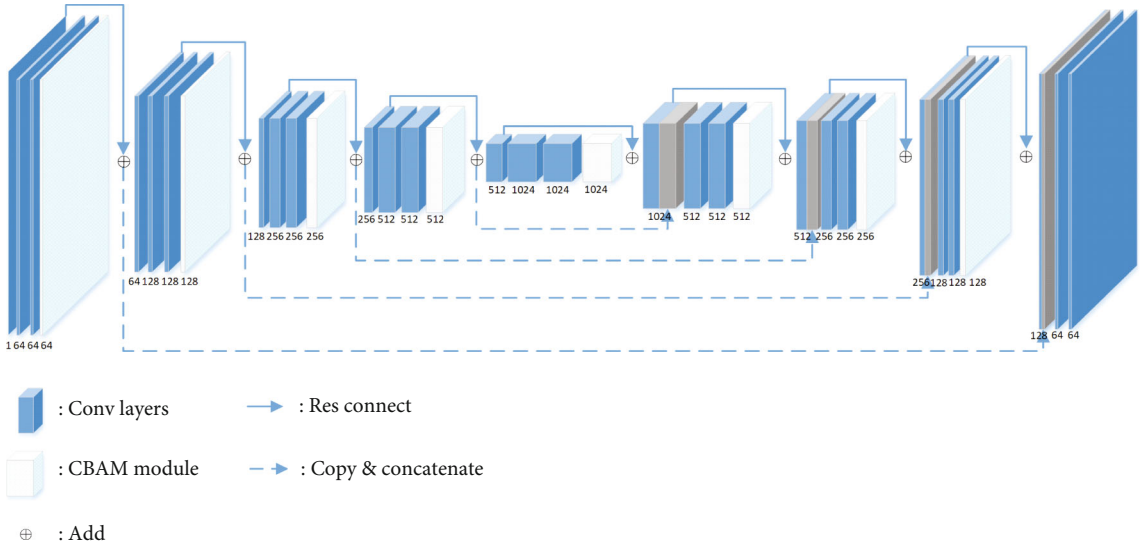


Figure 1: The model of Attention-Unet.

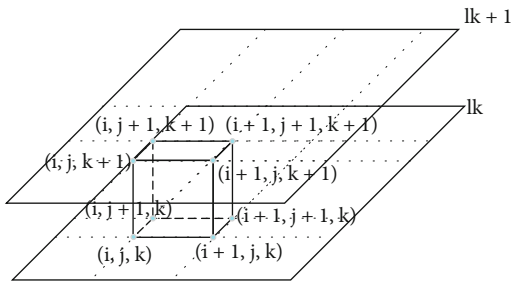


Figure 2: Schematic diagram of moving the cube.

reached the Dice coefficient of 0.84. However, the existing computer automatic measurement technology for the segmentation of patients' WMH does not have a good visualization and quantitative analysis and the reconstruction has not yet retrieved relevant reports.

Medical imaging equipment, such as CT and MRI, can obtain two-dimensional digital tomography images of internal organs of the human body but cannot display the three-dimensional structure of objects. The 3D reconstruction technology in computer image processing and graphics can reconstruct the 2D medical image sequence in 3D and can simulate and display the 3D structure of organs. The human body is scanned by medical scanning imaging equipment, and a continuous two-dimensional digital picture of the fault is obtained, which is then fed into a computer and read. Feature enhancement and segmentation were performed on the fault sequence images. After that, the segmented region is used to reconstruct the 3D image using the reconstruction algorithm. How to conduct accurate segmentation and 3D reconstruction of WMH is a problem that needs to be solved in this paper.

WMH accurate segmentation is the first step in 3D reconstruction. At present, most of the segmentation methods targeting the lesions are still in their infancy. Clustering method is one of the most commonly used traditional segmentation methods. It mainly realizes pixel clustering through the feature similarity of the pixels in the target region. In addition, traditional methods such as the Bayesian classifier and support vector machine (SVM) are also commonly used in lesion segmentation. However, the traditional methods of lesion segmentation usually have the characteristics of complex training and poor segmentation accuracy. With the development of deep learning technology, it is often used in medical image segmentation and detection and has achieved ideal results. In 2015, Long et al. [8] used the convolutional layer to extract features and used the deconvolutional layer to decode the feature image and restore it to the size of the input image, thus realizing image per-pixel classification. After the introduction of FCN, a series of convolutional network structures have been designed for image pixel classification, such as SegNet [9], DeepLabv3 [10], and Unet [11]. These networks are structurally composed of two parts, the coding layer and the decoding layer, and the structure is clear and easy to understand. Among them, Unet has been widely used in the segmentation of medical lesions [12–16] and achieved good results. Extraction of lesion features is the basis of accurate segmentation. VGG-16 [17] and GoogleNet [18] are commonly used lesion feature extraction networks, but their redundant network layer learns parameters that are not identity mapping, resulting in network degradation, while ResNet [19] solves the above model degradation problem by designing residual modules and achieve better feature extraction.

The current 3D reconstruction methods are generally divided into two kinds: surface rendering and volume rendering. In this paper, 3D reconstruction of segmented WMH is performed based on surface rendering. In the early stage of medical scanning imaging equipment, the distance between sections is relatively large and the slice-level surface reconstruction is generally adopted. With the improvement of medical scanning and imaging equipment technology, surface reconstruction based on voxel appears. The isosurfaces are extracted by some algorithms, and the isosurfaces are drawn by constructing the geometric elements of the isosurfaces in the voxels. Yu [20] used 3D reconstruction

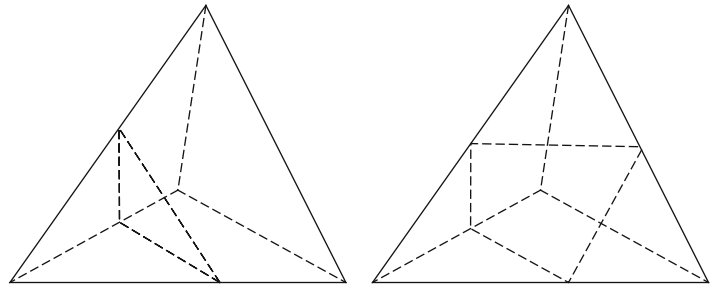


Figure 3: Schematic diagram of the quadrilateral surface.

Table 1: Results of each model experiment (mean ± s.d.%).

Method	Reference module		Evaluation coefficient	
	ResNet	CBAM	Dice (%)	MPA (%)
SegNet	✘	✘	87.43 ± 0.92	86.34 ± 0.58
DeepLabv3	✘	✘	91.31 ± 0.67	90.65 ± 1.08
Unet	✘	✘	88.90 ± 0.43	86.26 ± 1.12
ResUnet	✓	✘	91.05 ± 0.37	90.81 ± 0.49
Attention-Unet	✓	✓	92.52 ± 0.16	92.43 ± 0.82

technology to reconstruct traumatic atlantoaxial vertebra, so as to better help doctors specify appropriate surgical methods. He [21] reconstructed continuous heart tissue sections and realized virtual heart visualization.

In this paper, the improved Unet is used to segment the WMH and the ResNet structure is added on the basis of WMH. In addition, the attention mechanism of CBAM [22] was added in this paper to achieve feature extraction of WMH. In order to solve the ambiguity problem caused by 3D reconstruction with the moving cube method, we used the moving tetrahedron method for 3D reconstruction of a series of segmented images. In this paper, the author conducted data experiments on 23742 pathological images of 100 patients in Shandong Province Third Hospital. The location of the lesion was marked by a professional radiologist. According to the radiologist's labeling, 702 CT images with high white matter signal were selected. The data experiments show that the segmentation network proposed in this paper, compared with other popular segmentation networks, has at least 0.01 Dice coefficient improvement, which also provides the basis for accurate 3D reconstruction of WMH. In addition, the effect of 3D reconstruction in this paper is greatly improved compared with other methods. Through literature analysis, this is the first attempt to carry out accurate 3D reconstruction of WMH and provides a new idea for quantitative analysis and reconstruction of WMH.

2. Materials and Methods

2.1. Materials. The dataset in this paper was obtained from Shandong Province Third Hospital, China, including 100 patients with a total of 23742 pathological images. The locations of lesions were marked by professional radiologists. In this paper, 702 CT images with white matter hyperintensities were selected according to the radiologist's annotation. Due to the high contrast between WMH and surrounding tissues, FLAIR sequences have higher clarity and integrity compared with T1W and T2W sequences. In this paper, most of the DICOM images of high signal in white matter were selected from FLAIR sequence and a few were selected from other sequences.

In the experiment, the white matter hyperintensity data is divided into the train set and test set according to the ratio of 8 : 2. The train set is used for the training of white matter hyperintensity image segmentation, and the test set is used for segmentation test.

The accuracy of white matter hyperintensity reconstruction depends on the accuracy of segmentation. In this paper, the author use histogram equalization to enhance white matter hyperintensity data. The segmentation accuracy of high signal in white matter was greatly improved.

2.2. Methods

2.2.1. Model. The series of the Unet network is a kind of image segmentation network based on the convolutional neural network. Compared with the traditional full convolutional neural network, Unet has been improved to achieve full extraction of features through stronger connection between layers, plus upsampling and downconvolution. To better focus on dividing the area, ResUnet adds jump connection on the basis of Unet to better improve the accuracy of the deep convolutional neural network.

As shown in Figure 1, ResUnet is a U-shaped symmetrical structure, with the convolution layer on the left and the upper sampling layer on the right. Same as Unet, ResUnet contains 4 convolution layers and the corresponding 4 upper sampling layers and the feature map obtained from each convolution layer will be connected to the corresponding upper sampling layer, so that the feature map of each layer can be effectively used in the subsequent calculation. The corresponding upper sampling layer and convolution layer are calculated with a convolution kernel of 3×3 size and activated with ReLu. Among them, 4 convolution layers are connected by a 2×2 maximum pool and 4 upper sampling layers are upsampled by 2×2 convolution kernel. In addition, ResNet adds jump connections in each convolution layer and upper sampling layer, which are shown in Figure 1 as solid blue arrows. ResUnet with jump connection has better segmentation effect, which has been verified in the

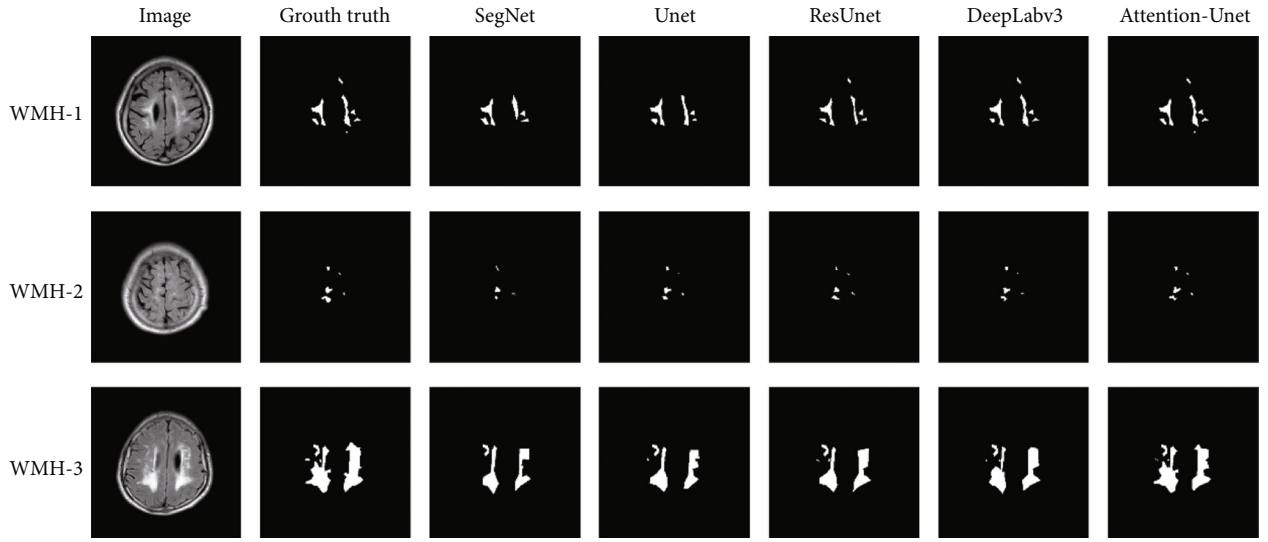


Figure 4: WMH segmentation effects of different models.

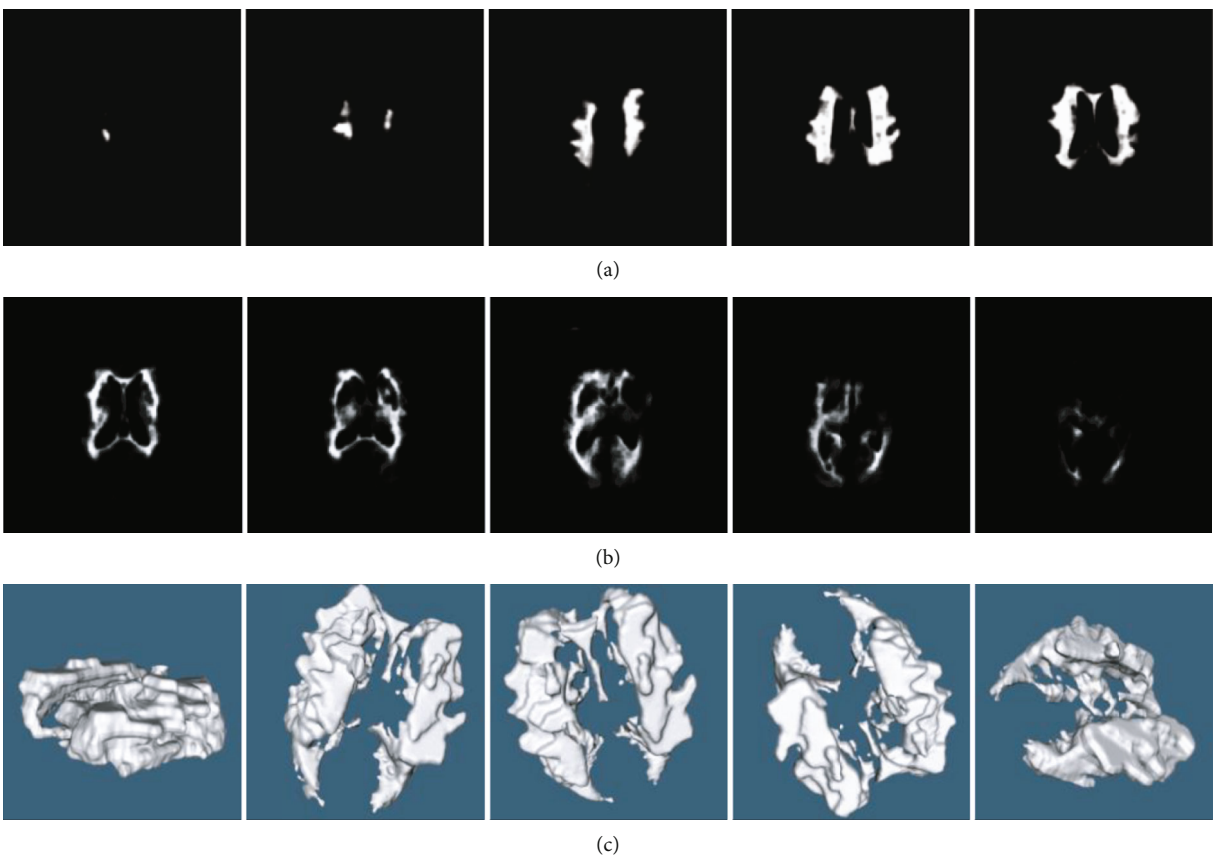


Figure 5: Segmentation and reconstruction renderings. (a, b) Are the results of segmentation of a group of WMH images. (c) Represents the 3D reconstruction result obtained from the segmented image.

experiment. In addition, this paper adds a CBAM attention module after each convolution layer and upper sampling layer, so that the network can obtain better segmentation effect. CBAM is covered in more detail in the next section. The white matter hyperintensity CT images enter the ResUnet network, pass through multiple convolution layers and upper sampling layers, and finally get the accurate segmentation results.

ResUnet uses the Dice loss function to achieve pixel-level segmentation of high signal in white matter. Dice coefficient

is derived from dichotomy and is essentially a measure of the overlap of two samples. The Dice function is shown as follows:

$$\text{Dice} = \frac{2\,|A \cap B|}{|A|+|B|}. \tag{1}$$

Among them, $|A \cap B|$ represents the common elements between set A and set B, $|A|$ represents the number of elements in the set A, and $|B|$ denotes the number of elements in set B. In this paper, the original target and the segmented target are overlapped at pixel level to obtain their Dice value.

2.2.2. CBAM Attention Mechanism. According to the experimental results of Yu [20], sequential addition of channel attention and spatial attention was carried out in this paper. For channel attention, feature graphs generated at the last layer will be maximized and average pored, to generate different spatial context descriptors $F^c_{\max}$ and F^c_{avg}. The descriptor is then entered into the shared network MLP, and the resulting eigenvectors are merged by a summation operation.

The formula for channel attention is as follows:

$$C(F) = \begin{matrix} \sigma(\text{MLP}(\text{AvgPool}(F) + \text{MLP}(\text{MaxPool}(F)) \\ \sigma\left(W_1\left(W\left(F^c_{\text{avg}}\right)\right) + W_1(W_0(F^c_{\max}))\right) \end{matrix}, \tag{2}$$

in which σ represents sigmoid function, W_0 and W_1 are weights in the MLP-shared network, where $W_0 \in \mathbb{R}^{C/r\times C}$ and $W_1 \in \mathbb{R}^{C\times C/r}$, and r represents the reduction rate, aiming at reducing parameter calculation in the shared network.

For spatial attention, the formula is as follows:

$$\begin{aligned} S(F) &= \sigma\left(f^{7\times7}([\text{AvgPool}(F)\,;\text{MaxPool}(F)])\right), \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{\text{avg}}\,;F^s_{\max}\right]\right)\right), \end{aligned} \tag{3}$$

in which σ represents sigmoid function and $f^{7\times7}$ represents convolution operation with a filter size of 7×7.

2.2.3. WMH Reconstruction. After accurate segmentation of white matter hyperintensities, the segmentation images of each group were reconstructed in 3D.

The Marching cubes algorithm is a classic algorithm among surface rendering algorithms. It is a voxel level reconstruction algorithm proposed by Lorense [23] in 1987, also known as the isosurface extraction algorithm. The main idea of the moving cube algorithm is to approximate the isosurface by linear difference in a three-dimensional discrete data field. In medical image segmentation and reconstruction, this isosurface is determined by defining a threshold value. First, define the concept of a "cell," as distinct from a "voxel." A voxel is a grid of eight pixels arranged in sequence, and each voxel (except the boundary) is shared by the eight voxels. There are three kinds of vertex values in a volume element: above or equal to the value is inside the surface, and below the value is outside the surface. Move the cube as shown in Figure 2.

As shown in Figure 2, l_k and l_{k+1} represent two different isosurfaces; (i, j, k) represent the vertices of voxels. There are two possible states for one vertex of a voxel, so a voxel (8 vertices) has a total of 28 or 256 states, where the grayscale value of a point within the voxel (as in Figure 2) can be calculated using the trilinear interpolation equation (4).

$$f(x,y,z) = a_0 + a_1x + a_2y + a_3z + a_4xy + a_5yz + a_6zx{+}{+}a_7xyz. \tag{4}$$

$a_0 - a_7$ represent the gray value of the eight vertices of the voxel; x, y, z represents the coordinate points in the voxel; $f(x, y, z)$ represents the gray value of the points (x, y, z) in the voxel. The 256 combinations can be reduced to 128 combinations by reversing the symmetry (the exchange of vertex 0 and 1 values in the boundary voxel). By rotating the symmetry (the position of vertexes 0 and 1 is the same after the rotation of the boundary voxel), the 256 combinations can be reduced to 15 cases. Each state of the volume element contains a number of three facets, and the vertices of the triangular facets in the volume element need to be calculated by linear interpolation according to the value of the isosurface and the value of the two vertices on the side.

In order to solve the problem of ambiguity in the connection of triangular surfaces in the moving stereoscopic method, this paper uses the moving tetrahedron method [24, 25] to carry out 3D reconstruction. Compared with the moving stereo method, the moving tetrahedron algorithm is to divide the cube element in the moving cube algorithm into tetrahedrons. There are many ways of splitting, usually dividing into 5 tetrahedrons, and then constructing isosurfaces in the tetrahedron. There are 24 and 16 combinations in total. By inverting and rotating symmetry, only one vertex in the boundary voxel is larger than the isosurface, so the triangular surface is generated. If two vertices are larger than the isosurface, a quadrilateral surface is generated, as shown in Figure 3.

Figure 3 shows the schematic diagram of the moving tetrahedron method. The article uses the moving tetrahedron method for 3D reconstruction of the white matter of the brain and is able to achieve better modeling accuracy.

3. Results

3.1. Evaluation Metrics. In this paper, segmentation and 3D reconstruction were performed on the white matter hyperintensity (WMH) dataset. In order to assess the accuracy of 3D reconstruction relative to the gold standard manual label, a number of corresponding measures for various volumes and spaces were used in this article, as no single measure reflects all the required information about the quality of the reconstruction. In this paper, Dice coefficient is mainly used as the evaluation index of spatial correspondence of each voxel between two segments. Dice coefficient can measure the similarity between WMH output from the

segmented network and real samples. Among them, the Dice coefficient value is between 0 and 1 and the larger the value is, the closer the segmentation is to the real value. The Dice coefficient formula is as follows:

$$\mathrm{Dice}(S, Y) = \frac{2 \mid S \cap Y \mid}{|S| + |Y|}. \tag{5}$$

Among them, the prediction result of the segmentation network is $|S|$ and the real result is $|Y|$. The intersection of the two results is represented by $|S \cap Y|$. In addition, in order to evaluate the accuracy of segmented pixels, MPA (average pixel accuracy) is also used to analyze the experimental results of the two types of pixels in the process of WMH segmentation. The MPA value is also between 0 and 1. The higher the value, the higher the pixel accuracy. MPA formula is as follows:

$$\mathrm{MPA} = \frac{1}{n_{cl}} \sum_{i=0}^{n_{cl}} \frac{P_{ii}}{\sum_{j=0}^{n_{cl}} P_{ij}}. \tag{6}$$

The number of categories is expressed in n_{cl}, the number of correctly classified pixels is expressed in p_{ii}, and the pixels with wrong classification are expressed in p_{ij}. In this paper, different models are used for experimental comparison to prove the effectiveness of the module added in this paper for WMH segmentation. All models are trained on the obtained WMH training set and evaluated on the verification set.

3.2. Implementation Details Jinan Science and Technology Bureau. For the sake of equality and comparison, all the real tests are carried out on PyTorch codes. The actual training was conducted on a Ubuntu 16.04 operating system with 6x Intel(R) Core (TM) i7-7700 CPU, and a NVIDIA GeForce RTX 2080 GPU was used for training. The image input size of each network is 512×512 pixels. In addition, this paper sets 30 epoch for the training of each network model and sets the initial learning rate of each model to 0.001. Unless otherwise stated, all models use the same parameters.

3.3. Main Results. In this paper, MPA and Dice coefficients are used as the evaluation indexes of the spatial correspondence of each voxel between the two segmentations. Due to the lack of WMH data, this paper uses TTA (test time augmentation) to enhance the results of the test set and uses the way of fivefold cross-validation. By observing the results obtained by using the method in this paper, the effect of corresponding WMH reconstruction is evaluated. In addition, in order to prove the effectiveness of the improvement of the Unet model in this paper, an experimental comparison was conducted on SegNet, DeepLabv3, and Unet.

The experimental results are shown in Table 1.

SegNet achieves a MPA of 86.34% and a Dice coefficient of 87.43% in the segmentation of white matter hyperintensity intensity. Similarly, DeepLabv3 has a MPA of 90.65% and a Dice coefficient of 91.31%, which is higher than SegNet, while the Attention-Unet used in this paper has a Dice coefficient of 92.52%, which is higher than other flow line segmentation models. In addition, compared with Unet, the residual module and CBAM attention module added in this paper increase the Dice coefficient by 2.15% and 1.47%, respectively. This is enough to prove the accuracy of the segmentation method used in this paper.

In order to better represent the content of the experimental results, the segmentation effects of five different models are compared and the comparison effect is shown in Figure 4.

Among them, WMH segmentation and reconstruction effects obtained by Attention-Unet and the moving tetrahedron method are shown in Figure 5, where WMH is generated from the segmented images of this sequence.

4. Discussion

In this paper, an accurate 3D reconstruction of WMH is attempted for the first time and a new idea for quantitative analysis and reconstruction of WMH is provided. In order to make the results of 3D reconstruction more accurate, this paper proposes an attention-based model called Attention-Unet. The model adds attention mechanism to achieve more accurate white matter hyperintensity segmentation and improve the accuracy of 3D reconstruction. The experimental results show that the Dice coefficient and MPA of the model on WMH dataset are 92.52% and 92.43%, respectively, which is better than the current popular segmentation model, thus laying an important foundation for the realization of accurate 3D reconstruction. In the following work, the author will try to achieve accurate measurement of the volume of 3D reconstruction of white matter hyperintensities, so that it can be applied to clinical work more quickly. In the next work, we will refer to the following novel work [26–32] to try to improve the accuracy of 3D reconstruction of the WMH and the speed of segmentation.

Data Availability

The data supporting the study are available in the database https://github.com/wls860707495/-White-Matter-Hyperintensities

Conflicts of Interest

The author declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (nos. 61972416, 61873280, and 61873281), Natural Science Foundation of Shandong Province (nos. ZR2019MF012 and ZR2021MH227), and Jinan Science and Technology Bureau plan (202019181).

References

[1] H. Qiao, L. Chen, and F. Zhu, "Ranking convolutional neural network for Alzheimer's disease mini-mental state examination prediction at multiple time-points," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106503, 2022.

[2] R. Solovyev, A. A. Kalinin, and T. Gabruseva, "3D convolutional neural networks for stalled brain capillary detection," *Computers in Biology and Medicine*, vol. 141, p. 105089, 2022.

[3] F. A. Sorond and P. B. Gorelick, "Brain white matter: a substrate for resilience and a substance for subcortical small vessel disease," *Brain Sciences*, vol. 9, no. 8, p. 193, 2019.

[4] K. M. E. Forsberg, Y. Zhang, J. Reiners et al., "Endothelial damage, vascular bagging and remodeling of the microvascular bed in human microangiopathy with deep white matter lesions," *Acta Neuropathologica Communications*, vol. 6, no. 1, p. 128, 2018.

[5] M. D. C. V. Hernández, F. M. Chappell, and S. M. Maniega, "Metric to quantify white matter damage on brain magnetic resonance images," *Springer Open Choice*, vol. 59, no. 10, pp. 951–962, 2017.

[6] F. R. A. B. Muhammad, D. C. Maria, and B. Valdés-Hernández, "Segmentation of white matter hyperintensities using convolutional neural networks with global spatial information in routine clinical brain MRI with none or mild vascular pathology," *Computerized Medical Imaging and Graphics*, vol. 66, pp. 28–43, 2018.

[7] M. Dadar, T. A. Pascoal, S. Manitsirikul, K. Misquitta, and D. L. Collins, "Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease," *IEEE Transactions on Medical Imaging.*, vol. 36, no. 8, pp. 1758–1768, 2017.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 39, no. 4, pp. 640–651, 2015.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 12, 2017.

[10] L. C. Chen, G. Papandreou, and F. Schroff, *Rethinking Atrous Convolution for Semantic Image Segmentation*, 2018.

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

[12] X. Li, C. Hao, and X. Qi, "H-DenseUnet: hybrid densely connected Unet for liver and liver tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging.*, vol. 37, 2018.

[13] H. Shaziya and K. Shyamala, "Pulmonary CT images segmentation using CNN and UNet models of deep learning," in *2020 IEEE Pune section international conference (PuneCon)*, Pune, India, 2020.

[14] S. Moradi, M. Ghelich-Oghli, and A. Alizadehasl, "A novel deep learning based approach for left ventricle segmentation in echocardiography," *Physica Medica*, vol. 67, pp. 58–69, 2019.

[15] Y. Zhao, X. Liu, and W. Sun, "A chain membrane model with application in cluster analysis," *International Journal of Adaptive and Innovative Systems.*, vol. 2, no. 4, pp. 324–348, 2019.

[16] K. Jain and A. Kumar, "An optimal RSSI-based cluster-head selection for sensor networks," *International Journal of Adaptive and Innovative Systems.*, vol. 2, no. 4, pp. 349–361, 2019.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, ArXiv preprint arXiv: 1409.1556.

[18] C. Szegedy, W. Liu, and Y. Jia, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015*, pp. 1–9, Boston, MA, 2015.

[19] K. He, X. Zhang, and S. Ren, "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.

[20] H. L. Yu, "Application of three-dimensional CT reconstruction in traumatic atlantoaxial destabilization," *Journal of Clinical Rehabilitative Tissue Engineering Research*, vol. 15, no. 30, 2011.

[21] H. T. He, *Research on Three-Dimensional Reconstruction of Heart Using Serial Sections Based on Computer 3D Printing*, CRC Press, 2015.

[22] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision*, pp. 3–19, 2018.

[23] W. E. Lorense, "Marching cubes: a high resolution 3D surface construction algorithm," *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[24] B. K. Natarajan, "On generating topologically consistent isosurfaces from uniform samples," *The Visual Computer.*, vol. 11, no. 1, pp. 52–62, 1994.

[25] A. Cignoni, F. Ganovelli, C. Montani, and R. Scopigno, "Reconstruction of topologically correct and adaptive trilinear isosurfaces," *Computers & Graphics.*, vol. 24, no. 3, pp. 399–418, 2000.

[26] S. Tao, Z. Pan, M. W. Dennis, and W. Xun, "Design of logic gates using spiking neural P systems with homogeneous neurons and astrocytes-like control," *Information Sciences.*, vol. 372, pp. 380–391, 2016.

[27] D. Barbuzzi, A. Massaro, A. Galiano, L. Pellicani, G. Pirlo, and M. Saggese, "Multi-domain intelligent system for document image retrieval," *International Journal of Adaptive and Innovative Systems.*, vol. 2, no. 4, pp. 282–297, 2019.

[28] S. Tao, R. P. Alfonso, Z. Pan, and X. X. Zeng, "Spiking neural P systems with colored spikes," *IEEE Transactions on Cognitive and Developmental Systems.*, vol. 10, 2018.

[29] S. Tao, W. Xun, Z. Pan, and X. Li, "A programming triangular DNA origami for doxorubicin loading and delivering to target ovarian cancer cells," *Oncotaget*, vol. 5, 2018.

[30] S. Tao, X. X. Zeng, Z. Pan, J. Min, and R. P. Alfonso, "A parallel workflow pattern modelling using spiking neural P systems with colored spikes," *IEEE Transactions on Nanobioscience*, vol. 17, 2018.

[31] S. Tao, S. C. Pang, S. H. Hao, R. P. Alfonso, and Z. Pan, "A parallel image skeletonizing method using spiking neural P systems with weights," *Neural Processing Letters.*, vol. 50, 2018.

[32] T. Song, L. Q. Pan, T. F. Wu, Z. Pan, M. L. Wong, and R. P. Alfonso, "Spiking neural P systems with learning functions," *IEEE Transactions on Nanobioscience*, vol. 18, no. 2, pp. 176–190, 2019.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 8501828, 11 pages
https://doi.org/10.1155/2022/8501828

Research Article

Compound W-Net with Fully Accumulative Residual Connections for Liver Segmentation Using CT Images

Mahmoud Abdelazim Khattab,[1] **Iman Yi Liao**,[1] **Ean Hin Ooi**,[2] **and Siang Yew Chong**[1]

[1]*School of Computer Science, University of Nottingham, Semenyih 43500, Malaysia*
[2]*School of Engineering, Monash University, Sunway, 47500, Malaysia*

Correspondence should be addressed to Iman Yi Liao; iman.liao@nottingham.edu.my

Received 19 August 2021; Revised 7 December 2021; Accepted 6 January 2022; Published 9 February 2022

Academic Editor: Cristiana Corsi

Computed tomography (CT) is a common modality for liver diagnosis, treatment, and follow-up process. Providing accurate liver segmentation using CT images is a crucial step towards those tasks. In this paper, we propose a stacked 2-U-Nets model with three different types of skip connections. The proposed connections work to recover the loss of high-level features on the convolutional path of the first U-Net due to the pooling and the loss of low-level features during the upsampling path of the first U-Net. The skip connections concatenate all the features that are generated at the same level from the previous paths to the inputs of the convolutional layers in both paths of the second U-Net in a densely connected manner. We implement two versions of the model with different number of filters at each level of each U-Net by maximising the Dice similarity between the predicted liver region and that of the ground truth. The proposed models were trained with 3Dircadb public dataset that were released for Sliver and 3D liver and tumour segmentation challenges during MICCAI 2007-2008 challenge. The experimental results show that the proposed model outperformed the original U-Net and 2-U-Nets variants, and is comparable to the state-of-the-art mU-Net, DC U-Net, and Cascaded UNET.

1. Introduction

The liver is the largest substantive organ of human body. It maintains important life activities such as detoxification and metabolism. Computed tomography (CT) is a common modality to detect liver and liver lesions. Liver segmentation is important for the formulation of treatment plan and for the evaluation of the follow-up treatment effect. As the manual segmentation is error-prone and time-consuming, automatic liver segmentation methods have been extensively studied [1–3].

Liver segmentation is still a challenging task, due to the variability of organ's shape and size and similar intensity values among neighbouring organs and tissues, such as the heart, the stomach, the kidneys, and the abdominal wall. On the other hand, a liver has a crucial role in metabolic processes; therefore, it is essential to perform a fast and accurate diagnosis in case of any disease. Moreover, with the improvement of different medical imaging techniques, the focus is placed on the application of noninvasive diagnostic methods, before performing a painful, invasive examination (e.g., biopsy). Out of different liver pathologies, liver cancer is the fifth most commonly occurring cancer in 2018 world-wide, according to the World Health Organisation and others [4–7]. Therefore, a continuous effort is required to develop efficient and automatic segmentation methods, which may support the diagnostic process and facilitate the treatment decision-making.

Liver segmentation task has been introduced as a challenge for many conferences, e.g., MICCAI 2007, MICCAI 2008, and ISBI 2017. During these challenges, 3Dircadb1 and LiTS datasets were introduced for training and evaluation of the proposed approaches. The existing automatic liver segmentation methods are divided into two

categories: pixel- or image-based segmentation and learning-based segmentation. Thresholding, region growing, edge detection, and graph cut [8–10] are some of the commonly used image-based segmentation methods, which directly segment images by the grey level, texture, and gradient. Most of them have low robustness, are prone to under or over segmentation, and are sensitive to initial seed selection; therefore, a sophisticated preprocessing is required. Therefore, in recent years, these methods are rarely applied to liver segmentation alone, but are usually used as a postprocessing for other methods.

The learning-based segmentation methods include statistical models, traditional machine learning methods, and deep learning methods. Supervised learning methods with pixel-wise binary classification usually performs better than the image-based segmentation methods, e.g., statistical models [11], active shape models (ASM), active appearance models (AAM) [9], level set-based methods [12], and atlas-based segmentation [13]. However, traditional machine learning methods, e.g., support vector machines (SVM) [14], and Adaboost [15], rely on extracting handcrafted image features, which is not efficient and subject to human bias. In contrast, deep learning methods extract image features automatically based on large training dataset without introducing human bias.

Convolutional neural networks have become the state-of-the-art in many fields especially for medical image analysis. U-Net was introduced in 2015 for the segmentation of neuronal structures in electron microscopic stacks. It works with binary crossentropy as a loss function for pixel wise classification, and the energy function is computed by a pixel-wise soft-max over the final feature map combined with the crossentropy loss function [16]. 2-bridged U-Net were proposed for prostate segmentation [17]. CNNs and deep learning approaches are widely used for liver segmentation and require significant number of training samples and preannotated masks as ground truth.

Segmentation processes are usually affected by the edge of the object. Although the skip connections in the conventional U-Net have effectively handled edge information to a certain extent, there are still room for improvement with the U-Net [18]. Firstly, the U-Net architecture duplicates low-resolution information of features. After pooling (i.e., down-sampling), low-resolution information of features passes on to the convolution layer in the next stage. However, this low-resolution information of features is transferred by the skip connection of the U-Net as well. Duplication of low resolution information may then cause smoothing of the object boundary information in the network, which is more critical in the case of fuzzy object boundaries [19]. Another drawback of the U-Net architecture is that it may not sufficiently estimate high level features for high-resolution edge information of the input object. The U-Net use the skip connection to transfer high-resolution information; however, high resolution edge information does not pass through any convolution layers during transfer by the skip connection. Thus, higher level feature maps learned by the network do not contain enough information of the high-resolution edges of the input object. Consequently, in the conventional U-Net, high-level features are extracted disproportionately from low-resolution information [17, 20].

Many new models based on U-Net have been introduced to overcome the drawbacks of the original U-Net [16]. Stacked U-Net [21], V-Net [22], and bridged 2U-Net [23] are examples of the variants of the U-Net. Zhang and Xu [24] added a separated path to extract the global features and local features separately by reducing the number of convolutional channels of the contraction and expansion paths. It has led to a faster training process and improved the efficiency of the convolution kernel feature extraction [24]. Whilst the adjacent network with less number of parameters sped up the training process, it has a limited accuracy [25]. U-Net has also been integrated with other traditional registration and segmentation techniques such as conditional random field (CRF) to segment the liver tumour with limited number of samples [26, 27]. Christ et al. [28] proposed a cascaded CNN in 2D with a 3D dense CRF as a postprocessing step, to achieve higher segmentation accuracy whilst preserving low computational cost and memory consumption. Albishri et al. [29] cascaded 2 U-Net, one for the liver and one for tumour segmentation, with preprocessing Hounsfield units (HU) and Contrast Limited Adaptive Histogram Equalization (CLAHE). Liu et al. [30] proposed CR-U-Net, where the cascade U-Net is combined with residual mapping, and the second-level of cascade network is deeper than the first-level to extract more detailed image features and adopted morphological algorithms as an intermediate-processing step to improve the segmentation accuracy. Lu et al. [31] combined a 3D CNN with a Graph Cut (GC) algorithm for liver segmentation. Wang et al. [32] transformed the Dicom image format to Hounsfield Unit, then used a window of the specific HU for the liver before training with CNNs. They replaced the convolutional layers at each level with dense connection blocks where each dense block contains 5 U-Nets of 2 levels. Zhou et al. [33] combined U-Nets of varying depths into one ensemble architecture where different U-Nets share the same encoder but have separate decoders to encourage knowledge sharing. However, such architecture still suffers from two drawbacks. Firstly, the decoders are disconnected, and deeper U-Nets do not offer a supervision signal to the decoders of the shallower U-Nets in the ensemble. Secondly, the common design of skip connections used in the U-Net is unnecessarily restrictive, requiring the network to combine the decoder feature maps with only the same-scale feature maps from the encoder. While striking as a natural design, there is no guarantee that the same-scale feature maps are the best match for the feature fusion.

Recently, residual mapping has been used in combination with image segmentation architectures, which is an effective way to prevent overfitting and meanwhile to improve accuracy. Milletari et al. [22] combine residual learning with U-Net to construct V-Net for 3D image segmentation. Bi et al. [34] proposed a cascaded deep residual networks (ResNet) approach to segment the liver and liver lesions. As preprocessing, it converts the images to HU and applies data augmentation strategies including random scaling, crops, and flips and used 3D CRF and multiscale fusion for postprocessing. The network is pretrained firstly

on the ImageNet dataset for parameter fine-tuning and is further fine-tuned with the liver dataset. On the other hand, Xu et al. [35] used HU as preprocessing and postprocessing 2D CRF and 3D CRF. ResU-Net added residual connections to each skip connection of the basic U-Net structure. Liu et al. [36] added attention block and residual block to the decoder path of the U-Net with adaptive dice loss function. It has helped increase the dice coefficient loss on LiTS dataset from 0.8365 to 0.9692 with all the residual and attention blocks. Seo et al. [20] included the residual path and a design of object-dependent upsampling to U-Net structure. The network avoids duplication of low-resolution information, estimates higher level feature maps that better represent high-resolution edge information of larger object inputs, and learns to extract even higher level global features for small object inputs. The testing accuracy on 3Dircadb dataset is $96.01 \pm 1.8\%$, marking a relatively superior performance compared to other state-of-the-art.

The previously reviewed models can be categorized to three categories based on the number of U-Nets and the type of connections between the U-Nets. The first category contains the models based on one U-Net, e.g., mUNet [20] introduced a decovolutional block before the skip connection; [24] proposed an extra path for global feature extraction; residual U-Net [35] added residual connection between each two consecutive layers; densely connected U-Net [32] replaced the conv layer at each level to of the contraction path with dens block; [36] added attention module to each level of the expansion path. The second category contains the models that used 2 cascaded U-Nets, e.g., [28, 29] consist of two separated U-Net one for liver and one for tumor segmentation, whereas [34] adds CRF as a post-processing technique. The third category contains the models that implemented 2 stacked U-Nets, e.g., [21] introduced 2 stacked U-Nets with N cut loss function; CRUNet [30] introduced 2 U-Nets with different depth and a morphological technique as intermediate process between the two; and [23] proposed one skip connection as a bridge between the 2 U-Nets.

In comparison, we propose a 2 stacked U-Net model. The model is not computationally expensive as it contains smaller number of layers compared to the Dense U-Net [32, 33] and less number of residual connections compared to [35]. The proposed model introduces 3 types of skip connections between the two U-Net in addition to the normal skip connection in each U-Net, whilst the bridge U-Net [23] used one bridge connection between the 2 U-Nets and [21, 30] contains only the skip connections of the original U-Net.

2. Materials and Methods

2.1. Data. 3Dircadb1 (3D Image Reconstruction for Comparison of Algorithm Database) is created by Hôpitaux Universitaires France as a public dataset for researchers in medical image segmentation. The dataset is composed of 3D CT-scans for 20 patients with hepatic tumours in 75% of cases. For each patient, there are number of CT scans in addition to manually annotated mask for several structures of interests, e.g., liver, left kidney, right kidney, and hepatic tumours performed by clinical experts. All CT scans and masks are in DICOM format with pixel size (512×512). The total numbers of CT scans are 2823. We adopted two augmentation techniques as in [37] to increase the number of samples. After applying horizontal and vertical flipping in addition to rotation with 15°, the total number of samples increased to 112,920 images. 13% of the images are used for testing while the remaining samples are divided into 75% for training and 25% for validation. From patient's point of view, the images are divided into 14 patients for training, 4 patients' data for validation, and 2 patients' data for testing (patients no. 5 and no. 20).

2.2. The Model. The proposed model extends the main feature of 2D U-Net [16] that concatenates the output of each layer in the contracting path to the inputs of the layer on the same level on the expansion pass to limit the effect of the loss in the high level feature during the convolution and pooling process. The model consists of two stacked U-Net with total 4 paths, B1 and B2 are the contracting and expansion path of the first U-Net while B3 and B4 are the components of the second U-Net as in Figure 1. Each U-Net consists of 4 levels at each path in addition to one level to connect the contracting and expansion paths. The number of filters at the contracting path starts with 64 and is increased by 200% for the next level until it reaches 1024 filter at level 5. As for the expansion path, the number of filters is decreased by 50% as the level goes up, reaching the initial 64 filters at the topmost level. The input image size is of 256×256 pixels and is decreased by 50% after each level on the contracting path due to the maxpooling process to reach the minimum image size with 16×16 at level 5, then start to increase by 50% with each level of upsampling on the expansion path. The output feature maps have equal size at each level on both U-Net paths as shown in Figure 2.

Each path consists of a series of layers to construct a block at each level. The blocks on B1 follow the typical architecture of a convolutional network. It consists of two convolutional layers with filter size 3×3 (unpadded convolutions), each followed by an exponential ReLU (ELU) or rectified linear unit (ReLU) at levels 4 and 5, a 2×2 max pooling operation with stride 2 for downsampling, and ends with dropout layer with 50% rate. After each block, the number of feature channels will be doubled. The maxpooling and dropout layers are excluded from the block at level 4. The blocks in the expansion path (B2) consists of an upsampling of the feature map followed by a 2×2 convolution (upconvolution) that halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path, and two 3×3 convolutions, each followed by an ELU or ReLU and a dropout with 45% rate. The blocks in B3 are similar to B1 except that the former starts with a concatenation of the feature map from the corresponding expansion path in B2 with the feature maps from the previous level in the contracting path in B3, then followed by similar layers as in the blocks on B1 (i.e., Conv, Conv, maxpooling, and dropout). The building blocks on B4 have the same structure as B2 except that it concatenates 4 feature

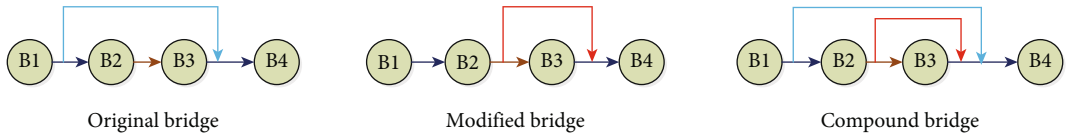


Figure 1: The architecture overview of the proposed 2 stacked U-Net main structure. B1 and B2 are contracting and expansion path for the first U-Net while B3 and B4 are the contracting and expansion path of the second U-Net. The colored lines represent the bridging connections between the different paths. Original, modified, and compound are three different structures of the bridging connections.

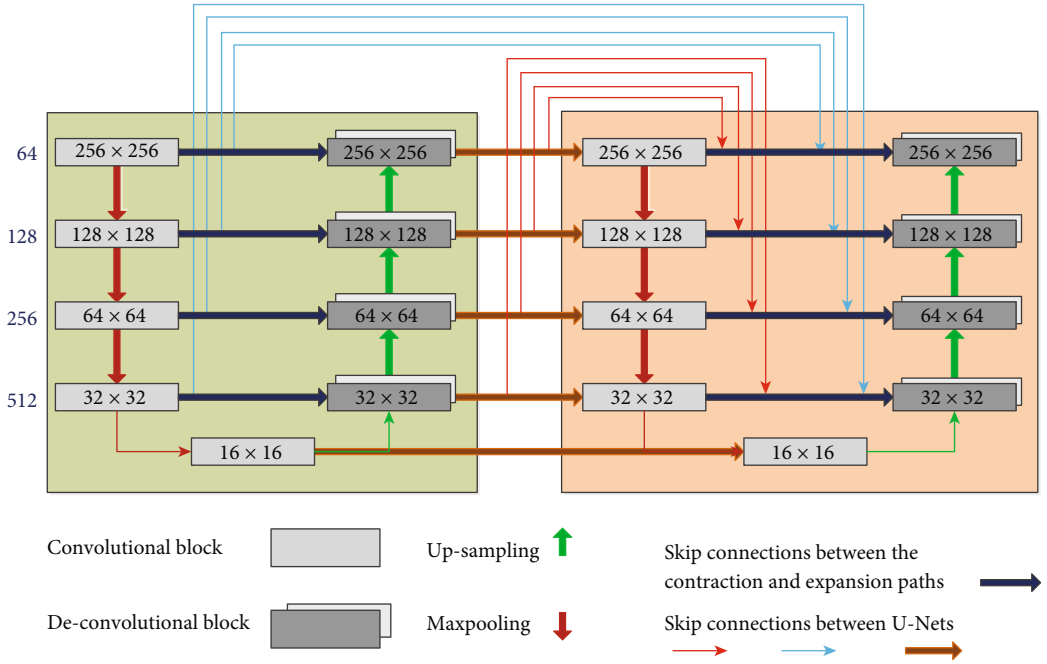


Figure 2: The proposed compound W-Net model structure. The model connects two U-Nets with 3 bridging connections (blue, red, and brown) and one skip connection in each U-Net (dark blue).

maps before applying the sequence of (Conv, Conv, and Dropout). The first feature map that comes from the previous upsampled level in B4 will be concatenated with the feature maps at the corresponding levels from B1, B2, and B3. At the final layer, a 1×1 convolution is used to map each 64-component feature vector to the 256×256 image mask as shown in Figure 3.

2.3. Skip Connections. A model with 2 bridged U-Nets for prostate segmentation [17] introduced a bridge to add the output features from each level at the contracting path of the first U-Net to the inputs of the expansion path of the second U-Net at the same level (blue lines). Our modified model introduces two new bridging connections. One bridge concatenates the output of the expansion path of the first U-Net (B2) to the inputs of the contracting path of the second U-Net (brown lines). The second bridge concatenates the output of B2 to the inputs of B4 (red lines). The final model proposed in this research is a compound model that contains all type of bridging connections. The novel architecture has used all the previously generated feature maps from all paths of the two U-Nets and concatenated them to the inputs of the last expansion path. We hypothesize that by concatenating all previously generated feature maps, the proposed model can decrease the loss of both high-level and low-level features (see Figure 3).

2.4. Feature Concatenation. There are two types of operations to combine the features through the bridge and skip connections. Addition operator applies a pixel-wise summation operation and generates one layer for all the input layers. Concatenation operator stacks all the feature maps together along the feature map dimension with depth equal to the number of input layers. We used concatenation with all skip connections and bridge connections since concatenation operation increases the features space by combining the high-level and low-level features. Therefore, the subsequent convolutional operation is able to learn new features that are dependent on both high-level as well as low-level features (see Figure 4).

In comparison, we created another version of the model started with 32 filters applied on the first layer then increased by 200% on the next level to reach 512 filters applied at the deepest level.

2.5. Objective and Loss Function. In fact, the U-net, an end-to-end segmentation network, is a classification of each pixel. Most of the deep learning networks use the cross-

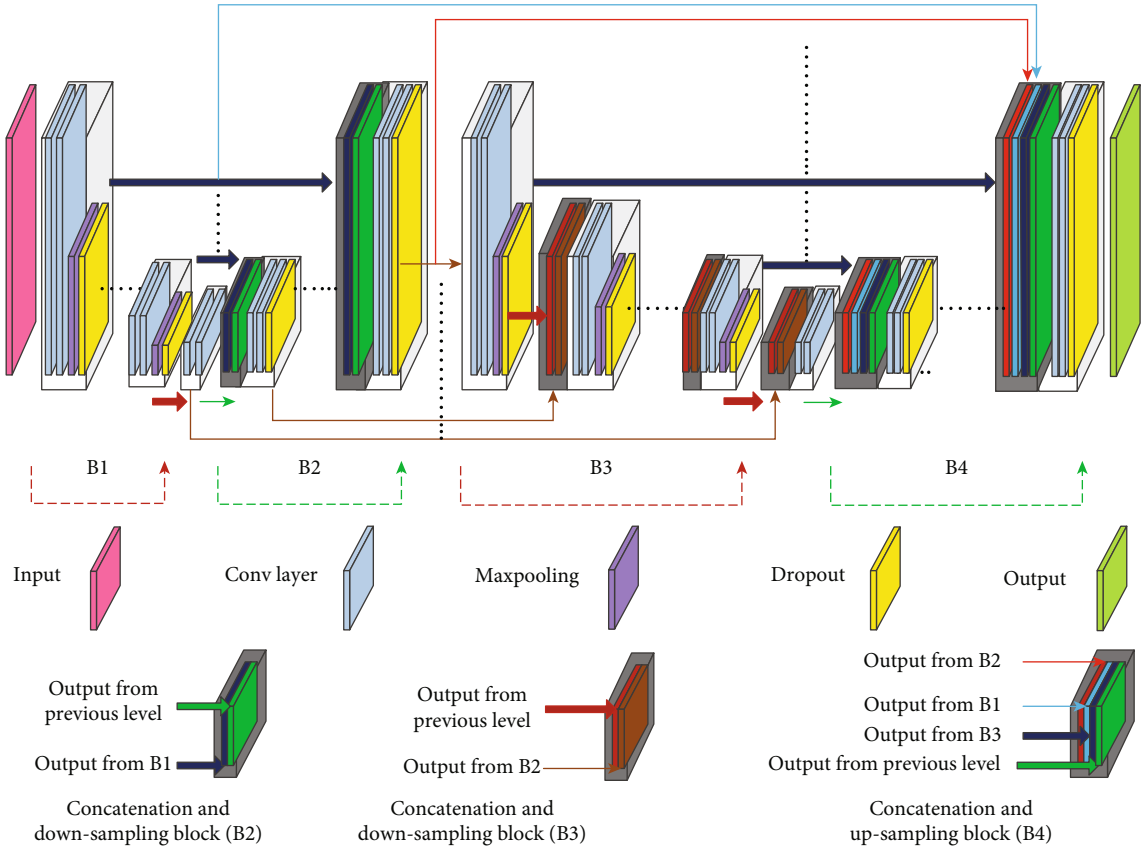


FIGURE 3: Detailed architecture of the model with the layers in each block for all branches B1, B2, B3, and B4.

entropy as a loss function for pixel-wise classification to segment an image into different regions. However, the samples in the dataset we use are only for liver area, and the ratio of positive and negative samples is about 1 : 15, an extremely uneven distribution. If crossentropy function is used in our training process, the result will be biased towards the negative samples.

In view of the above, we implement the Dice loss, a loss function based on Dice similarity coefficient (DSC). The Dice similarity coefficient is a statistical metric that measures the similarity between two sets of data. It has become one of the broadly used metrics in the validation of image algorithms as in

$$\text{Dice} = \frac{2*|X \cap Y|}{|X| + |Y|}, \tag{1}$$

where X and Y are two sets of pixels of ground truth (masks) and the predicted liver, respectively. $|X|$ is the number of elements in set X. $\cap$ represents the set intersection. The loss function L_{Dice} is the difference between the ground truth and the predicted mask as in

$$L_{\text{Dice}} = 1 - \text{Dice}. \tag{2}$$

2.6. Training and Testing. The testing samples represent 13% (14,560 images) of the total samples of 3Dircadb1, whilst the remaining samples are divided into training and validation with percentage 75% (74,680) and 25% (14,560). All models trained for 10 epochs with learning rate $1e^{-5}$ using Adam optimizer [38]. All training and testing images were scaled from 512×512 to 256×256 due to the limitation of computing resources. Training parameters are shown in Table 1.

During the testing phase, each model was evaluated with two groups of data. Firstly, the models were tested using the original testing data without applying any augmentation technique (normal data). Secondly, the models were evaluated using the data after augmentation including all vertically and horizontally flipped images in addition to all 15° step rotated images (augmented data). The results are shown in Tables 2 and 3, respectively.

2.7. Hardware and Software. For training and testing, we used Intel® Core™ i7-6700 CPU @ 3.40 GHz×8, with 16 GB RAM and GPU GeForce GTX 1080/PCIe/SSE2 with 8 GB RAM. Our model was implemented using python 2.7.3, for Keras 2.1.1, with TensorFlow 1.2.1, Theano 0.8 for Ubuntu 14.04.05.

3. Results and Discussion

We refer to the model by the number of filters at the deepest level. For example, (compound-512) which represents the compound model with filters starts with 32 and ends with 512 at the deepest level, whilst (compound1024) which represents the compound model with filters starts with 64 and

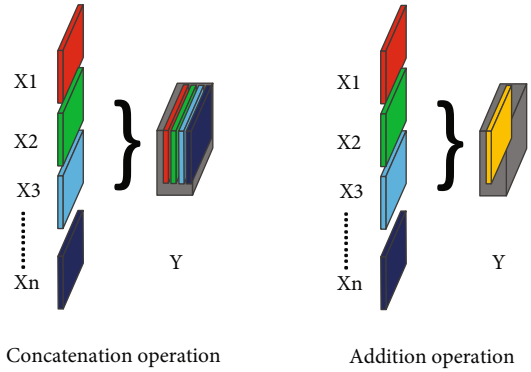


Figure 4: The difference between concatenation and addition operations.

ends with 1024 at the deepest level. To compare our results with the original U-Net [16] and 2-Bridged U-Net [17], we conducted two groups of experiments based on the filter structure of the base U-Net. The first group used base U-Net with (32-➔512) while the second group used base U-Net with (64➔1024).

The evaluation metrics of the medical image segmentation algorithms include but not limited to Intersection over Union (IOU) or Jaccard index, Dice similarity coefficient (DSC or Dice), precision, and recall. We use Dice to evaluate and compare our approach with other related work because it is the common metric for most of liver segmentation methods [10, 19, 20, 24, 28–30, 32–36].

The results in Table 2 highlight the key findings. The model U-Net-1024 recorded higher accuracy than U-Net-512. The models based on the modified connections and compound connections recorded higher accuracy than the original connections with both structures 32-512 and 64-1024. The models compound-512 and compound-1024 recorded the best accuracy over original and modified connections for both filter structures 32-512 and 64-1024, except that the modified model with 64-1024 recorded higher accuracy than that of the models with 64-1024 filters when testing with augmented data. Using the filter structure 64-1024 recorded higher accuracy over the filter structure 32-512 except for compound connections. The best overall accuracy for testing with and without augmented data were recorded for compound model using 32-512 filters structure as shown in Figure 5.

The findings came in line with the main hypotheses that by adding the high level feature-maps from the contraction path of the U-Net to the feature-maps on the same level of the expansion path will reduce the loss of the feature that may be caused due to the convolution and pooling operations. The modified model and the compound models add extra bridging connections from the first U-Net to the final expansion path of the second U-Net and have shown better performance over the original model. Although the original 2-bridged U-Net used filters structure 32-512, the results show that using filter structure of 64-1024 recorded higher accuracy because the number of training parameters is higher than the parameters of the model with filters 32-512 for both original and modified models where only one skip

Table 1: Model parameters for training.

Parameters	Values
Image size	256×256
Number of epochs	10
Learning rate	$1e^{-5}$
Filter size	3×3
Pooling size	2×2
Dropout rate contracting path	0.5
Dropout rate expanding path	0.4
Filters per layer	Previous layer's filters $\times 2$

connection had been used to transfer the feature maps from the first U-Net to the second U-Net. The compound model has demonstrated that using two skip connections to transfer the feature maps to the second U-Net doubles the size of feature space at the second U-Net, enhances the model performance, and decreases the need of increasing the number of filters to 64-1024 (Figure 5). On the other hand, when the size of the feature space at the second U-Net is doubled due to the compound connections, the number of learnable parameters would also be doubled for all the deconvolutional layers in the second U-Net. It has more significant impact on the filter structure of 64-1024 than that of 32-512. That is, the compound model with the filter structure of 64-1024 is more likely to overfit than the compound model with the filter structure of 32-512. This could explain as to why the compound model with the filter structure of 32-512 had better testing performance in terms of the Dice coefficient than the compound model with the filter structure of 64-1024.

The sample results in Figure 6 illustrated that using 64-1024 filters with all models decreased the over segmentation because the number of the high-level and low-level features on the contracting path increased by 200% that will allow the models to learn more global features. In case of the modified and compound, adding the extra skip connections that concatenate all the previously generated features to the final expansion path reduced the false positive and over segmentation artefacts. Although a few images are segmented with better accuracy using the modified model with 64-1024 filters (h column), the compound model in general recorded the best accuracy with both filters' structure.

Column b shows the results of liver segmentation using the original structure of U-Net with 32-512 filters. The results suffer from artefacts near the boundaries of the liver because the loss in the global features during the downsampling propagated to the following convolutional layers of the next deeper levels. Similar loss happened due to the fusion of the features on the expansion path which indicates that the skip connections in the original U-Net is not enough to overcome the features loss. The segmentation using the original structure of the 2 Bridged U-Net enhanced the accuracy for the images where the liver size is relatively smaller in rows 1 and 2 at column c. On the other hand, the output for the images where the liver size is larger have suffered from oversegmentation because the skip connections from

Table 2: The accuracies of training, validation, testing using normal images, and testing using augmented images for all four types of models (U-Net, original, modified, and compound) using both versions with (32➔ 512) and (64➔ 1024) filters structure. The accuracy represents the value of Dice similarity coefficient.

Model	Filters	Training Dice	Validation Dice	Testing	
				Normal data Dice	Augmented data Dice
U-Net [16]	32-512	0.9650	**0.9800**	0.4649	0.6007
Bridge U-Net [17]	32-512	0.9755	0.7065	0.7321	0.6031
Our modified bridge	32-512	*0.9785*	0.6702	0.7912	0.6010
Our compound model	32-512	0.9752	0.9011	**0.8988**	**0.9442**
U-Net [16]	64-1024	0.9740	0.9150	0.5873	0.7593
Bridge U-Net [17]	64-1024	0.9738	0.8775	0.7989	0.7534
Our modified bridge	64-1024	*0.9812*	*0.9250*	0.8137	*0.8368*
Our compound model	64-1024	**0.9812**	0.9113	*0.8303*	0.7836

Table 3: Quantitative comparison between our models and other models using 3Dircadb dataset.

Model	Dice %	Testing	
		Normal data Dice %	Augmented data Dice %
U-Net [16]		46.49	60.07
Bridge U-Net [17]		73.21	60.31
Our modified bridge		79.12	60.10
Our compound model		**89.88**	**94.42**
U-Net [32]	92.6 ± 2.2		
FCN-8 s [32]	92.1 ± 1.5		
3D DSN [32]	92.8 ± 1.4		
DecNet [32]	90.1 ± 1.0		
FCN [32]	93.30		
Cascaded UNet+CRF [32]	93.10		
DC U-Net [32]	**94.9 ± 2.0**		
mU-Net [20]	**96.01 ± 1.8**		
UNET [28]	72.90		
Cascaded UNET [28]	93.10		
Cascaded UNET +3D CRF [28]	**94.30**		

the expansion path at the first U-Net (B2) were concatenated to the contracting path at the second U-Net only, whilst the feature maps from (B1) were transferred directly to (B4) only Figure 1 (original bridge). Column d illustrates the improvement of the output and decreased the oversegmented artefacts because the skip connections of the modified bridge model transferred the low-level features from (B2) directly to be concatenated to the expansion path (B4) at the second U-Net. The compound bridge results in column (e) show the significant enhancement of the performance for all images with small and large liver size because the concatenation of all output feature maps from all the previous paths to the final expansion path of the second U-Net has recovered the loss of the high-level and low-level features. The accumulative residual connections increased the feature space and allowed the model to learn more features by concatenating all the previously generated high-level and low-level feature maps from all the previous paths. Columns (f, g, h, and i) show the results of the models using filter structure 64-1024. Generally, the overall accuracy with 64-1024 models is higher than the models with filter structure 32-515 because of the number of filters, and feature maps generated are increased by 100%. In rare cases, the compound model suffers from undersegmentation at some images with small size liver rows 1 and 2 at column (i) because the number of features transferred from the previous paths to the final path is large and might contain redundant features which increased the false negative for the images with smaller liver size.

The results in Table 3 compared our models with the state-of-the-art approaches that used the same dataset 3Dircadb for liver segmentation. In this paper, we used 14 patients' data for training and 4 patients' data for validation with 80 : 20 ratio and 2 patients' data for testing. In [20], it used 15 patients' data for training and validation and 5 patients for testing; however, it did not specify which patient numbers were used for testing. In [28], 15 patients' data were used for training and testing with 2-fold crossvalidation. In

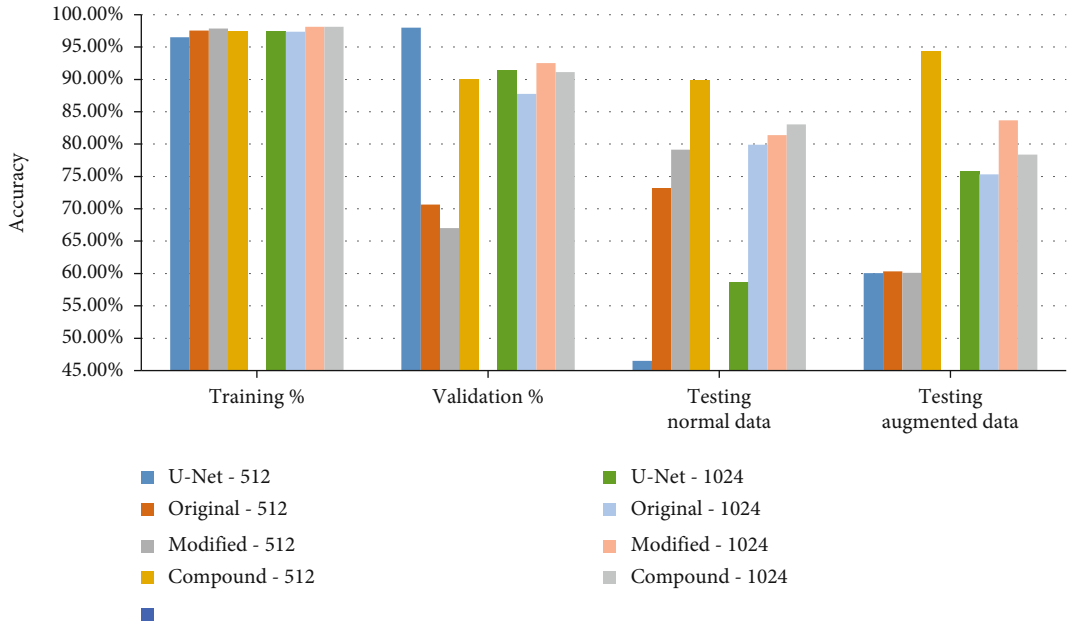


Figure 5: The accuracies in percentage for trading, validation, and testing using normal data and testing using augmented data for all models' structure (U-Net, original bridge, modified bridge, and compound bridge) with both filter structures (32 ➔ 512) and (64➔ 1024).

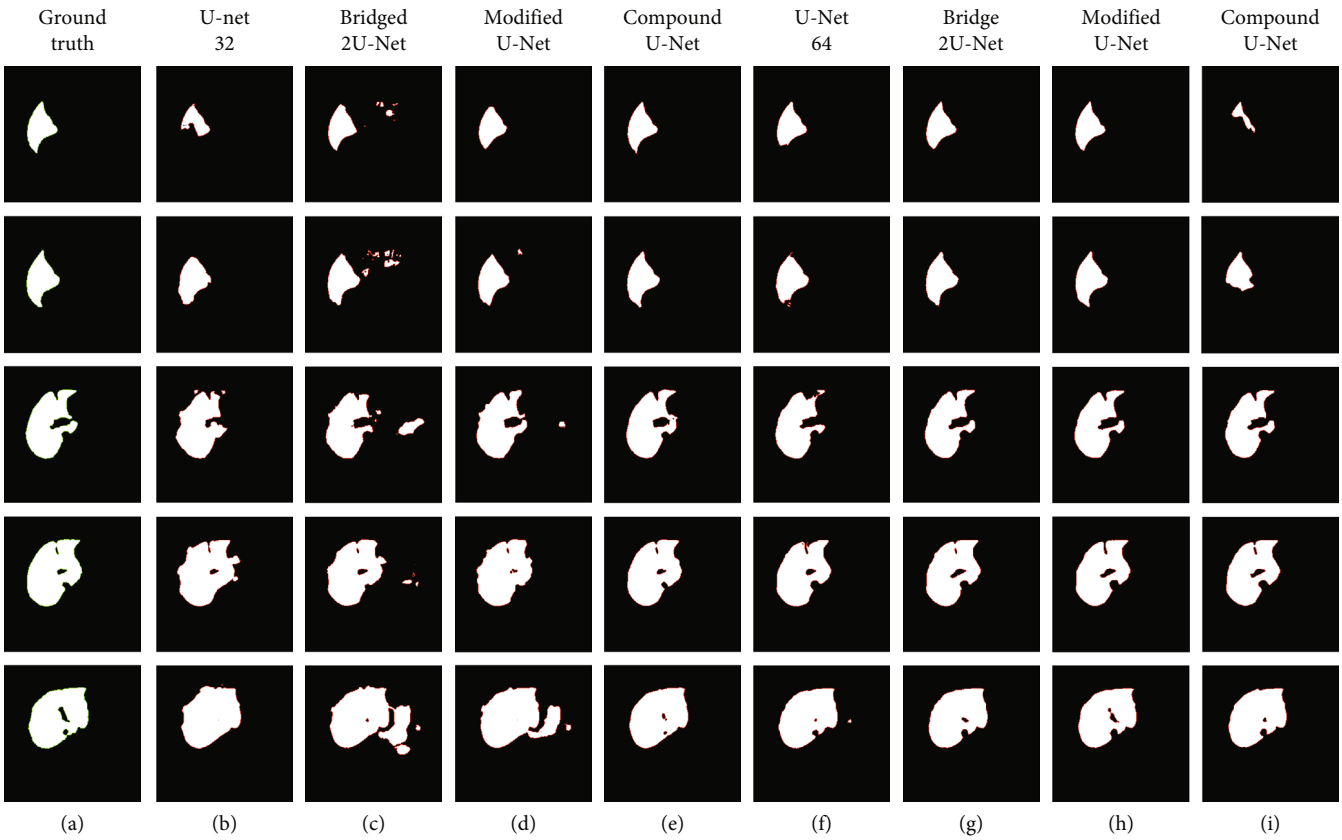


Figure 6: Comparison of testing results. (a) Ground truth, (b) U-Net, (c) bridged 2U-Net, (d) our modified stacked U-Net, (e) our compound 2U-Net. All (b–e) with filter structure 32-512. All (f–i) with filter structure 64-1024.

contrast, [32] did not mention any details about the data splitting approach for training, validation, and testing. We used the mean values of the Dice of all the tested images as the accuracy without calculating the STD or tolerance as the way the results were presented in [24, 28].

Christ et al. [28] proposed a cascaded CNN for liver and lesion segmentation with preprocessing the image to convert it to HU values and postprocessing in 2D with a 3D dense conditional random field (CRF) approach. Although their recorded results exceeded the U-Net performance with 94.30%, our compound model outperformed their results with 94.42%. Wang et al. [32] converted the DICOM images to Hounsfield Unit, then used a window of the specific HU for the liver. Replacing the Conv layers at each level with a dense connection blocks significantly increased the number of layers, and the network ended up similar to 5 dense blocks; each block contains 5 U-Nets of 2 levels. Their model is very complicated and computationally expensive with very large number of layers and connections. It is not clear if the stated results were based on testing with 3Dircadb dataset or LiTS as the model was tested on both datasets. It is evident that our compound model outperformed all the benchmarked models stated in [32]. The performance of the proposed compound model is comparable to that of the densely connected U-Net (DC U-Net). Seo et al. [20] included the residual path and a design of object-dependent upsampling to U-Net structure. The network tried to avoid duplication of low-resolution information by adding block or residual layers on the skip connection while it used only one U-Net.

The original architecture of the U-Net model [16] was reimplemented for comparison purpose in this paper as it was the case in [28, 32]. Note that Wang et al. [32] applied preprocessing technique to convert the Dicom images into Hounsfield Unit (HU) to prevent the loss of information when the whole image pixel values are scaled into the range 0-255. Specifically, the raw CT slices are windowed to a Hounsfield Unit range of -100 to 400 HU to neglect organs and tissues that are not of interest. Christ et al. [28] followed the same preprocessing technique and additionally applied histogram equalization for contrast enhancement followed by augmenting the images to increase the number of samples using translation, rotation, and adding Gaussian noise. On the contrary, we did not use any preprocessing techniques but directly converted the Dicom images to pixel values within the range 0-255. Furthermore, whilst [28] used 15 out of 20 patients' data for training, validation, and testing with 2-fold cross-validation, we used 14 patients' data for training, 4 patients' data for validation, and 2 patients' data for testing. Our testing data might be not included in the 15 patients' data used in [28]. On the other hand, Wang et al. [32] did not mention the methods that used for data splitting. Hence, it explains the discrepancy between the results of our implementation of the U-Net and the results presented in [24, 28].

Our training and testing data were rescaled to 256×256 which is 50% of the size of the original images. While [20] used the original image size 512×512, [28] and [32] did not mention if they used the original data size or rescaled it. Although scaling the images to smaller size might cause the loss of some features, our proposed models outperformed most of the other models in terms of Dice coefficient (see Table 3). We also plan to work with original image size 512×512 in the future for potentially more accurate liver segmentation.

Our approach did not apply any preprocessing technique to the images but directly normalized the DICOM images' pixel intensity to the range 0-255. It thus produced different pixel intensity mapping for the liver region because 3Dircadb1 database contains DICOM image with various HU range. Effectively, the variations of the pixel intensity for the liver region act as an image augmentation technique. Note that image augmentation techniques, such as adding noises to image intensity values amongst others, have been shown to be very useful for deep learning based models. We have compared our results with methods that applied image calibration or windowing process based on Hounsfield Units [28, 32]. Our results are comparable or better in terms of the Dice coefficient. We plan to add preprocessing step that includes windowing process in our future work to compare its effectiveness with the models that do not apply the windowing process.

Although 3D CNNs can process volumetric information, they have some disadvantages. Due to the increased dimension, 3D CNNs require higher computational cost. Besides, the large number of parameters may result in higher risk of overfitting, especially when encountering small datasets. Moreover, the GPU requirement of 3D CNNs is impractically expensive, which hinders their further clinical application. Our model is implemented based on 2D U-Net architecture where each CT slice is treated as input image that is independent of other slices. The proposed model has much lighter computational cost but higher inference speed. In our future work, the information between adjacent slices could be taken into consideration.

4. Conclusions

In this paper, we proposed a novel segmentation method, and the experimental results show that stacking two U-Nets and adding three bridging connections from the first U-Net to the second U-Net can significantly improve the accuracy of liver segmentation. The accumulative approach of concatenating the feature maps from the previous layer with all the previously generated features at the same level from all the previous paths of the 2 stacked U-Nets significantly reduced the loss of global features and low level features during the pooling and upsampling and outperformed most of the recent approaches. The model results are robust against noise as it did not use any preprocessing or postprocessing. Our model used augmentation techniques to overcome the shortage of the medical data with manually annotated masks, and it showed a significant improvement in the performance when testing with the augmented data.

Data Availability

The dataset 3Dircadb1 for liver CT scans and the associated masks for each liver are publicly available on [37] in DICOM format.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors are grateful to the Ministry of Science, Technology and Innovation, Malaysia, for funding this work through the eScienceFund grant scheme under the project 06-02-10-SF0317. IYL would like to thank the Ministry of Higher Education Malaysia for supporting the research through Fundamental Research Grant Scheme under the project FRGS/1/2014/ICT07/UNIM/02/1.

References

[1] S. Priyadarsini, "Survey on segmentation of liver from CT images," in *2012 IEEE international conference on advanced communication control and computing technologies (ICACC CT)*, vol. 978, pp. 234–238, Ramanathapuram, India, 2012.

[2] A. M. Mharib, A. Rahman, S. Mashohor, and R. Binti, "Survey on liver CT image segmentation methods," pp. 83–95, 2012.

[3] S. V. Vanmore, "Survey on automatic liver segmentation techniques from abdominal CT images," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, vol. 37, pp. 1030–1035, Madurai, India, 2019.

[4] "Cancer.Net, liver cancer statistics," 2020, https://www.cancer.net/cancer-types/liver-cancer/statistics.

[5] World Health Organization, "Our world in data," 2020, https://ourworldindata.org/cancer#deaths-from-cancer.

[6] World Health Organization, "World health ranking," 2020, https://www.worldlifeexpectancy.com/cause-of-death/liver-cancer/by-country/.

[7] World Health Organization, "Global Cancer Observatory, International Agency for Research on Cancer, World Health Organization," 2020, https://gco.iarc.fr/.

[8] R. Beichel, C. Bauer, A. Bornik, E. Sorantin, and H. Bischof, "Liver segmentation in CT data : a segmentation refinement approach," in *Proceedings of" 3D Segmentation in The Clinic: A Grand Challenge (2007)*, pp. 235–245, New York, NY, USA, 2007.

[9] Xinjian Chen, J. K. Udupa, U. Bagci, Ying Zhuge, and Jianhua Yao, "Medical image segmentation by combining graph cuts and oriented active appearance models," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2035–2046, 2012.

[10] W. Wu, S. Wu, Z. Zhou, R. Zhang, and Y. Zhang, "3D liver tumor segmentation in CT images using improved fuzzy C-means and graph cuts," *BioMed Research International*, vol. 2017, Article ID 5207685, 11 pages, 2017.

[11] T. Heimann, H. Meinzer, and I. Wolf, "A statistical deformable model for the segmentation of liver CT volumes using extended training data," in *3D Segmentation in the clinic: A grand challenge*, pp. 161–166, New York, NY, USA, 2007.

[12] D. Furukawa, S. Akinobu, and H. Kobakate, "Automatic liver segmentation method based on maximum a posterior probability estimation and level set method," in *3D Segmentation in The Clinic: A Grand Challenge*, pp. 117–124, New York, NY, USA, 2007.

[13] P. Slagmolen, A. Elen, D. Seghers, D. Loeckx, F. Maes, and K. Haustermans, "Atlas based liver segmentation using nonrigid registration with a B-spline transformation model," in *Proceedings of MICCAI workshop on 3D segmentation in the clinic: a grand challenge*, pp. 197–206, New York, NY, USA, 2007.

[14] J. Zhou, W. Xiong, Q. Tian et al., "Semi-automatic segmentation of 3D liver tumors from CT scans using voxel classification and propagational learning," in *2008 MICCAI Work. Gd. Chall. Liver Tumor Segmentation*, New York, NY, USA, May, 2008.

[15] A. Shimizu, T. Narihira, D. Furukawa, H. Kobatake, S. Nawano, and K. Shinozaki, "Ensemble segmentation using AdaBoost with application to liver lesion extraction from a CT volume," in *Proc. MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge II*, vol. 2014, NY, USA, 2008.

[16] 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, Munich, Germany, 2015.

[17] W. Chen, Y. Zhang, J. He et al., "Prostate segmentation using 2D bridged U-net," in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 14–19, Budapest, Hungary, 2018.

[18] Y. Han and J. C. Ye, "Framing U-net via deep convolutional framelets: application to sparse-view CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[19] V. Czeipczr and A. Manno-Kovacs, "Automatic liver segmentation on CT images combining region-based techniques and convolutional features," in *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*, vol. 2019, pp. 2–7, Dublin, Ireland, 2019.

[20] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-net (mU-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1316–1325, 2020.

[21] X. Xia and B. Kulis, "W-Net: a deep model for fully unsupervised image segmentation," 2017, https://arxiv.org/abs/1711.08506.

[22] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, Stanford, CA, USA, 2016.

[23] W. Chen, Y. Zhang, J. He et al., "W-net: bridged U-net for 2D medical image segmentation," pp. 1–13, 2018, https://arxiv.org/abs/1807.04459.

[24] L. Zhang and L. Xu, "An automatic liver segmentation algorithm for CT images U-net with separated paths of feature extraction," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pp. 294–298, Chongqing, China, 2018.

[25] I. Astono, J. S. Welsh, and S. Chalup, "Adjacent network for semantic segmentation of liver CT scans," in *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 35–40, Taichung, Taiwan, 2018.

[26] K. Saito, H. Lu, H. Kim, S. Kido, and M. Tanabe, "ROI-based fully automated liver registration in multi-phase CT images," in *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, pp. 645–649, PyeongChang, Korea (South), 2018.

[27] H. S. Hoang, C. Phuong Pham, D. Franklin, T. Van Walsum, and M. Ha Luu, "An evaluation of CNN-based liver segmentation methods using multi-types of CT abdominal images from multiple medical centers," in *2019 19th international symposium on communications and information technologies (ISCIT)*, pp. 20–25, Ho Chi Minh City, Vietnam, 2019.

[28] P. F. Christ, M. E. Elshaer, F. Ettlinger et al., "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 415–423, Athens, Greece, 2016.

[29] A. A. Albishri, S. J. H. Shah, and Y. Lee, "CU-Net: cascaded U-net model for automated liver and lesion segmentation and summarization," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1416–1423, San Diego, CA, USA, 2019.

[30] Y. Liu, N. Qi, Q. Zhu, and W. Li, "Cr-u-net: cascaded u-net with residual mapping for liver segmentation in ct images," in *2019 IEEE Visual Communications and Image Processing (VCIP)*, pp. 6–9, Sydney, NSW, Australia, 2019.

[31] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong, "Automatic 3D liver location and segmentation via convolutional neural networks and graph cut," *International journal of computer assisted radiology and surgery*, vol. 12, no. 2, pp. 171–182, 2017.

[32] Z. H. Wang, Z. Liu, Y. Q. Song, and Y. Zhu, "Densely connected deep U-net for abdominal multi-organ segmentation," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1415–1419, Taipei, Taiwan, 2019.

[33] Z. Zhou, M. M. Rahman 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.

[34] L. Bi, J. Kim, A. Kumar, and D. Feng, "Automatic liver lesion detection using cascaded deep residual networks," 2017, https://arxiv.org/abs/1704.02703.

[35] W. Xu, H. Liu, X. Wang, and Y. Qian, "Liver segmentation in CT based on ResUNet with 3D probabilistic and geometric post process," in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pp. 685–689, Wuxi, China, 2019.

[36] Y. C. Liu, D. S. Tan, J. C. Chen, W. H. Cheng, and K. L. Hua, "Segmenting hepatic lesions using residual attention U-net with an adaptive weighted dice loss," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 3322–3326, Taipei, Taiwan, 2019.

[37] I. France, "3D-IRCADb (3D image reconstruction for comparison of algorithm database)," 2016, https://www.ircad.fr/research/3dircadb/.

[38] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015*, pp. 1–15, San Diego, CA, 2015.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 7137524, 13 pages
https://doi.org/10.1155/2022/7137524

Research Article

Investigating the Role of Image Fusion in Brain Tumor Classification Models Based on Machine Learning Algorithm for Personalized Medicine

R. Nanmaran,[1] S. Srimathi,[2] G. Yamuna,[2] S. Thanigaivel,[3] A. S. Vickram,[3] A. K. Priya,[4] Alagar Karthick,[5] J. Karpagam,[5] V. Mohanavel,[6] and M. Muhibbullah[7]

[1]*Department of Biomedical Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, 602105 Tamil Nadu, India*
[2]*Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, Annamalai University, Annamalainagar, 608002 Tamil Nadu, India*
[3]*Department of Biotechnology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, 602105 Tamil Nadu, India*
[4]*Department of Civil Engineering, KPR Institute of Engineering and Technology, Coimbatore, 641407 Tamil Nadu, India*
[5]*Department of Electrical and Electronics Engineering, KPR Institute of Engineering and Technology, Coimbatore, 641407 Tamil Nadu, India*
[6]*Centre for Materials Engineering and Regenerative Medicine, Bharath Institute of Higher Education and Research, Chennai, 600073 Tamil Nadu, India*
[7]*Department of Electrical and Electronic Engineering, Bangladesh University, Dhaka 1207, Bangladesh*

Correspondence should be addressed to M. Muhibbullah; m.muhibbullah@bu.edu.bd

Received 10 October 2021; Revised 13 December 2021; Accepted 20 December 2021; Published 7 February 2022

Academic Editor: Iman Yi Liao

Image fusion can be performed on images either in spatial domain or frequency domain methods. Frequency domain methods will be most preferred because these methods can improve the quality of edges in an image. In image fusion, the resultant fused images will be more informative than individual input images, thus more suitable for classification problems. Artificial intelligence (AI) algorithms play a significant role in improving patient's treatment in the health care industry and thus improving personalized medicine. This research work analyses the role of image fusion in an improved brain tumour classification model, and this novel fusion-based cancer classification model can be used for personalized medicine more effectively. Image fusion can improve the quality of resultant images and thus improve the result of classifiers. Instead of using individual input images, the high-quality fused images will provide better classification results. Initially, the contrast limited adaptive histogram equalization technique preprocess input images such as MRI and SPECT images. Benign and malignant class brain tumor images are applied with discrete cosine transform-based fusion method to obtain fused images. AI algorithms such as support vector machine classifier, KNN classifier, and decision tree classifiers are tested with features obtained from fused images and compared with the result obtained from individual input images. Performances of classifiers are measured using the parameters accuracy, precision, recall, specificity, and $F1$ score. SVM classifier provided the maximum accuracy of 96.8%, precision of 95%, recall of 94%, specificity of 93%, $F1$ score of 91%, and performed better than KNN and decision tree classifiers when extracted features from fused images are used. The proposed method results are compared with existing methods and provide satisfactory results.

1. Introduction

Early detection of cancer plays a vital role in the healthcare industry because when abnormal tissue or cancer is found early, it is easy to plan successful treatment [1, 2]. If cancer spreads to neighbor cells, it is challenging to treat, and survival chances are much lower. Many machine learning techniques were developed to detect cancer at early stages [3, 4]. Still, a tool with more accuracy and less processing time is needed. This research is aimed at detecting cancer irrespective of its types by combining two imaging modalities, such as CT/PET or MRI/SPECT, which provides better accuracy than existing methods. Finally, the concept is extended to cancer classification to predict the tumor type, whether it belongs to benign or malignant tumors [5, 6]. While there are many studies in image fusion for the visual enhancement of images, very few types of research focus on the influence of image fusion in other applications such as image classification. The role of image fusion in remote sensing is significant, and almost every recent sensor development for earth applications considers channels with different spatial resolutions [7, 8]. Image fusion improves the information content by combining two or more images using a specific algorithm [3]. A case study on the influence of image fusion approaches on classification accuracy for remote sensing applications is more relevant to this research work. They took Landsat 7 ETM+ image for the analysis. They analyzed various image fusion approaches such as adaptive image fusion (AIF), wavelet-based fusion, multisensor multiresolution fusion technique (MMT), principal component analysis (PCA), hue saturation-value transform (HSV), and Brovey method of fusion. The fusion results were employed for maximum likelihood classifier, object-based classification, and support vector machine classifier [9]. The Brovey Transform (BT) was created to visually boost contrast between the image histogram's low and high ends. BT produce the spectral degradation, which should not be employed if the original scene radiometry must be preserved. Brovey Transform resulted in low dynamic range resultant image and significant misclassification detected when using pixel-based classifiers. HSV fusion was not assigning the discrete classes from the training dataset. Another method called principal component analysis performed better because of its capability of separately high- and low-frequency parts of an image. MMT fusion technique produced "Speckle" noise in the classification, which results in poor edge detection. They concluded in the case study that wavelet transform-based fusion improves the classification accuracy, and they recommended the classification application. Wavelet transforms are used to portray abrupt peaks and discontinuities. Wavelet transform has a number of drawbacks, including shift sensitivity and directional selectivity [10, 11].

The research on the effects of image fusion algorithms on classification accuracy in remote sensing applications was conducted. They took Quickbird-02 panchromatic and multispectral images over the city of Wuhan (China) and analyzed the effects of fusion on unsupervised ISODATA (iterative self-organizing data analysis) classifier accuracy. They considered eight image fusion techniques and analyzed the effects of fusion on improving classification accuracy. They found that the high pass filter-based fusion method performs worst because of its injection of high-frequency information [12]. They concluded that the region-based image fusion method improves the accuracy of the classifier. Even though colour distortion occurs, the contrast of the fused image is higher, which is very helpful in interpretation and classification, and they recommended a region-based fusion technique for classification application. Region-based image fusion has several advantages over pixel-based image fusion methods, such as being less susceptible to noise, more resilient, and avoiding misregistration, but it is also more difficult.

Kumar et al. have proposed a novel MSLN-CNN method to solve the HSI classification problems, multilayer spatial-spectral feature fusion and sample augmentation with local and nonlocal constraints (MSLN-CNN). The authors make full use of complementary spatial-spectral information among different layers, and compared with other deep learning models, CNN has two unique structures: local connection and weight sharing. The representative classifiers include k-nearest neighbors, logistic regression, support vector machine (SVM), sparse representation-based classification, and extreme learning machine [13, 14]. Among these classifiers, SVM seeks to separate the samples with different classes. Finally, the authors concluded that MSLN-CNN could achieve end-to-end classification by optimizing multilayer spatial-spectral feature fusion jointly. It is a promising method to deal with the overfitting problem by considering local spatial and nonlocal spectral constraints.

Chandran et al. have developed an approach for fusing features obtained from multisensor compressive measurements for spectral image classification. This fusion method merged the components extracted from data captured by sensors that satisfy the Nyquist–Shannon sampling theorem. They developed a low-resolution feature as degraded versions of the high-resolution features [15]. Also, they formulated an inverse problem that aims at estimating high-resolution characteristics, including both a sparsity-inducing term and a total variation (TV) regularization term to exploit the correlation between neighboring pixels of the spectral image. Therefore, they improved the performance of pixel-based classifiers. Besides, they introduce the mathematical model of the high-resolution features. This model describes the relationship between the high-resolution elements and the spectral image under test. The authors described an algorithm based on the alternating direction method of multipliers (ADMM) for solving the fusion problem. The proposed feature fusion approach is tested for two CSI architectures: three-dimensional coded aperture snapshot spectral imaging (3D-CASSI) and colored CASSI (C-CASSI). Finally, they compared the proposed feature fusion method concerning state-of-the-art feature extraction and fusion techniques which improves the accuracy and robustness to noise [16, 17]. They concluded that the ADMM algorithm performs better when compared to existing methods. Because of its fast convergence speed in many applications, the ADMM technique has sparked a lot of academic attention in recent years.

Most of the related research was attempted only in remote sensing, focusing on medical imaging applications. Based on the above literature, many researchers reported that image fusion could improve image quality, and no research analyzes with experiment results. This research work shows how image fusion improves the visual quality of medical images and how these images further improve classification accuracy with experiment results. This research work combines brain tumor images from MRI and SPECT modalities using the discrete cosine transform (DCT) method. The magnetic resonance interference (MRI) imaging technique can produce incredibly detailed diagnostic images of most of body's major organs and tissues that other imaging techniques cannot. Because MRIs do not involve radiation, they are safe for children and pregnant people to undergo. The SPECT (single-photon emission computed tomography) imaging technology can be used to determine whether or not there is enough blood flow to various parts of the brain. SPECT scans can be used to acquire data on changes in brain function as a result of disease. Benefits of two imaging modalities can be attained in a single image when these two modalities are merged. The fused image provides more information to the doctors than the individual input images. Apart from that, when compared to individual input images, the number of features or attributes recovered from fused images is higher. As a result, cancer classification will be more precise, allowing doctors and other health care providers to better plan therapy for their patients. In order to combine MRI and SPECT images, discrete cosine transform-based image fusion technique has developed with improved fusion parameters such as peak signal to noise ratio (PSNR), structural similarity index measure (SSIM), and normalized correlation (NC). High values of PSNR in dB, SSIM in %, and NC in % indicates improved fusion between two images. This proposed DCT-based fusion technique provides high PSNR, SSIM, and NC which is used for fusing MRI and SPECT images in this research work.

Three classifiers, namely, support vector machine, K-nearest neighbour classifier, and decision tree classifiers, are used in this research work to classify the brain tumour images into benign or malignant classes. The performance of these classifiers is measured and compared based on accuracy, precision, recall, specificity, and $F1$ score. In order to analyze the effect of image fusion, the experiment results are compared with the result obtained when MRI input image alone was used, SPECT input image alone was used for classification. Algorithms of all these three classifiers are explained in the "Research Methods" section, and the performance metrics are explained in the next section.

2. Research Methods

The proposed brain tumour classification model is shown in Figure 1.The input images such as MRI and SPECT images are collected from http://kaggle.com for analysis. Image registration was applied to input images before applying the image fusion method. Image fusion using discrete cosine transform was developed, and fused images were obtained. Thirty features were extracted from fused images, and they were given as input to three classifiers. Classifier's performance is measured using accuracy, precision, recall, specificity, and $F1$ score. These results were compared with the classifier results when features extracted from individual input images were given to the classifiers. High values of performance metrics indicate the better performance of classifiers.

2.1. Preprocessing of Brain Tumour Images. The visual quality of input medical images such as MRI and SPECT images are improved by applying contrast limited adaptive histogram equalization (CLAHE) technique [18, 19]. Instead of operating on entire image pixels, CLAHE operates only on a small region of the image called tiles [20, 21]. Image registration needs to be applied before applying the image fusion technique. A parametric transformation $\mathrm{Ta}(.)$ is applied to the target image to maximize the similarity between the target image and the reference image to make it similar to I_r [22]. The optimization target can be represented as in Equation (1).

$$\mathrm{Ta}(.) = \arg\max \rho(\mathrm{I_r}, \mathrm{T_a}(\mathrm{I_t})). \tag{1}$$

Various steps involved in image registration include similarity measure, point detection and extraction, applying image descriptors, point selection, pattern matching, image resampling, and compositing [23].

2.2. Image Fusion Using Discrete Cosine Transform. Fusion can be performed in spatial or frequency-domain methods, and DCT belongs to the frequency domain category. When DCT is applied to input images, it decomposes the images into DCT coefficients, and in the transform domain, the fusion rules are applied [24–26]. After applying inverse DCT, a fused image is obtained in the spatial domain [27, 28]. Input images such as MRI and SPECT images are applied with discrete cosine transform, which provides DCT coefficients, and averaging fusion rule is applied in the frequency domain, which provides fused DCT coefficients. Inverse discrete cosine transform is applied on fused DCT coefficients to obtain fused images at spatial domain [25, 29–32]. Image fusion using discrete cosine transform is shown in Figure 2.

Discrete cosine transform of two-dimensional image $X(n_1, n_2)$ of size $M \times N$ is represented by Equation (2). Here, M represents number of rows in an image, and N represents number of columns in an image.

$$X(k_1, k_2) = \sum_{n_1=0}^{M-1} \sum_{n_2=0}^{N-1} x(n_1, n_2) \cos\left(\frac{\pi(2n_1+1)k_1}{2M}\right) \cos\left(\frac{\pi(2n_2+1)k_2}{2N}\right), \tag{2}$$

where n_1, n_2 are the spatial domain coordinates and k_1, k_2 are the frequency domain coordinates.

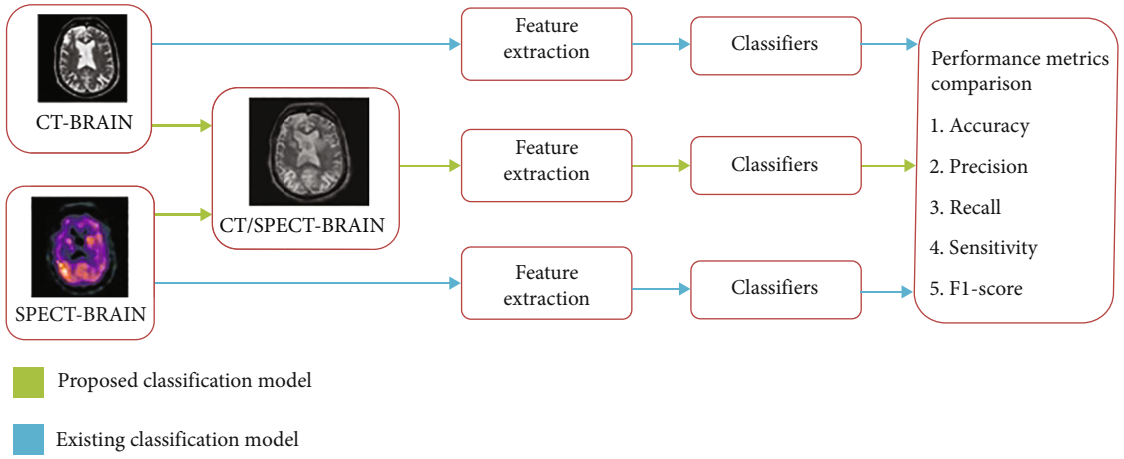


Figure 1: Proposed fusion-based brain tumor classification model.

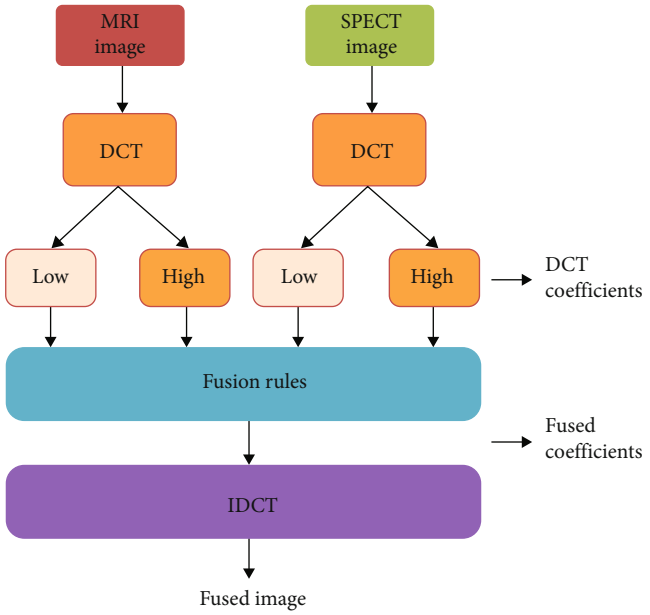


Figure 2: Discrete cosine transform-based image fusion methodology.

Similarly, inverse discrete cosine transform of two-dimensional image $X(k_1, k_2)$ of size $M \times N$ is represented by

$$X(k_1, k_2) = \sum_{k_1=0}^{M-1} \sum_{k_2=0}^{N-1} x(k_1, k_2) \cos\left(\frac{\pi(2n_1+1)k_1}{2M}\right) \cos\left(\frac{\pi(2n_2+1)k_2}{2N}\right), \quad (3)$$

where k_1 ranges from 0 to $M-1$ and k_2 ranges from $N-1$.

2.3. Support Vector Machine Classifier. SVM classifiers are supervised learning methods that are used for regression and classification.

The SVM classifier can maximize the geometric margin and minimize the classification error, and hence, the SVM classifier can also be called maximum margin classifiers [33–40]. SVM classifiers are not biased by outliers and not sensitive to overfitting but they are not suitable when huge number of features used and for nonlinear problems.

Let us consider a dataset $(A_1, B_1, ..A_n, B_n)$, where $(A_1, A_2..A_n)$ is the set of the input variable, $(B_1, B_2.., B_n)$ is the output variable, and 'C' is the intercept, then the SVM classifier is given as in

$$\text{SVM} = \sum_{m=1}^{i} \beta_m - \frac{1}{2} \sum_{m,n=1}^{i} b_m b_n C(a_m, a_n) \beta_m \beta_n, \quad (4)$$

where $m = 1, 2, 3 \cdots .i$ and $C = b_m \beta_m + b_n \beta_n$.

2.4. K-Nearest Neighbour Classifier. K-nearest neighbor classification is more suitable for large datasets, and it takes more computation time for testing than training the dataset. The K-nearest neighbor classification technique is the most straightforward technique that provides good classification accuracy and stability [41, 42]. The K-NN algorithm is based on distance functions such as Manhattan, Minkowski, Tanimoto, Jaccard, Mahalanobis, and Euclidean distance in which Euclidean distance is more common [43, 44]. It is mathematically given in

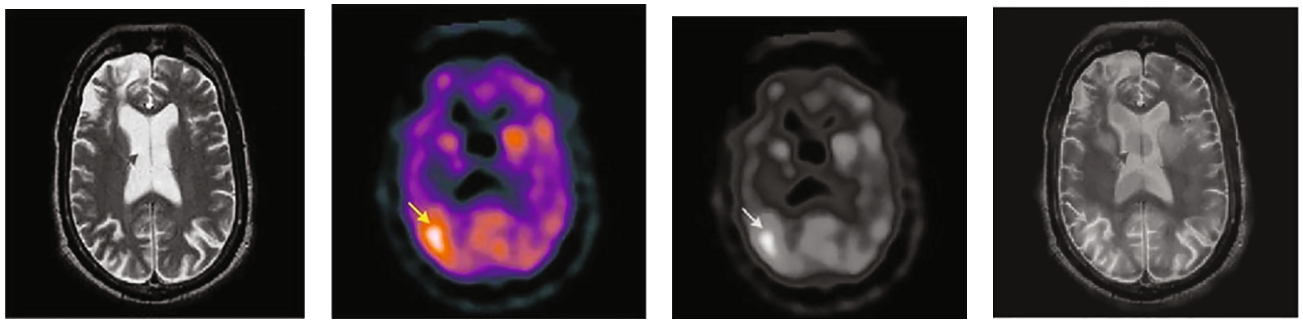
$$d(a, b) = \sqrt{\sum_{i=1}^{m} (b_{i-} a_i)^2}, \quad (5)$$

where

a, b = Two points in Euclidean distance
a_i, b_i = Euclidean Vectors
m = m-space.

Even though the KNN classifier runs slowly, its computational power is more. KNN algorithm consists of the training phase and testing phase. Features or attributes are stored during the training phase, and while in the testing phase, the features from testing images are compared with stored features, and the class will be determined [45].

K-NN algorithm consists of the following steps:



(a) (b) (c) (d)

Figure 3: (a) Input image-1 (MRI-brain tumor. (b) Input image-2 (SPECT-brain tumor. (c) Converted grey scale image of input image-2. (d) Fused image (MRI-SPECT).

(1) Determine suitable distance metrics like Euclidean distance

(2) Store all the training datasets in the training phase

(3) During the testing phase, compute the distances between the new feature and stored features

(4) The correct classification access given in the test phase is used to assess the correctness of the algorithm

2.5. Decision Tree Classifier. A decision tree is a supervised learning technique that can be used for both regression and classification problems, but it is mainly used for classification problems. Decision tree classifiers are appropriate for both linear and nonlinear systems; however, they are ineffective when dealing with limited datasets. In a decision tree classifier, overfitting is a common occurrence. Nodes in the tree-like structure represent the features of a dataset, branches represent the decision rules, and leaf node represents the output [46, 47]. The decision trees usually resemble human thinking ability; hence, it is easy to understand. The algorithm starts at the root node and compares the attributes of root nodes with attributes of record nodes. Based on the comparison, it jumps to the next node. Many algorithms are proposed for learning decision tree from a given dataset, but commonly, ID3 algorithm is preferred due to its simplicity for implementation [48–51]. ID3 algorithm is a top-down greedy search of possible branches, and it uses information gain and entropy to build the tree.

The $H(Y)$ Shannon entropy of a random discrete variable Y with possible $Y_1, Y_2 \cdots .Y_n$ and probability mass function $P(Y)$ is defined as in

$$H(Y) = -\sum_{i=1}^{n} P(y_i) \log_2 P(y_i). \quad (6)$$

Entropy is equal to zero for a completely homogeneous dataset, and entropy is equal to one of the datasets equally divided. A branch with entropy more than one needs splitting.

2.6. Experimental Setup. MATLAB version 2021 software is used for this research work. Proposed work consists of steps such as preprocessing, image fusion, feature extraction, and image classification. For preprocessing using CLAHE, DCT-based image fusion technique and feature extraction MATLAB script have developed, and result is obtained. For testing with different classifiers and analysis, MATLAB inbuilt classification learner application is used. The result which obtained feature extraction is stored in excel file, and this file is directly given as input to classification learner application in order to analyze the effect of image fusion on SVM, K-NN, and decision tree classifiers.

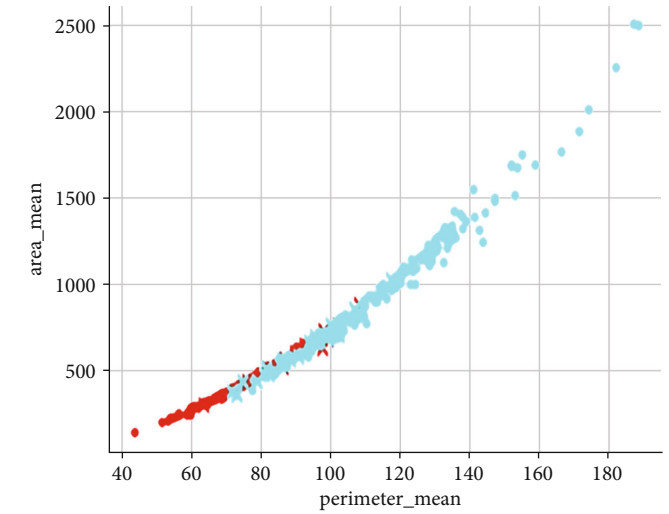
3. Results and Discussion

3.1. Comparison of Our Method with KNN Classifier and Decision Tree Classifier. Input images such as brain tumors from CT modality and SPECT modality are shown in Figures 3(a) and 3(b), respectively. A converted grayscale image of SPECT image is shown in Figure 3(c). After applying with CLAHE for image enhancement and registration, the two input images are applied with DCT-based image fusion technique and the resultant fused image Figure 3(d).

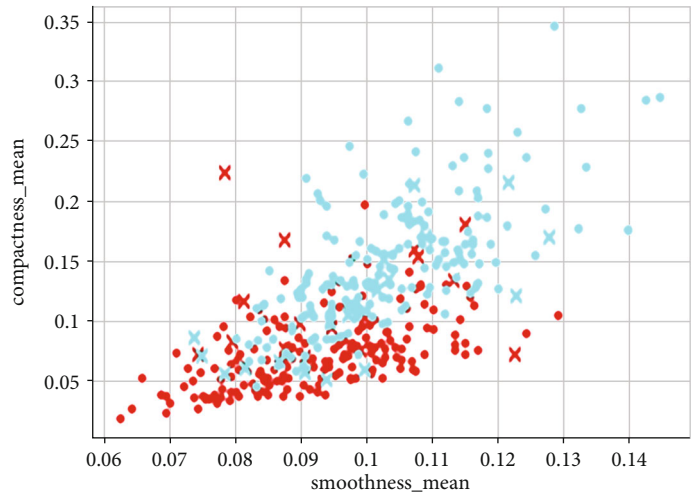
Two hundred MRI images and 200 SPECT images are fused, providing 200 fused images. Features such as radius, area, perimeter, texture, smoothness, compactness, concave points, and concavity are considered, and thus, six features in total are extracted from fused images and individual input images such as MRI and SPECT images. A scatter plot representing some of the features is shown in Figure 4. These features are given as input to SVM, KNN, and decision tree classifiers, and the results are compared to identify the best classifier for brain tumor detection.

The performances of classifiers are measured using accuracy, precision, recall, specificity, $F1$ score, confusion matrix, and ROC curve. The confusion matrix typically consists of four different results, namely, true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Performance measures of classifiers are described in this section.

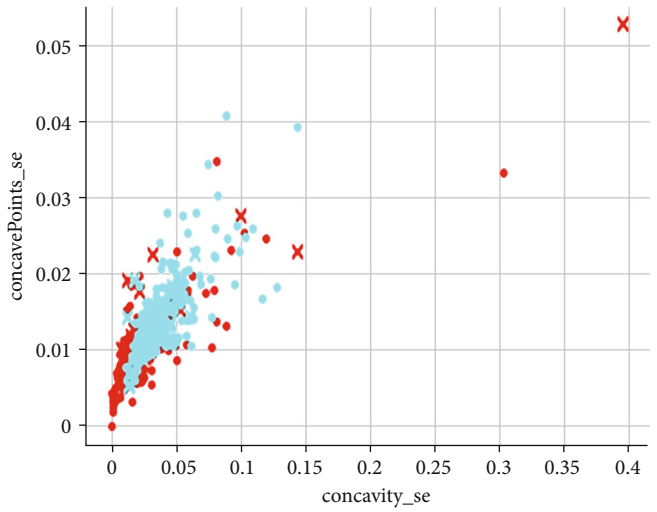
3.2. Accuracy. The accuracy of the classifier is defined as the ratio of total number of correct predictions to the total



(a)

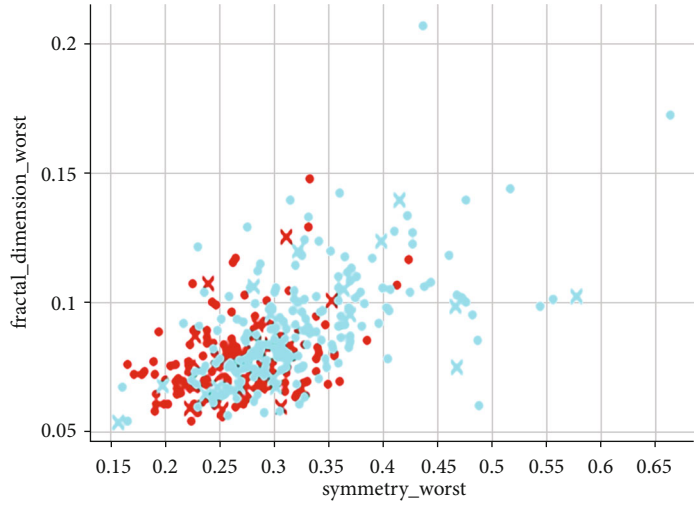


(b)



(c)

Figure 4: Continued.



(d)

Figure 4: (a) Scatter plot representation of mean perimeter and mean area. (b) Scatter plot representation of mean smoothness and mean compactness. (c) Scatter plot representation of standard error mean concavity and standard error mean concave points. (d) Scatter plot representation of worst symmetry and worst fractal dimension.

number of predictions. Accuracy is defined in Equation (7), as shown below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}. \tag{7}$$

3.3. Precision. The precision of the classifier is defined as the correct prediction ratio to the actual prediction of brain tumor cases represented

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{8}$$

3.4. Recall. Recall of the classifier is defined as the ratio of correctly predicted brain tumor cases to overall brain tumor cases, including nonbrain tumor cases. The recall is defined in

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{9}$$

3.5. Specificity. The specificity of the classifier is defined as the ratio of correctly identified nonbrain tumor cases to negative brain tumor cases, which is represented as in

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{10}$$

3.6. F1 Score. It is the measure of average between precision and recall of classifiers defined in

$$F1\ \text{Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{11}$$

3.7. Confusion Matrix. The performance of the classifier in matrix form is given by the confusion matrix of the prediction model. It consists of correctly identified brain tumor cases, missclassified brain tumor results, correctly identified nontumor cases, and missclassified nonbrain tumor cases.

3.8. Receiver-Operating Characteristic (ROC). The ROC curve graph graphically illustrates the performance of the classifier. This graph shows the relationship between the actual positive rate and the false-positive rate.

Extracted features from fused images are tabulated and plotted using parallel coordinates plot as shown in Figure 5.

The confusion matrix of the SVM classifier, *K*-NN classifier, and decision tree classifier is shown in Figures 6(a)–6(c), respectively. Performance measures such as accuracy, precision, recall, specificity, and *F*1 score for SVM, KNN, and decision tree classifiers are tabulated in Table 1.

The classifiers are stored with extracted feature values during training phase. During testing phase, classifiers are tested with different input images. Initially, the classifiers are tested with cancerous input images. If the classifier predicts the cancerous image output correctly, this condition can be taken as true positive (TP) but if it is wrongly predict as noncancerous image, then the condition is known as false positive (FP). When the classifiers are tested with noncancerous input images, if the classifier predicts the noncancerous image output correctly, this condition can be taken as true negative (TN) but if it is wrongly predict as cancerous image, then the condition is known as false negative (FN). Based on this TP, FP, TN, and FN values obtained from confusion matrix as shown in Figures 6(a)–6(c), the performance measures of classifiers are calculated using Equations (7) to (11) and tabulated as shown in Table 1.

From tabulated results, it is inferred that the SVM classifier provides high accuracy, precision, recall, specificity, and *F*1 score parameters and thus performs better than *K*-NN and decision tree classifier.

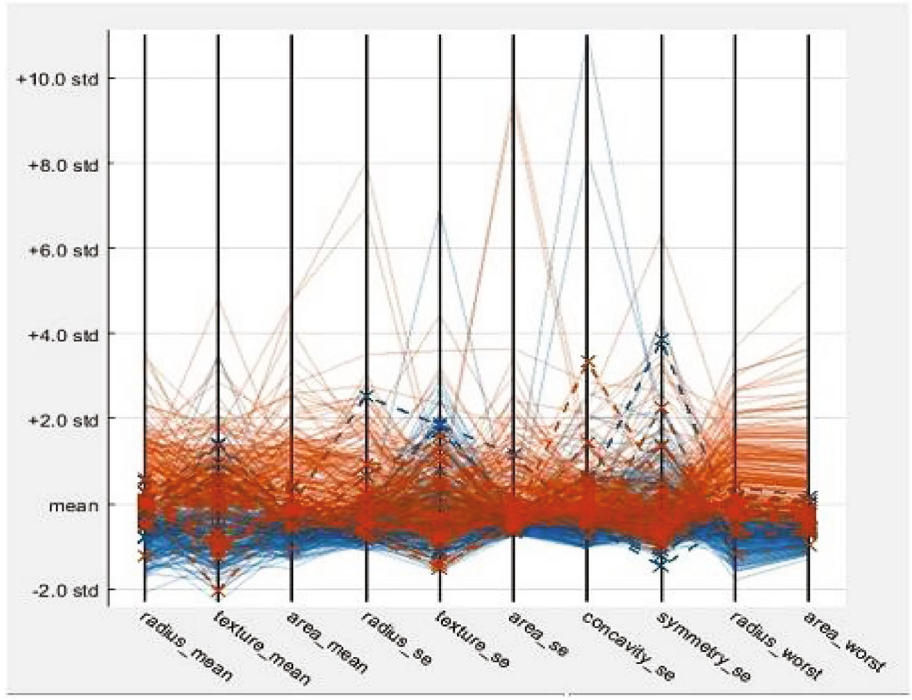


Figure 5: Parallel coordinates plot of features extracted from malignant type tumor.

The performance of classifiers can also be predicted using the response of the ROC curve based on the value of the area under the curve. The area under curve values of the SVM, KNN, and decision tree classifier is 0.994, 0.932, and 0.921, as shown in Figures 7(a)–7(c). From observation, it is noted that the SVM classifier gives the maximum area under the curve value of 0.994, which indicates the better performance of the classifier over the other two classifiers.

3.9. Comparison of Our Method with Features Extracted From MRI Image Alone Given as Input to Classifiers. Features are extracted from 200 samples of MRI images collected from the medical database, and these features are given as input to SVM, KNN, and decision tree classifiers. The results of classifiers are tabulated in Table 2.

From tabulated results, it is inferred that SVM classifier when fused images are considered for feature extractions provides an accuracy of 96.3%, the precision of 97.5%, recall of 95.12%, specificity of 97.13%, and F1 score of 96.29%. These five parameters are high values and thus perform better than SVM, K-NN, and decision tree classifiers when MRI image alone is considered for feature extraction.

3.10. Comparison of Our Method with Features Extracted From SPECT Image Alone Given as Input to Classifiers. Features are extracted from 200 SPECT images collected from a medical database, and these features are given as input to SVM, KNN, and decision tree classifiers. The results of classifiers are tabulated in Table 3.

From tabulated results, it is inferred that SVM classifier when fused images are considered for feature extractions provides an accuracy of 96.8%, the precision of 97.5%, recall of 95.12%, specificity of 97.13%, and $F1$ score of 96.29%. These five parameters are high values and thus perform better than SVM, K-NN, and decision tree classifiers when SPECT image alone is considered for feature extraction.

3.11. Comparison of Our Method with SVM, KNN Classifier, and Decision Tree Classifier In Terms of Consumed Time. In this test, we compare the implementation time of our proposed method and SVM, KNN, and decision tree classifiers when features from SPECT image alone are given as input to classifiers, and the obtained result is shown in Table 4.

Ten independent experiments are taken on PC with Intel Core 3 processor and PC with Celeron 3.06G/1G processor, and the average time is calculated. From the tabulated results, it is observed that the proposed SVM classifier when features from fused images are given as input to classifier consumes longer time, almost 420 seconds, to execute the results, whereas SVM, KNN, and decision tree classifiers when features from SPECT image alone given as input to classifier consume 128, 160, and 180 seconds, respectively. The existing methods involve steps like preprocessing, feature extraction, segmentation, and classification, whereas the proposed method involves steps like preprocessing, image fusion, feature extraction, segmentation, and classification. Because of the additional step involved, the proposed method consumes more time to execute the results when compared to existing research methods. The experiment results of the proposed method are compared with relevant literature and presented in Table 5.

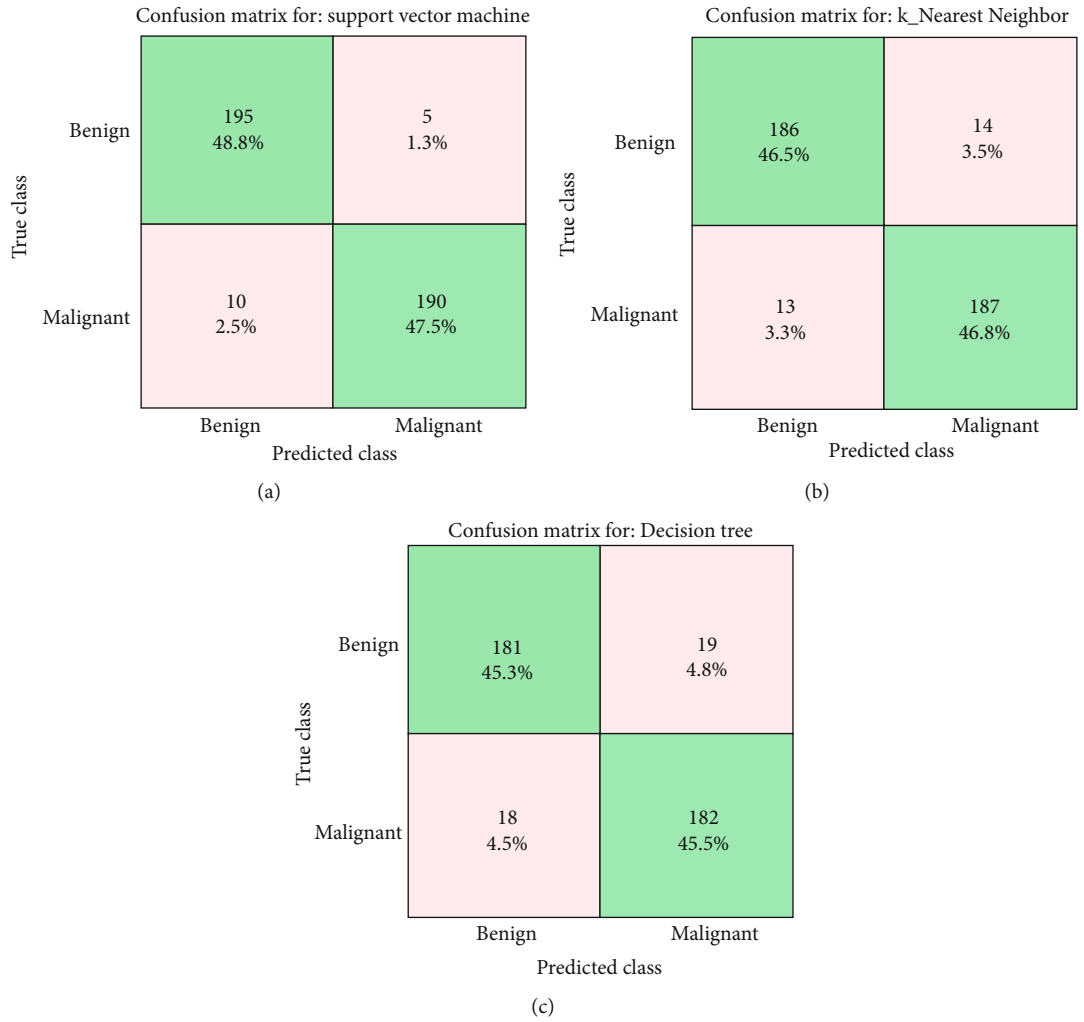


Figure 6: (a) Confusion matrix of the SVM classifier. (b) Confusion matrix of the K-NN classifier. (c) Confusion matrix of the decision tree classifier.

Table 1: Performance measures of SVM, KNN, and decision tree classifiers when features extracted from fused images are given as input.

S. no.	Classifier name	TP	FP	TN	FN	Accuracy	Precision	Recall	Specificity	$F1$ score
1	SVM	195	5	190	10	96.8	97.5	95.12	97.43	96.29
2	K-NN	186	14	187	13	93.3	93	93.46	93.03	93.23
3	Decision tree	181	19	182	18	90.8	90.5	90.95	90.54	90.72

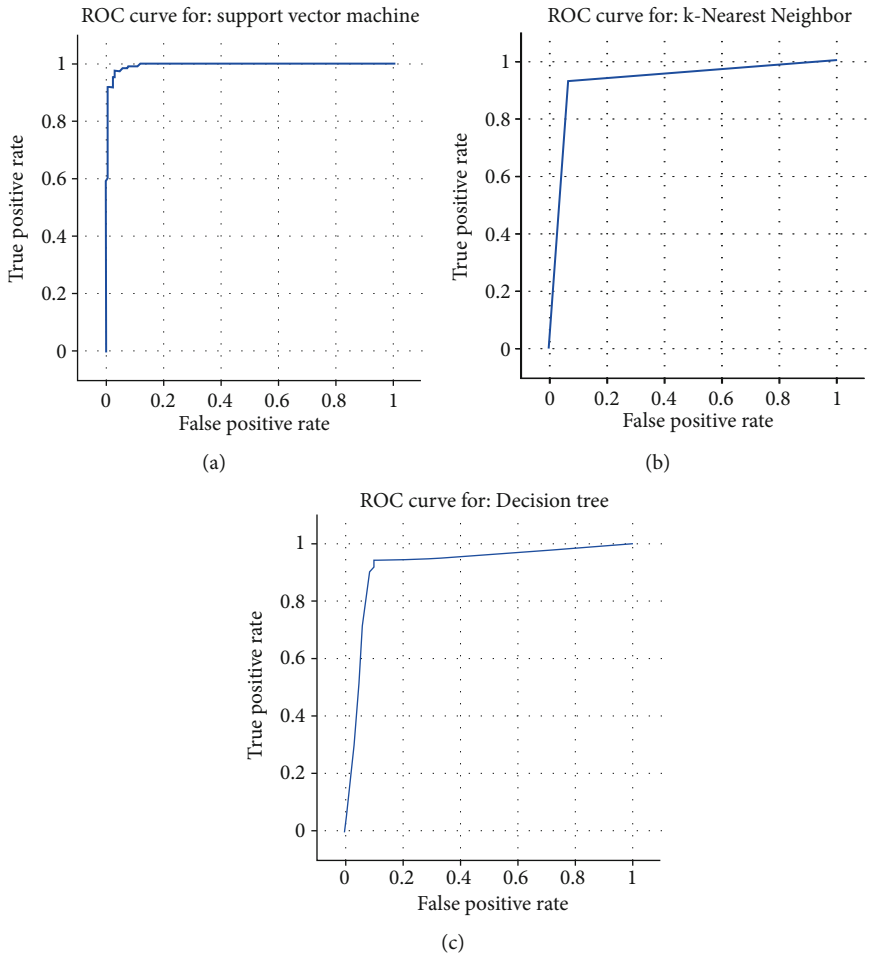


Figure 7: (a) ROC curve of the SVM classifier. (b) ROC curve of the KNN classifier. (c) ROC curve of the SVM classifier.

Table 2: Performance comparison of proposed method and SVM, KNN, and decision tree classifiers when MRI image alone given as input.

S. no.	Classifier name	TP	FP	TN	FN	Accuracy	Precision	Recall	Specificity	*F*1 score
1	SVM (proposed)	195	5	190	10	96.8	97.5	95.12	97.43	96.29
2	SVM	190	10	186	14	94	95	93.21	94.89	94.09
3	*K*-NN	180	20	181	19	90.25	90	90.45	90.04	90.20
4	Decision tree	174	26	172	28	86.50	87	86.86	86	87

Table 3: Performance comparison of proposed method and SVM, KNN, and decision tree classifiers when SPECT image alone given as input.

S. no.	Classifier name	TP	FP	TN	FN	Accuracy	Precision	Recall	Specificity	*F*1 score
1	SVM (proposed)	195	5	190	10	96.8	97.5	95.12	97.43	96.29
2	SVM	192	8	188	12	95	96	94.11	95.91	95.04
3	*K*-NN	182	18	185	15	91.75	91	92.38	91.13	91.68
4	Decision tree	178	22	179	21	89.25	89	89.44	89.05	89.21

Table 4: Performance comparison of proposed method and SVM, KNN, and decision tree classifier based on consumed time.

Parameter	Classifiers			
	SVM (proposed)	SVM	*K*-NN	Decision tree
Time consumed (seconds)	420	128	160	180

Table 5: Comparison of our method with existing methods from literature.

Reference	Classifiers name	Accuracy (%)	Sensitivity (%)	Precision (%)
Present study	SVM, KNN, decision tree	96.80	97.43	97.5
Masoudi S et al. 2021 [52]	Resnet-101C	86.3	NA	NA
Welikala RA et al. 2020 [53]	R-CNN	NA	89.51	84.77
Anupama et al. 2019 [54]	CNN-capsule network	92.5	93	96
T Nguyen et al. 2019 [55]	CNN	73.68	NA	NA
Erkal B et al. 2020 [56]	Multilayer perceptron	97	NA	NA

4. Conclusion

Brain tumor images from MRI and SPECT modalities are considered and applied with CLAHE method to preprocess the images, and then, DCT-based fusion technique was applied to obtain fused images. Features from fused images are extracted and inputted to SVM, KNN, and the decision tree classifier. SVM classifier provides the maximum accuracy of 96.8%, precision of 97.5%, recall of 95.12%, sensitivity of 97.43%, and *F*1 score of 96.29%, which is higher when compared to SVM, KNN, and decision tree classifier when features from either MRI or SPECT image is alone given as input to classifiers. As this research involved image fusion and preprocessing feature extraction and image classification, it took a long time to execute the results compared to conventional cancer classification models. These novel fusion-based AI algorithms can be more suitable for personalized medicine. In the future, transform-based image fusion approaches such as Curvelet transform and Shearlet transform can be applied to input images, and the classification performance can be measured.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] C. Pohl and J. L. Van Genderen, "Review article multisensor image fusion in remote sensing: concepts, methods and applications," *International journal of remote sensing*, vol. 19, no. 5, pp. 823–854, 1998.

[2] C. Pohl and J. L. van Genderen, "Multisensor fusion: optimization and operationalization for mapping applications," in *Signal Processing, Sensor Fusion, and Target Recognition III*, vol. 2332, pp. 17–25, International Society for Optics and Photonics, 1994.

[3] R. R. Colditz, T. Wehrmann, M. Bachmann et al., "Influence of image fusion approaches on classification accuracy: a case study," *International Journal of Remote Sensing*, vol. 27, pp. 3311–3335, 2006.

[4] V. ME, V. DR, M. MK, and S. MP, "A novel technique for optimizing panchromatic and multispectral image fusion using discrete wavelet transform," *International Journal of Engineering and Technology*, vol. 10, pp. 247–260, 2018.

[5] D. Sara, A. K. Mandava, A. Kumar, S. Duela, and A. Jude, "Hyperspectral and multispectral image fusion techniques for high resolution applications: a review," *Earth Science Informatics*, vol. 14, pp. 1685–1705, 2021.

[6] X. Feng, L. He, Q. Cheng, X. Long, and Y. Yuan, "Hyperspectral and multispectral remote sensing image fusion based on endmember spatial information," *Remote Sensing*, vol. 12, pp. 1–17, 2020.

[7] A. I. Yuhendra, J. T. Sri Sumantyo, and H. Kuze, "Spectral quality evaluation of pixel-fused data for improved classification of remote sensing images," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 63, pp. 483–486, 2011.

[8] R. Singh and R. Gupta, "Improvement of classification accuracy using image fusion techniques," in *2016 International Conference on Computational Intelligence and Applications (ICCIA)*, pp. 36–40, Jeju, Korea (South), 2016.

[9] S. Li and Z. Li, "Effects of image fusion algorithms on classification accuracy," in *2010 18th International Conference on Geoinformatics*, Beijing, China, 2010.

[10] U. Subramaniam, M. M. Subashini, D. Almakhles, A. Karthick, and S. Manoharan, "An expert system for COVID-19 infection tracking in lungs using image processing and deep learning techniques," *BioMed Research International*, vol. 2021, 17 pages, 2021.

[11] S. S. Ganesh, G. Kannayeram, A. Karthick, and M. Muhibbullah, "A novel context aware joint segmentation and classification framework for glaucoma detection," *Computational and Mathematical Methods in Medicine*, vol. 2021, 19 pages, 2021.

[12] M. Elkholy, M. M. Hosny, and H. M. Farid El-Habrouk, "Studying the effect of lossy compression and image fusion

on image classification," *Alexandria Engineering Journal*, vol. 58, pp. 143–149, 2019.

[13] P. M. Kumar, R. Saravanakumar, A. Karthick, and V. Mohanavel, "Artificial neural network-based output power prediction of grid-connected semitransparent photovoltaic system," *Environmental Science and Pollution Research 29*, pp. 10173–10182, 2022.

[14] N. Jeevanand, P. A. Verma, and S. Saran, "Fusion of hyperspectral and multispectral imagery with regression kriging and the Lulu operators; a comparison," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII–5, pp. 583–588, 2018.

[15] V. Chandran, M. G. Sumithra, A. Karthick et al., "Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images," *BioMed Research International*, vol. 2021, 15 pages, 2021.

[16] D. Jiang, D. Zhuang, Y. Huang, and J. Fu, "Survey of multispectral image fusion techniques in remote sensing applications," *New Advances in Image Fusion*, pp. 1–23, 2011.

[17] R. Kabilan, V. Chandran, J. Yogapriya et al., "Short-term power prediction of building integrated photovoltaic (BIPV) system based on machine learning algorithms," *International Journal of Photoenergy*, vol. 2021, 11 pages, 2021.

[18] B. K. Umri, M. Wafa Akhyari, and K. Kusrini, "Detection of COVID-19 in chest X-ray image using CLAHE and convolutional neural network," in *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 14–18, Manado, Indonesia, 2020.

[19] V. Chandran, C. K. Patil, A. M. Manoharan et al., "Wind power forecasting based on time series model using deep machine learning algorithms," *Materials Today: Proceedings*, vol. 47, no. 1, pp. 115–126, 2021.

[20] C. D. N. Kumar and R. Aruna, "Contrast limited adaptive histogram equalization (Clahe) based color contrast and fusion for enhancement of underwater images," *Journal of Engineering (IOSRJEN)*, vol. 5, pp. 63–69, 2018.

[21] B. Bhan and S. Patel, "Efficient medical image enhancement using CLAHE enhancement and wavelet fusion," *International Journal of Computers and Applications*, vol. 167, no. 5, pp. 1–5, 2017.

[22] V. Chandran, K. Patil, A. Karthick, D. Ganeshaperumal, R. Rahim, and A. Ghosh, "State of charge estimation of lithium-ion battery for electric vehicles using machine learning algorithms," *World Electric Vehicle Journal*, vol. 12, no. 1, p. 38, 2021.

[23] F. E. Z. A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," *Egyptian Informatics Journal*, vol. 17, no. 1, pp. 99–124, 2016.

[24] V. P. S. Naidu, "Discrete cosine transform based image fusion techniques," *Journal of Communication, Navigation and Signal Processing*, vol. 1, no. 1, pp. 35–45, 2012.

[25] M. Wang and X. Shang, "A fast image fusion with discrete cosine transform," *IEEE Signal Processing Letters*, vol. 27, no. 8, pp. 990–994, 2020.

[26] N. Paramanandham and K. Rajendiran, "Infrared and visible image fusion using discrete cosine transform and swarm intelligence for surveillance applications," *Infrared Physics & Technology*, vol. 88, pp. 13–22, 2018.

[27] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019.

[28] K. Wang, M. Zheng, H. Wei, G. Qi, and Y. Li, "Multi-modality medical image fusion using convolutional neural network and contrast pyramid," *Sensors*, vol. 20, pp. 1–17, 2020.

[29] K. Koonsanit, S. Thongvigitmanee, and N. P. PT, "Image enhancement on digital x-ray images using n-clahe x-ray," in *2017 10th Biomedical Engineering International Conference (BMEiCON)*, pp. 1–4, Hokkaido, Japan, 2017.

[30] S. S. Chavan and S. N. Talbar, "Multimodality image fusion in frequency domain for radiation therapy," in *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, pp. 174–178, Greater Noida, India, 2014.

[31] L. Junwu, B. Li, and Y. Jiang, "An infrared and visible image fusion algorithm based on LSWT-NSST," *IEEE Access*, vol. 8, pp. 179857–179880, 2020.

[32] B. Chacko, S. L. Agrwal, S. K. Gupta, H. Chahar, S. R. Srivastava, and N. Srivastav, "Performance of image fusion technique using 4×4 block wavelet cosine transformation," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, pp. 618–622, Noida, India, 2017.

[33] J. Zhi, J. Sun, Z. Wang, and W. Ding, "Support vector machine classifier for prediction of the metastasis of colorectal cancer," *International Journal of Molecular Medicine*, vol. 41, no. 3, pp. 1419–1426, 2018.

[34] N. Liu, J. Shen, M. Xu, D. Gan, E. S. Qi, and B. Gao, "Improved cost-sensitive support vector machine classifier for breast cancer diagnosis," *Mathematical Problems in Engineering*, vol. 2018, 13 pages, 2018.

[35] H. Wang, Y. Shi, X. Zhou, Q. Zhou, S. Shao, and A. Bouguettaya, "Web service classification using support vector machine," in *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, vol. 1, pp. 3–6, Arras, France, 2010.

[36] X. Zeng, S. Yuan, Y. Li, and Q. Zou, "Decision tree classification model for popularity forecast of Chinese colleges," *Journal of Applied Mathematics*, vol. 2014, 7 pages, 2014.

[37] T. Yang, J. Song, and L. Li, "A deep learning model integrating SK-TPCNN and random forests for brain tumor segmentation in MRI," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 3, pp. 613–623, 2019.

[38] P. M. Siva Raja and A. V. Rani, "Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 440–453, 2020.

[39] P. A. Charde and S. D. Lokhande, "Classification using K nearest neighbor for brain image retrieval," *International Journal of Scientific and Engineering Research*, vol. 4, no. 8, pp. 760–765, 2013.

[40] R. Chaves, J. Ramirez, J. M. Gorriz et al., "SPECT image classification based on NMSE feature correlation weighting and SVM," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, vol. 1no. 3, pp. 2715–2719, Orlando, FL, USA, 2009.

[41] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of heart disease using K-nearest neighbor and genetic algorithm," *Procedia Technology*, vol. 10, pp. 85–94, 2013.

[42] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *OTM Confederated*

International Conferences On the Move to Meaningful Internet Systems, vol. 2888, pp. 986–996, Springer, Berlin, Heidelberg, 2003.

[43] J. M. Górriz, F. Segovia, J. Ramírez, A. Lassl, and D. Salas-Gonzalez, "GMM based SPECT image classification for the diagnosis of Alzheimer's disease," *Applied Soft Computing*, vol. 11, no. 2, pp. 2313–2325, 2011.

[44] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh, "Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3333–3342, 2014.

[45] G. Amato and F. Falchi, "KNN based image classification relying on local feature similarity," in *Proceedings of the Third International Conference on Similarity Search and Applications*, pp. 101–108, Istanbul, Turkey, 2010.

[46] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.

[47] A. F. Mashat, M. M. Fouad, S. Y. Philip, and T. F. Gharib, "A decision tree classification model for university admission system," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 10, pp. 17–21, 2012.

[48] S. K. Baranwal, K. Jaiswal, K. Vaibhav, A. Kumar, and R. Srikantaswamy, "Performance analysis of brain tumour image classification using CNN and SVM," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 537–542, Coimbatore, India, 2020.

[49] P. Nagaraj, V. Muneeswaran, L. V. Reddy, P. Upendra, and M. V. Reddy, "Programmed multi-classification of brain tumor images using deep neural network," in *2020 4th international conference on intelligent computing and control systems (ICICCS)*, pp. 865–870, Madurai, India, 2020.

[50] A. Biswas and M. S. Islam, "Brain tumor types classification using K-means clustering and ANN approach," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 654–658, DHAKA, Bangladesh, 2021.

[51] H. T. Zaw, N. Maneerat, and K. Y. Win, "Brain tumor detection based on Naïve Bayes classification," in *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pp. 1–4, Luang Prabang, Laos, 2019.

[52] S. Masoudi, S. Mehralivand, S. A. Harmon et al., "Deep learning based staging of bone lesions from computed tomography scans," *IEEE Access*, vol. 9, pp. 87531–87542, 2021.

[53] R. A. Welikala, P. Remagnino, J. H. Lim et al., "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer," *IEEE Access*, vol. 8, pp. 132677–132693, 2020.

[54] M. A. Anupama, V. Sowmya, and K. P. Soman, "Breast cancer classification using capsule network with preprocessed histology images," in *2019 International conference on communication and signal processing (ICCSP)*, pp. 143–147, Chennai, India, 2019.

[55] P. T. Nguyen, "Multiclass breast cancer classification using convolutional neural network," in *2019 International symposium on electrical and electronics engineering (ISEE)*, pp. 130–134, Ho Chi Minh City, Vietnam, 2019.

[56] B. Erkal, S. Basak, A. Çiloglu, and D. D. Sener, "Multiclass classification of brain cancer with machine learning algorithms," in *2020 Medical Technologies Congress (TIPTEKNO)*, Antalya, Turkey, 2020.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 4295985, 5 pages
https://doi.org/10.1155/2022/4295985

Research Article

Functional Magnetic Resonance Imaging Study of Electroacupuncture Stimulating Uterine Acupoints

ChengChao Xu,[1] XiaoHua Yu,[2] Liang Yin,[1] Xiang Li,[2] WanLi Zhang,[2] Fei Li,[3] and TianYu Bai[1]

[1]*Shandong Provincial Third Hospital, Shandong University, Jinan, Shandong 250031, China*
[2]*Shandong University of Traditional Chinese Medicine, Shandong, Jinan 250355, China*
[3]*School of Economics, Beijing Technology and Business University, 100048, China*

Correspondence should be addressed to TianYu Bai; baitianyu19820217@163.com

Received 6 November 2021; Revised 12 December 2021; Accepted 20 December 2021; Published 19 January 2022

Academic Editor: Tao Song

Objective. Based on resting-state functional magnetic resonance imaging (rs-fMRI), to observe the changes of brain function of bilateral uterine points stimulated by electroacupuncture, so as to provide imaging basis for acupuncture in the treatment of gynecological and reproductive diseases. *Methods*. 20 healthy female subjects were selected to stimulate bilateral uterine points (EX-CA1) by electroacupuncture. FMRI data before and after acupuncture were collected. The ReHo values before and after acupuncture were compared by using the analysis method of regional homogeneity (ReHo) of the whole brain, so as to explore the regulatory effect of acupuncture intervention on brain functional activities of healthy subjects. *Results*. Compared with before acupuncture, the ReHo values of the left precuneus lobe, left central posterior gyrus, calcarine, left lingual gyrus, and cerebellum decreased significantly after acupuncture. *Conclusion*. Electroacupuncture at bilateral uterine points can induce functional activities in brain areas such as the precuneus, cerebellum, posterior central gyrus, talform sulcus, and lingual gyrus. The neural activities in these brain areas may be related to reproductive hormone level, emotional changes, somatic sensation, and visual information. It can clarify the neural mechanism of acupuncture at uterine points in the treatment of reproductive and gynecological diseases to a certain extent.

1. Introduction

Functional magnetic resonance imaging (fMRI) technology is a new noninvasive, nonradioactive, and multiangle method to detect local brain functional activities. FMRI technology has made many achievements in the study of brain functional changes and has been gradually applied to the study of the mechanism of acupuncture and moxibustion in recent years [1, 2]. Uterine acupoint (ex-ca1) is one of the odd acupoints outside the meridian. It has the effects of warming the uterus and regulating menstruation, promoting qi and relieving pain, raising Yang, and lifting depression. It is widely used to treat gynecological diseases such as irregular menstruation, dysmenorrhea, infertility, and uterine prolapse [3]. Because it is located in the less abdomen and is outside the uterus, it is also the main acupoint for regulating the uterus [4]. However, at present, there are few studies on the effective mechanism of uterine acupoints in the treatment of gynecological diseases. Therefore, through functional magnetic resonance imaging technology, we preliminarily explore the effect of electroacupuncture stimulating bilateral uterine acupoints on brain function, so as to provide imaging basis for further exploring the mechanism of acupuncture and moxibustion.

2. Materials and Methods

2.1. Observation Object. Through the official account of WeChat, the public health volunteers were recruited openly. Inclusion criteria were as follows: (1) age 20-40 years; (2)

female, all right-handed; (3) no history of mental or nervous system diseases; (4) menstruation is generally normal; (5) no drug-taking history and acupuncture treatment history within the first two weeks of participating in this project; and (6) those who sign the informed consent form. Exclusion criteria were as follows: (1) those with contraindications to MRI scanning, such as foreign bodies such as metal products, cardiac pacemakers, and metal dentures (which cannot be removed), those with claustrophobia, or those who cannot accept MRI scanning for other reasons; (2) pregnant and lactating women; (3) have a history of alcohol or drug abuse; and ④ those with contraindications to acupuncture and moxibustion. This study was approved by the ethics committee of the Third Hospital of Shandong Province (ethics No.: kyll-2021038). All volunteers were informed of the whole experimental process and signed informed consent.

2.2. Electroacupuncture Stimulation Method. Volunteers lie flat on the treatment bed and expose their small abdomen. Huatuo brand disposable acupuncture needle of $0.25 * 40$ mm is selected as the needle. After routine disinfection of acupoints, the needle is directly inserted into the skin for about 25 mm. After needling qi, connect the Huatuo brand electroacupuncture instrument to both uterine points; select continuous wave, frequency 1 Hz; and keep the needle for 20 min. All 20 volunteers were operated by the same acupuncturist. FMRI scans were performed before and after acupuncture.

2.3. MRI Scanning Program and Parameters. The Philips Ingenia 3.0 T MRI scanner of our hospital was used to collect data with 16-channel head neck combined coil, with anti-noise earplugs inside and sponge outside. Before examination, the subjects lay flat on the examination bed and had a full rest to eliminate psychological factors such as fear and anxiety. Keep the head still during the examination, wear eye masks, and guide the subjects to relax, not fall asleep, and not do any thinking activities.

Before acupuncture treatment, subjects with intracranial organic lesions were excluded by routine sequence MRI scanning. The scanning sequence included T1WI, T2WI, FLAIR, and DWI. Then, 3D structure data acquisition and finally rs-fMRI (FE-EPI sequence) scanning were done, scanning parameters:TR = 2000 ms,TE = 30 ms,FA = 90 °,FOV = 220 mm × 220 mm,matrix = 64 × 64 ,number of layers = 24,layer thickness = 3 mm,thickness interval = 1 mm; a total of 180 time points were collected. After acupuncture treatment, 3D structural data and RS fMRI were collected.

2.4. fMRI Image Processing. In order to reduce the impact of data acquisition error on subsequent analysis results, DPABI software based on MATLAB 2018b platform is used in this project to preprocess the collected MRI data. The preprocessing steps include eliminating the first 10 scanning time points, time horizon correction, and head movement correction (excluding subjects with average head movement amplitude > 1 mm or rotation parameter > 1° in the X-, Y-, and Z-axes). The corrected data will be registered on the EPI template for spatial standardization, and the fMRI image will be resampled with the size of 3 mm × 3 mm × 3 mm voxel. The 24 friston head movement parameters, cerebrospinal fluid, white matter, and whole brain mean signals were removed as covariates. A band-pass filter with a filtering range of 0.01-0.1 Hz was used to eliminate the impact of noise caused by subjects' physiological activities such as breathing and heartbeat on the research results.

2.5. Statistical Analysis. After fMRI data preprocessing, DPARSF, SPM8, and other toolkits based on MATLAB platform were used for whole brain regional homogeneity (ReHo) analysis. ReHo analysis can consider local spatial and temporal information at the same time [5]. Kendall concordance coefficient (KCC) can be used to explain the voxel consistency of brain functional regions in different time series. The Kendall harmony coefficient is used to measure the ReHo value. By calculating the time series consistency (KCC value) between each voxel in the brain and its adjacent 26 voxels, the ReHo diagram of the whole brain of each subject can be obtained [6]. Then, the whole brain is averaged; that is, the ReHo value of each voxel is divided by the whole brain ReHo mean to achieve standardization. In order to further reduce noise and improve signal-to-noise ratio, Gaussian smoothing kernel with half height and width of 6 mm is used for spatial smoothing. The ReHo values of the two sets of subjects were statistically analyzed by paired t-test to obtain the different brain regions before and after acupuncture. Gaussian random field theory (GRF) is used for multiple comparison correction.

3. Research Results

A total of 20 female healthy subjects were included in this study, and all completed fMRI scanning before and after acupuncture. Using the method of ReHo analysis to analyze the fMRI data, it was found that the ReHo values of the left precuneus lobe, left central posterior gyrus, calcarine, left lingual gyrus, and cerebellum induced by electroacupuncture at uterine point were significantly lower than those before acupuncture. After GRF correction (voxel $P < 0.05$, cluster $P < 0.05$), there was significant difference (see Table 1 and Figure 1 for details).

4. Discussion

Acupuncture and moxibustion are more and more recognized by the medical community, and the mechanism of acupuncture and moxibustion is also being explored. With the development of science and technology, fMRI is a new imaging method in recent years. It has the unique advantages of no radiation, no bone artifacts, and multiparameter imaging. It has become one of the hotspots of acupuncture research. The principle is mainly to observe the change characteristics of image signal intensity according to the change of blood oxygen level by calculating the ratio of oxygenated hemoglobin to deoxyhemoglobin and the change of relative

Table 1: Different brain regions before and after acupuncture intervention.

Brain region	Hemisphere	Condition	MNI			Cluster size	Peak t
			X	Y	Z		
Precuneus Posterior gyrus	Left	After acupuncture < before acupuncture	-27	-66	48	456	-5.3196
Calcarine Lingual gyrus	Bilateral Left	After acupuncture < before acupuncture	0	-93	0	327	-5.7168
Cerebellum	Bilateral	After acupuncture < before acupuncture	-6	-72	-18	357	-5.4967

MNI: Montreal Neurological Institute, Montreal Institute. The cluster of precuneus also includes functional brain areas such as the central posterior gyrus and superior parietal lobule. The cluster of the talform sulcus also includes functional brain regions such as the lingual gyrus, middle occipital gyrus, and inferior occipital gyrus.

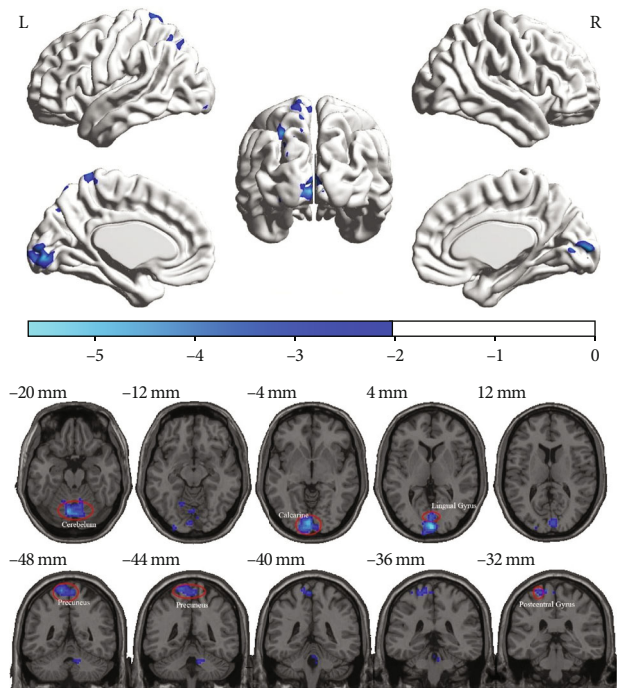


Figure 1: Different brain regions before and after acupuncture intervention.

concentration of deoxyhemoglobin, combined with the paramagnetic characteristics of deoxyhemoglobin [7].

ReHo is a method for measuring similarity or coherence in voxel analysis of the whole brain. This method has been used to explain the relationship between neurovascular coupling and task activation [8], as well as the functional regulation of cognitive changes in subjects in a resting state [8]. However, there are few studies on the changes of ReHo in brain region caused by acupuncture. In this study, we found that acupuncture at uterine point can inhibit the neural activities of the precuneus, cerebellum, talform sulcus, lingual gyrus, and posterior central gyrus. In a previous study, Lin et al. reported that the higher the ReHo value, the stronger the synchronization of local neuronal activities, which is a compensatory response to metabolic abnormalities and blood flow changes related to adverse clinical outcomes [9]. Their results suggest that the reduction of ReHo may have a protective effect. In this study, we found that electroacupuncture at uterine points can reduce ReHo values in brain regions such as the precuneus lobe in healthy subjects. However, whether electroacupuncture at uterine point has neuroprotective effect needs further study.

Precuneus is considered to be one of the important regions related to emotion, cognition, and memory and related to mental disorders. It participates in the distributed network of cortical and subcortical regions and integrates self-generation and external information [10]. The precuneus can produce significant neural changes to inconsistent stimulation information [11]. The nerves distributed at uterine points have the same part as the ganglion segments of female reproductive organs such as the uterus, and studies have proved that acupuncture can cause changes in sex hormones [12]. Therefore, we speculate that the reduction of ReHo in precuneus is related to sex hormone fluctuations and emotional changes. The regulatory effect and clinical efficacy of acupuncture and moxibustion also show that acupuncture and moxibustion can maintain internal balance. Therefore, we speculate that the potential mechanism of acupuncture at uterine points in the treatment of gynecological and reproductive diseases may be related to the neural activity of precuneus, regulate the changes of female emotion, and restore the normal level of hormone.

As we all know, the basic function of the cerebellum is to maintain postural balance and coordinate random movement. Studies have also shown that the cerebellum plays a potential role in many functions such as cognition and emotion [13]. The literature suggests that [14] the cerebellum has a wide impact on the emotion, pain, and other negative emotions of female patients with premenstrual anxiety. Some scholars have found that [15], in patients with premenstrual syndrome, bilateral cerebellar falff decreases, and acupuncture at Sanyinjiao can induce bilateral cerebellar falff to increase. This study also proves that electroacupuncture at bilateral uterine points can affect the ReHo value of the cerebellum. Therefore, we believe that acupuncture at uterine points can regulate the activity of the cerebellum and then affect the changes of women's emotion and cognition.

The posterior central gyrus belongs to the parietal lobe, also known as the somatosensory center, and is part of the pain matrix. It may play a key role in regulating pain perception, including pain localization and pain intensity recognition. Ke et al. [16] reported that ReHo increased due to

adaptive neuronal changes caused by long-term pain stimulation and biological response. Electroacupuncture at bilateral uterine points can reduce the ReHo value of the posterior central gyrus. We speculate that electroacupuncture at bilateral uterine points can affect the pain regulation function of the posterior central gyrus to a certain extent, which may be the effective mechanism of acupuncture at uterine points in the treatment of dysmenorrhea.

The calcarine sulcus and lingual gyrus are located in the visual cortex and play an important role in vision [17]. In addition, the study of Rehbein et al. [17, 18] also proved that, at the neural level, the increase of estradiol level was related to the decreased activation of the right lingual gyrus and left calcarine sulcus. In the present study, visual structures were involved in electroacupuncture at uterus points, suggesting that uterus points may be related to the treatment of visual impairment. Therefore, we speculated that electroacupuncture of bilateral uterine acupoints interferes with the changes of talus sulcus and lingual gyrus, thus improving the level of estradiol, which may be the neuroimaging mechanism of uterine acupoints for the treatment of reproductive diseases.

In conclusion, electroacupuncture at bilateral uterine points can affect the neural activities of brain areas such as the precuneus lobe, cerebellum, posterior central gyrus, talform sulcus, and lingual gyrus, which may be related to the fact that acupuncture at uterine points can regulate the level of reproductive hormone, emotional changes, and somatic sensation, so as to clarify the neural mechanism of acupuncture at Zigong point in the treatment of reproductive and gynecological diseases to a certain extent. However, further experiments are needed to prove whether this is related and consistent with the brain effect of acupuncture at uterine points under pathological conditions. In addition, the sample size of this study is too small, and only 20 volunteers were recruited. Although the number of participants is no less than that in previous similar studies [19–21], we believe that the number of volunteers in fMRI research is more than 20, which may obtain more convincing evidence. These problems will be solved in our future research.

Data Availability

The datasets are available at http://www.oasis-brains.org/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Shandong Provincial TCM High-level Personnel Training Program supported by special Funds, Qilu Health and Health Leading Talents Training Project, and Shandong Traditional Chinese Medicine Science and Technology Development Plan Project (No. 2019-0335).

References

[1] A. Li, X. L. Li, F. Zhang et al., "A functional magnetic resonance imaging study of the neuronal specificity of an acupoint: acupuncture at Rangu (ILI 2) and its sham point," *Internal medicine journal*, vol. 46, no. 8, pp. 973–977, 2016.

[2] T. He, W. Zhu, S. Q. Du et al., "Neural mechanisms of acupuncture as revealed by fMRI studies," *Autonomic Neuroscience*, vol. 190, pp. 1–9, 2015.

[3] X. Y. Shen and C. Z. Liu, *Meridian Yu Acupoints*, China traditional Chinese Medicine Press, Beijing, Fifth edition, 2021.

[4] X. Zhang and W. Li, "Efficacy on endometriosis treated with electroacupuncture," *Zhongguo Zhen Jiu*, vol. 35, no. 4, pp. 323–326, 2015.

[5] M. Lei and J. Zhang, "Brain function state in different phases and its relationship with clinical symptoms of migraine: an fMRI study based on regional homogeneity (ReHo)," *Annals of translational medicine*, vol. 9, no. 11, p. 928, 2021.

[6] Y. Zhu, M. Huang, Y. Zhao et al., "Local functional connectivity of patients with acute and remitting multiple sclerosis: a Kendall's coefficient of concordance- and coherence-regional homogeneity study," *Medicine (Baltimore)*, vol. 99, no. 43, p. e22860, 2020.

[7] M. Waltmann, O. O'Daly, A. Egerton et al., "Multi-echo fMRI, resting-state connectivity, and high psychometric schizotypy," *NeuroImage: Clinical*, vol. 21, p. 101603, 2019.

[8] J. Shi, J. Teng, X. Du, and N. Li, "Multi-modal analysis of resting-state fMRI data in mTBI patients and association with neuropsychological outcomes," *Frontiers in Neurology*, vol. 12, p. 639760, 2021.

[9] W. C. Lin, T. W. Hsu, C. L. Chen, C. H. Lu, H. L. Chen, and Y. F. Cheng, "Resting state-fMRI with ReHo analysis as a non-invasive modality for the prognosis of cirrhotic patients with overt hepatic encephalopathy," *PLoS One*, vol. 10, no. 5, p. e0126834, 2015.

[10] N. Mashal, T. Vishne, and N. Laor, "The role of the precuneus in metaphor comprehension: evidence from an fMRI study in people with schizophrenia and healthy participants," *Frontiers in Human Neuroscience*, vol. 8, p. 818, 2014.

[11] Y. Pang, H. Liu, G. Duan et al., "Altered brain regional homogeneity following electro-acupuncture stimulation at sanyinjiao (SP6) in women with premenstrual syndrome," *Frontiers in Human Neuroscience*, vol. 12, p. 104, 2018.

[12] X. R. Liu, Z. W. Zang, X. L. Li, and J. H. Ma, "Effectiveness of electroacupuncture of Zigong (EX-CA 1), etc. in treatment of perimenopausal syndrome women," *Zhen Ci Yan Jiu*, vol. 41, no. 3, pp. 247–250, 2016.

[13] A. J. Rapkin, S. M. Berman, and E. D. London, "The cerebellum and premenstrual dysphoric disorder," *AIMS neuroscience*, vol. 1, no. 2, pp. 120–141, 2014.

[14] A. J. Rapkin, S. M. Berman, M. A. Mandelkern, D. H. Silverman, M. Morgan, and E. D. London, "Neuroimaging evidence of cerebellar involvement in premenstrual dysphoric disorder," *Biological Psychiatry*, vol. 69, no. 4, pp. 374–380, 2011.

[15] G. Duan, Y. Chen, Y. Pang et al., "Altered fractional amplitude of low-frequency fluctuation in women with premenstrual syndrome via acupuncture at Sanyinjiao (SP6)," *Annals of General Psychiatry*, vol. 20, no. 1, p. 29, 2021.

[16] J. Ke, R. Qi, C. Liu et al., "Abnormal regional homogeneity in patients with irritable bowel syndrome: a resting-state

functional MRI study," *Neurogastroenterology and Motility*, vol. 27, no. 12, pp. 1796–1803, 2015.

[17] E. Rehbein, L. Kogler, J. Hornung et al., "Estradiol administration modulates neural emotion regulation," *Psychoneuroendocrinology*, vol. 134, p. 105425, 2021.

[18] E. Rehbein, J. Hornung, I. S. Poromaa, and B. Derntl, "Shaping of the female human brain by sex hormones: a review," *Neuroendocrinology*, vol. 111, no. 3, pp. 183–206, 2021.

[19] L. Liu, S. Chen, D. Zeng, H. Li, C. Shi, and L. Zhang, "Cerebral activation effects of acupuncture at Yanglinquan (GB34) point acquired using resting-state fMRI," *Computerized Medical Imaging and Graphics*, vol. 67, pp. 55–58, 2018.

[20] B. Zhu, Y. Wang, G. Zhang et al., "Acupuncture at KI3 in healthy volunteers induces specific cortical functional activity: an fMRI study," *BMC Complementary and Alternative Medicine*, vol. 15, p. 361, 2015.

[21] Y. Shi, S. Zhang, Q. Li et al., "A study of the brain functional network of Deqi via acupuncturing stimulation at BL40 by rs-fMRI," *Complementary Therapies in Medicine*, vol. 25, pp. 71–77, 2016.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 9773917, 9 pages
https://doi.org/10.1155/2021/9773917

Research Article

RDFNet: A Fast Caries Detection Method Incorporating Transformer Mechanism

Hao Jiang[1], **Peiliang Zhang**[1], **Chao Che**[1] and **Bo Jin**[2]

[1]*Key Laboratory of Advanced Design and Intelligent Computing (Dalian University), Ministry of Education, Dalian 116622, China*
[2]*School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Chao Che; chechao@gmail.com

Received 20 August 2021; Accepted 25 October 2021; Published 10 November 2021

Academic Editor: Pan Zheng

Dental caries is a prevalent disease of the human oral cavity. Given the lack of research on digital images for caries detection, we construct a caries detection dataset based on the caries images annotated by professional dentists and propose RDFNet, a fast caries detection method for the requirement of detecting caries on portable devices. The method incorporates the transformer mechanism in the backbone network for feature extraction, which improves the accuracy of caries detection and uses the FReLU activation function for activating visual-spatial information to improve the speed of caries detection. The experimental results on the image dataset constructed in this study show that the accuracy and speed of the method for caries detection are improved compared with the existing methods, achieving a good balance in accuracy and speed of caries detection, which can be applied to smart portable devices to facilitate human dental health management.

1. Introduction

Dental caries is a prevalent disease of the human oral cavity that has a great impact on human quality of life. Data from the National Health and Nutrition Examination Survey, 2011–2012, showed that among children aged 2–8 years, 37% had dental caries in their primary teeth. Among adolescents aged 12–19 years, the prevalence of dental caries in permanent teeth was 58%. Approximately 90% of adults aged ≥20 years had dental caries [1]. Therefore, the detection of dental caries can provide reliable clinical reference to doctors and effectively avoid the onset or the further severity of dental caries, which is of great significance to improving the quality of human life.

With the advances of deep-learning methods in computer vision, many such methods have been applied to caries detection to improve the accuracy of detection and relieve the dentists' workload. For example, Suryani et al. [2] used mask R-CNN to detect objects in dental panoramic X-ray images, which saved time and improved the quality of dentists' diagnoses by automatically detecting panoramic X-ray images. Majanga et al. [3] proposed a deep learning-based technique for dental caries detection named blob detection. The proposed technique automatically detects hidden and inaccessible dental caries lesions. The process of detection and classifying dental caries achieved the results of 97% and 96% for the precision and recall values, respectively. The above methods for caries detection are mainly based on X-ray images. However, the X-ray images must be captured and acquired by professional equipment and specialized technicians, which is expensive and cumbersome. It has become increasingly easy to acquire digital images due to the popularity of portable devices and the development of biological storage technology [4]. Caries detection becomes quicker and more convenient based on these digital images, making it possible to conduct detection anytime and anywhere. Therefore, the use of digital images for caries detection has become a new demand in human dental health management.

Saini et al. [5] used digital images of dental caries for early classification and prevention of dental caries. They used four convolutional neural networks including ResNet50, all of which achieved good classification accuracy. However, their study only classified the images of caries

without detecting the type of caries and lesion area. To our knowledge, no study has been conducted on the detection of dental caries based on digital images, mainly because of the lack of large-scale digital image datasets of annotated dental caries. Caries detection is an object detection task, and the popular object detection algorithms are all supervised deep-learning methods, which require a large amount of annotated data to train the model. The caries images must be annotated by professional dentists, which is troublesome and laborious. Our research group has accumulated a large number of caries images annotated by professional doctors during the preliminary research process. To address the problem of the lack of a digital image dataset of dental caries, we performed data cleaning and image enhancement operations based on the aboveannotated images of dental caries. We deleted the low-quality images in the dataset and further enhanced and expanded the images to build the annotated dataset for dental caries detection. Meanwhile, we here propose use of the RDFNet model (rapid network with deep features for dental caries detection) based on the single-stage deep-learning detection method YOLOv5s to detect dental caries in digital images. Since the caries parts of the images in the dataset are dark and their features are not obvious, RDFNet incorporates the transformer mechanism in the backbone to better extract the complex features of tooth decay and improve the accuracy of detection. To run the caries detection algorithm on portable devices and improve the inference speed, the RDFNet model uses the FReLU activation function to activate the complex visual-spatial information of the images to meet the demand of computing speed.

Overall, the main contributions of this paper are the following.

(1) A caries image dataset is constructed, and all images are annotated by professional dentists.

(2) The transformer mechanism based on the original YOLOv5 backbone network is incorporated to better extract the complex features of dental caries.

(3) The FReLU activation function is adopted to activate the complex visual-spatial information of the images, which improves the inference speed of the model.

2. Related Works

2.1. Medical Image Object Detection. Object detection is a computer technology related to computer vision and image processing for detecting specific classes of semantic objects (e.g., people, buildings, or cars) in images and videos, which has promising applications in areas such as video security and autonomous driving, among others [6]. Medical images are mainly used to help doctors to make judgments about medical information. Their production has grown exponentially due to the increase of image acquisition devices and advances in camera technology. In recent years, with the continuous development and progress in medical technology, modern hospitals have used medical images to predict the intensity of patients' diseases. Medical object detection based on deep learning has gradually become a current research hotspot. In 2020, Tavakoli et al. [7] used a deep-learning method to detect microaneurysms in retinal images. The experimental results showed that the accuracy of microaneurysm detection was approximately 90%, and the performance of the method using top-hat preprocessing was greater than 80%. During the COVID-19 epidemic, Loey et al. [8] used ResNet50 and YOLOv2 for medical mask-wearing detection. Their proposed model consists of two parts, a deep transfer-learning model based on ResNet50 for feature extraction and a YOLOv2 framework for medical mask detection, achieving a best detection accuracy of 81%.

2.2. X-Ray Image-Based Caries Detection. X-ray images are widely used in stomatological research because they show the full details of the teeth and gums. At present, most of the caries detection methods are carried out using X-ray images. Relevant studies have used the method of caries segmentation to detect caries in X-ray images and achieved good accuracy. Rad et al. [9] proposed a caries segmentation and recognition method that uses integral projection technology to extract local feature mapping information of teeth. Lakshmi et al. [10] segmented dental X-ray images using a deep convolutional neural network (CNN) to predict caries in dental images. Both methods demonstrated high accuracy on caries X-ray image datasets. Some studies, such as that of Suryani et al. [2], have also directly used deep-learning methods for caries detection and identification, which saved time and process steps and improved the simplicity of the method.

2.3. Transformer. The transformer network is a well-known and efficient deep-learning model proposed by Google in 2017 [11]. The basic version of transformer is based on the attention mechanism and consists of a decoder and encoder. This structure was originally proposed in the sequence-to-sequence model of machine translation [12]. Currently, transformer has been widely used in natural language processing [11], computer vision [13–15], medical disease detection [16], etc. The transformer structure mainly utilizes the self-attention mechanism [11] to extract intrinsic features and has shown great potential in various fields.

Transformer has received significant attention in the field of object detection due to its advantageous capability in extraction of intrinsic features. There are two main categories of transformer-based object detection, a transformer-based set prediction for detection and a transformer-based backbone for detection. DEtection TRansformer (DETR) [13] is a typical example of the first category. A clear disadvantage of this category is the high training cost and poor detection accuracy for small objects. The second type of method takes the transformer structure as a part of the backbone network of common detection methods and has achieved good performance [17].

3. Materials and Methods

3.1. Dental Caries Annotation Dataset Construction. During preliminary research conducted by our research group, a large number of caries images annotated by professional

Table 1: Statistics of caries instances in the dataset constructed in the present work.

Categories	Total number	Percentage
Mild caries	5,570	38.1%
Moderate caries	4,370	29.9%
Severe dental caries	4,670	32.0%

doctors have been accumulated, but two types of problems exist in the original images. First, there are some low-quality images. The image acquisition equipment contains different models of cameras, smartphones, etc., and the acquired images exhibit large variability in pixels, size, and resolution. Some of the images contain not only the oral part but also the actual living environment of users. Second, some noise exists in the acquired images, and the lighting and contrast in the images are significantly different.

To address the problem of low-quality images in the original data, we conducted data cleaning on the original images, manually selecting and removing unreasonable labels and low-quality pictures one by one according to the annotations in the dataset and only keeping the high-quality images with clear and correct annotations. Then, 7.6% of the original caries images were removed, and a total of 4,277 high-quality caries images were obtained. To address the noise problem in the data, we performed image enhancement operations, e.g., adjusting brightness and contrast, random flipping, and randomly adding noise to the images, to improve the robustness of the model. Specifically, the images in the dataset have been flipped according to 50% probability, and in the process of image flipping, horizontal and vertical flipping were performed separately according to 50% probability. For each image, the brightness, contrast, and saturation were adjusted separately with 33.3% probability. We added noise to the images with 30% probability, and in the process of adding noise, Gaussian, pepper, and salt noise were separately added with 33.3% probability. At the end of all operations, a total of 8,554 images were obtained, including the processed and original images, which realizes the enhancement and expansion of the dataset.

After preprocessing the original data, the labeled dental caries instances were divided into three different categories, i.e., mild, moderate, and severe dental caries according to the clinical manifestations of dental caries at different stages, clinical diagnoses of patients in the hospital, and practical experience of the dentists. The statistics of caries instances for each category are shown in Table 1.

3.2. RDFNet. Given the lack of research on digital images for caries detection, the RDFNet model, based on YOLOv5s, which is inspired by the YOLO [18] series algorithm, taking the characteristics of real-time and accuracy into account, is proposed in this study.

The structure of RDFNet is shown in Figure 1. The model contains backbone, neck, and prediction modules.

3.2.1. Overview. The backbone module is used to perform the extraction of caries features. Among them, focus [19] is a slicing operation that can replace convolution and reduce the loss of feature information caused by convolution. C3Modified is a convolution module activated by the FReLU function that is used to extract complex visual-spatial information of caries images. SPP [20] is a spatial feature pyramid pooling structure that can expand the perceptual field, fuse the local and global features, and enrich the information of the feature map. To better extract the depth features of dental caries, RDFNet adds an improved transformer-encoder module after the original SPP structure, which is used to increase the network's ability to extract dental caries features.

The neck module is used to fuse feature maps of different sizes as well as extract high-level semantic features. The module uses the structure of feature pyramid network (FPN [21]) and path aggregation network (PAN [22]), where FPN is performed top-down, the predicted feature maps are obtained by passing fusion of information through upsampling, and PAN is performed bottom-up to obtain the feature pyramids. Feature fusion is performed using top-down as well as bottom-up approaches, thus reducing information loss. To better extract the high-level semantic features of caries images, in this study, the improved C3Modified convolutional module is adopted in the neck module.

The prediction module uses the high-level semantic features generated by the neck module to classify and regress the class and location of objects. It consists of three detection heads to detect large, medium, and small objects, which is a good way to overcome the shortcomings of the single-stage detection method with low detection accuracy for small target objects.

3.2.2. Feature Extraction Module Integrated with Transformer Structure. Since transformer has strong extraction capability for complex features, the feature extraction module shown in Figure 2 is proposed based on transformer's structure.

To extract features better, we perform stacking of three transformer encoders. In our experiments, we remove the original normalization layer from the transformer encoder in order to simplify the model.

We fed the feature map into this structure for the extraction of deep features. The attention values of the different heads were calculated separately and then concatenated. The calculation formulas for calculating the value of multi-head attention are

$$\begin{aligned}\text{MultiHeadAttention}(Q, K, V) &= \text{Concat}(\text{head}_1, \cdots, \text{head}_4),\\ \text{head}_i &= \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right).\end{aligned} \tag{1}$$

The formula for calculating the value of attention is

$$\text{Attention}(Q, K, V) = \text{soft max}\left(\frac{Q \times K^T}{\sqrt{d_k}} \times V\right), \tag{2}$$

where d_k denotes the input dimension, and Q, K, and V denote the query, key, and value matrices, respectively. Q,

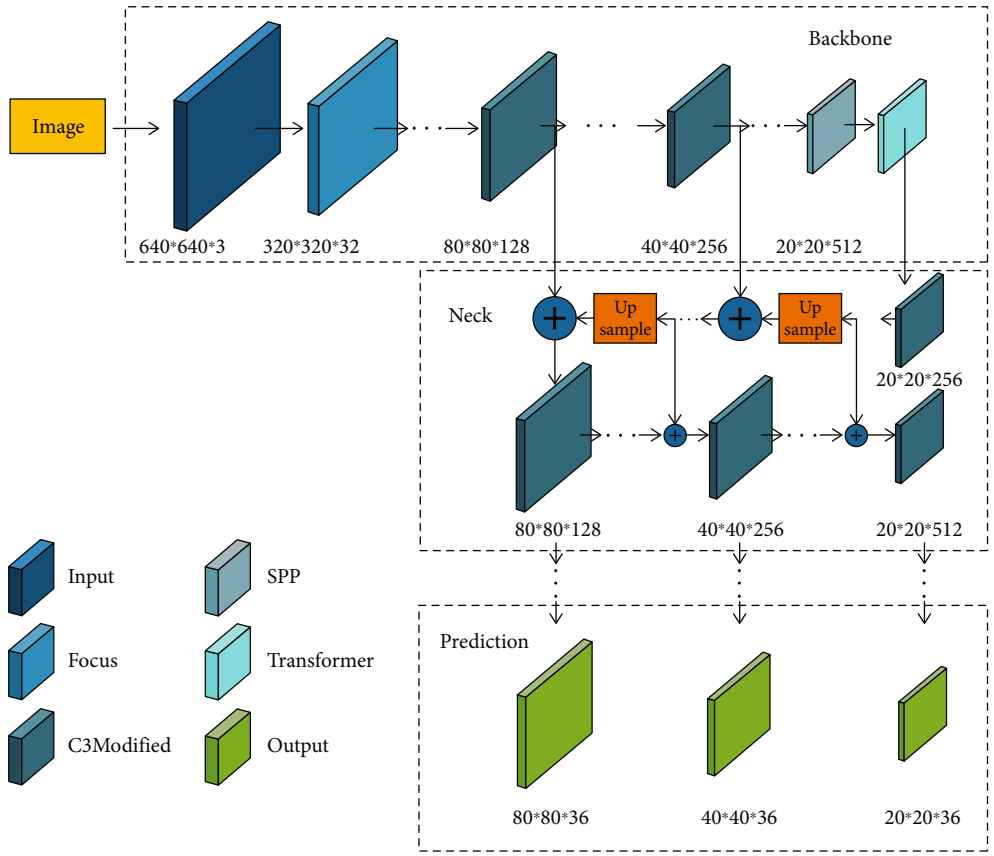


Figure 1: RDFNet structure.

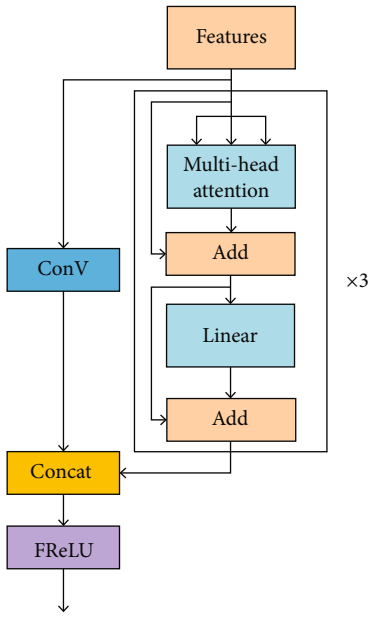


Figure 2: Feature extraction module integrated with transformer structure.

K, and V in the multihead self-attention mechanism in this modular structure have the same values, which are the values of the input feature maps.

3.2.3. Improved Convolutional Module for Extracting Complex Visual-Spatial Information. Ma et al. [23] identified spatially insensitiveness in activations as the main obstacle impeding visual tasks from achieving significant improvements in vision tasks and proposed a new visual activation function, i.e., FReLU, based on that. Inspired by the ability of FReLU to capture complex visual-spatial information, in the present work, we used the FReLU activation function to improve the original convolution module. The structure of the improved convolution module is shown in Figure 3.

We adopted max $(*)$ as the nonlinear function and spatial context for each pixel as the condition part. A parametric pooling window was used to create spatial dependency.

$$f\left(x_{c,i,j}\right) = \max\left(x_{c,i,j}, T\left(x_c, i, j\right)\right),$$
$$T\left(x_{c,i,j}\right) = x^{\omega}_{c,i,j} \bullet p^{\omega}_{c}. \tag{3}$$

Here, $x_{c,i,j}$ is the input pixel of the nonlinear activation $f(\cdot)$ on the c th channel at the two-dimensional spatial position (i, j). The function $T(\cdot)$ denotes the funnel condition, $x^{\omega}_{c,i,j}$ a $k_h \times k_w$ parametric pooling window centered on $x_{c,i,j}$, p^{ω}_{c} the coefficient on this window that is shared in the same channel, and $(\cdot)$ dot multiplication.

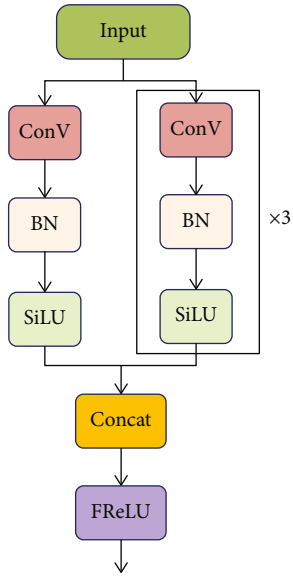


Figure 3: Improved convolution module structure.

4. Results and Analysis

4.1. Experimental Settings. The dataset of digital images constructed in the present work was used in the experiments, with 8,554 caries images divided into training, verification, and test sets according to the ratio of 7 : 2 : 1. The numbers of images in the training, verification, and test sets are 5,987, 1,711, and 856, respectively. To improve the detection accuracy and model robustness, we combined all caries instances in the dataset into one category and detected the caries of this category.

The model was trained using the Pytorch deep-learning framework on an NVIDIA GeForce RTX 3090 graphics card with 24 GB of RAM. The model training parameter settings are shown in Table 2.

4.2. Evaluation Criteria. We used three measurements to evaluate the models, namely, accuracy P, recall R, and average precision mAP@0.5. The calculation formulas are defined as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{4}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{5}$$

$$\mathrm{mAP@0.5} = \frac{\sum_1^n \int_0^1 p(r)dr}{n} (\mathrm{threshold} = 0.5). \tag{6}$$

In formulas (4) and (5) TP, FP, and FN represent true positive, false positive, and false negative, respectively.

In formula (6), n denotes the number of classes, $p(r)$ is the curve between recall and accuracy, and threshold is the Intersection over Union (IoU) threshold of the ground-truth and predicted boxes.

We used frames per second (FPS), i.e., the number of pictures that can be processed per second, as the evaluation criteria of the model detection speed. The calculation formula is defined as follows:

Table 2: Training parameter settings for proposed method.

Parameter	Value	Parameter	Value
Input size	640	Epoch	80
Initial learning rate	0.001	Decay	0.0005
Momentum	0.937	Batch size	32
Optimizer	Steepest gradient descent	Multiscale training	False

$$\mathrm{FPS} = \frac{\mathrm{frameNum}}{\mathrm{elapsedTime}}, \tag{7}$$

where frameNum denotes the total number of pictures that need reasoning, and elapsedTime denotes the total elapsed time of the reasoning process (seconds).

4.3. Comparison with Existing Methods. Six deep learning models, i.e., YOLOv5s, YOLOv5m, YOLOv5l, YOLOv3-tiny [24], faster R-CNN [25], and mask R-CNN [26], were used to facilitate separate comparisons with RDFNet, and the experimental results are shown in Table 3. YOLOv5s, YOLOv5m, and YOLOv5l are different improved versions of YOLOv5, and the number of modules and parameters of all three is in the form of incremental increase. YOLOv3-tiny removes some feature layers from YOLOv3 [27] and keeps only two independent prediction branches. Compared with YOLOv3, YOLOv3-tiny has faster computing speed. Faster R-CNN integrates feature extraction, proposal extraction, bounding box regression, and classification in a single network, which has improved the comprehensive performance of the model, especially the detection speed. Mask R-CNN is improved from faster R-CNN by adding a branch of prediction segmentation mask.

As can be seen from the table, the results of the RDFNet method in accuracy and average precision are significantly better than those of the region proposal-based methods, i.e., faster R-CNN and mask R-CNN, but the experimental recall results are the opposite. This is because, at the early stage of dental caries, the difference between lesion and background images is small, and the region proposal-based methods divide the background, etc. into recognition targets when proposing regions, which leads to the reduction of recognition accuracy of the model and to the recall rate deviating from the normal range. Region proposal-based methods take a significant amount of time in proposing regions and thus are slightly inferior to the single-stage object detection methods in terms of recognition speed. However, faster R-CNN effectively improves the speed of proposing regions by using region proposal networks instead of the previous selective search method, resulting in a large improvement in the performance of the model, especially in detection speed, and thus the best detection speed was achieved in the experiments.

Table 3: Comparison between RDFNet and other object detection methods. The italic value indicates the top performance.

Methods	*P*	mAP@0.5	*R*	FPS
YOLOv5s	0.597	0.552	0.578	20.12
YOLOv5m	0.608	0.556	0.585	18.94
YOLOv5l	0.606	0.563	0.582	17.98
YOLOv3-tiny	0.589	0.5	0.566	22.74
Faster R-CNN	0.421	0.421	*0.825*	*23.14*
Mask R-CNN	0.42	0.42	0.824	17.89
RDFNet	*0.623*	*0.569*	0.579	20.24

In the experiments described herein, YOLOv5s, YOLOv5m, and YOLOv5l did not differ significantly in accuracy, average precision, and recall, but YOLOv5s showed the best results in terms of detection speed. This is because YOLOv5s has a smaller number of parameters compared to YOLOv5m and YOLOv5l; so, the model has a better detection speed overall. Meanwhile, there was no significant difference in other evaluation criteria, indicating that the feature extraction of the model tends to saturate. Because caries in all images are small, this results in limited features extracted by the model, which affects the performance of the method. This situation is not related to the improvement of model complexity and the number of parameters. The detection speed of the RDFNet method is maintained at the same level as in YOLOv5s. This is because the transformer mechanism effectively captures the deeper features in the caries images, while the FReLU activation function accelerates the speed of detection. From the above experimental results, it can be seen that the RDFNet method can effectively improve the accuracy of caries detection while maintaining a fast detection speed, achieving a good balance between accuracy and speed.

4.4. Ablation Study. To verify the effectiveness of each part of the model, a series of ablation experiments were designed and conducted; the experimental results are shown in Table 4.

4.4.1. Transformer Mechanism. It can be seen that in the ablation experiments, the accuracy of RDFNet without incorporating the FReLU function was improved by 1.6% compared with RDFNet without incorporating the FReLU function and transformer mechanism. This is because the transformer mechanism extracts deeper features in the image, and these features effectively improve the performance of the model. The accuracy and average precision of RDFNet are improved by 5% and 1.7%, respectively, compared with RDFNet without incorporating the transformer mechanism. The comparison of the results also proves that the transformer mechanism can effectively enhance the model's ability to extract features and improve accuracy.

4.4.2. FReLU Function. RDFNet without incorporating transformer mechanism improved the recall rate by 1.6%, and the detection speed was also improved compared with RDFNet without incorporating the transformer mechanism and FReLU function. This is because the FReLU activation function accelerates the convergence speed of the model and improves its prediction efficiency, while the FReLU activation function can also affect the feature extraction of the model and reduce the missed detection rate. Comparison of the experimental results of RDFNet with and without incorporating the FReLU function show that the RDFNet method alone outperforms the former in all evaluation criteria, which also further verifies the accuracy of the above conclusions.

4.5. Comparison of Test Results. To test the caries detection effect of RDFNet, six annotated images were randomly selected in the caries test dataset and compared with the prediction results of the model, arranged according to the real annotated caries images at the top and the prediction results of RDFNet at the bottom. The arrangement is shown in Figure 4.

From the overall view of the six images, the detection result of the six images is satisfying. In Figure 4(a), the lesion area of caries is accurately located. For caries with no obvious characteristics in Figure 4(c) and caries with a small lesion area in Figure 4(e), both instances of caries can still be accurately detected, indicating that RDFNet can effectively detect the caries position in the images. Figure 4(f) is the image after horizontal flipping; brightness, contrast, and saturation adjustments; and the addition of noise based on the original image. A good detection result is achieved, indicating that RDFNet has good robustness and can cope with more complex application scenarios.

However, errors and omissions are also evident in the test results. In Figures 4(b) and 4(d), some caries in of the image are not detected correctly because they are not particularly obvious compared with other caries, and there is insufficient illumination near caries, which leads to the omission. In Figure 4(e), there is insufficient illumination inside the oral cavity and the inner part of the cavity is dark, resulting in incorrect detection of dental caries by RDFNet.

5. Discussion

The objectiveness of this study is to construct a dataset of digital images of dental caries annotated by professional doctors and propose an end-to-end detection method RDFNet for the detection of dental caries.

Therefore, the main contributions of this paper are as follows: a caries dataset annotated by professional dentists was constructed. The RDFNet method is proposed to achieve fast and accurate detection of caries images. In the analysis in the previous section, we can find that the transformer mechanism effectively extracts the feature of dental caries, which is an important reason for RDFNet to prevail. And the FReLU activation function is added to the convolutional module to improve the convergence speed of the model, which makes up for the decrease of the model running speed after the transformer mechanism is added.

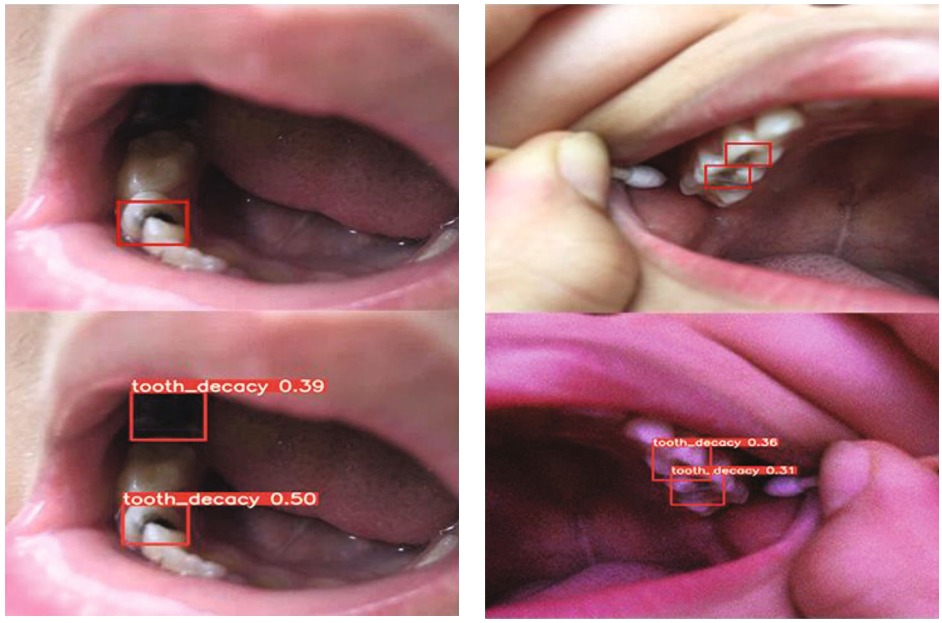
The limitation of this study includes that the method does not work well when the internal illumination of oral

Table 4: Results of RDFNet ablation experiments. The italic value indicates the top performance, and "w/o" indicates "without."

Methods	*P*	mAP@0.5	*R*	FPS
RDFNet w/o transformer and FReLU	0.597	0.552	0.578	20.12
RDFNet w/o FReLU	0.613	0.554	0.578	20.19
RDFNet w/o transformer	0.573	0.552	*0.604*	*20.56*
RDFNet	*0.623*	*0.569*	0.579	20.24

(a) (b) (c) (d)

Figure 4: Continued.



(e) (f)

Figure 4: Comparison of real annotated caries images and predicted results using the proposed method.

image is insufficient, and the detection accuracy and speed are improved compared with the original method, but the detection speed is not the fastest among all comparison methods. Hence, future work will concentrate on how to further improve the running speed of the model and improve the performance of the model on images with insufficient illumination.

Since the proposed method has only been validated in an experimental setting, we plan to deploy the model to various devices, including personal computers, smartphones, and embedded devices, in the future. We expect that the method will be applied in subsequent industrial applications and become fundamental for oral care devices, such as smart toothbrushes with caries detection, to promote human dental health.

6. Conclusions

Aimed at the problem of a lack of research on caries detection methods based on digital images, we construct a caries image dataset annotated by professional dentists and propose a caries detection method, i.e., RDFNet, according to the requirements of performing caries detection on portable devices. Based on the original YOLOv5s backbone network, this method integrates the transformer mechanism to extract features, improving the detection accuracy. The FReLU activation function is used to activate the complex visual-spatial information of the image, which accelerates the reasoning speed of the model. RDFNet method can effectively improve the accuracy of caries detection while maintaining a fast detection speed at the same time. In the future, the RDFNet method will be applied to various devices for dental caries detection and promote human dental health.

Data Availability

Considering the privacy of patients, we cannot open access to our clinical data.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 62076045) and the Guidance Program of Liaoning Natural Science Foundation (No. 2019-ZD-0569).

References

[1] C. Heng, "Tooth decay is the most prevalent disease," *Federal Practitioner*, vol. 33, no. 10, pp. 31–33, 2016.

[2] D. Suryani, M. N. Shoumi, and R. Wakhidah, "Object detection on dental x-ray images using deep learning method," *IOP Conference Series: Materials Science and Engineering*, vol. 1073, no. 1, article 012058, 2021.

[3] V. Majanga and S. Viriri, "Automatic blob detection for dental caries," *Applied Sciences*, vol. 11, no. 19, p. 9232, 2021.

[4] B. Cao, X. Zhang, J. Wu, B. Wang, Q. Zhang, and X. Wei, "Minimum free energy coding for DNA storage," *IEEE Transactions on Nanobioscience*, vol. 20, no. 2, pp. 212–222, 2021.

[5] D. Saini, R. Jain, and A. Thakur, "Dental caries early detection using convolutional neural network for tele dentistry," in *2021 7th International Conference on Advanced Computing and Communication Systems*, pp. 958–963, Coimbatore, India, 2021.

[6] 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.

[7] T. Meysam, J. Sina, and N. Mahdieh, "Automated detection of microaneurysms in color fundus images using deep learning with different preprocessing approaches," in *Medical Imaging 2020: Imaging informatics for healthcare, Research, and Applications*, pp. 113–180E, Houston, United States, 2020.

[8] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: a novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, article 102600, 2021.

[9] A. E. Rad, M. S. M. Rahim, H. Kolivand, and A. Norouzi, "Automatic computer-aided caries detection from dental x-ray images using intelligent level set," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28843–28862, 2018.

[10] M. M. Lakshmi and P. Chitra, "Tooth decay prediction and classification from X-ray images using deep CNN," in *2020 International Conference on Communication and Signal Processing*, pp. 1349–1355, Chennai, India, 2020.

[11] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[12] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, vol. 27, pp. 3104–3112, 2014.

[13] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020*, pp. 213–229, Cham, 2020.

[14] N. Parmar, A. Vaswani, J. Uszkoreit et al., "Image transformer," in *the 35th International Conference on Machine Learning*, pp. 4055–4064, Stockholm, Sweden, 2018.

[15] K. Lin, L. Wang, and Z. Liu, "End-to-end human pose and mesh reconstruction with transformers," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1954–1963, 2021.

[16] C. Che, P. Zhang, M. Zhu, Y. Qu, and B. Jin, "Constrained transformer network for ECG signal processing and arrhythmia classification," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 184, 2021.

[17] X. Pan, Z. Xia, S. Song, L. E. Li, and G. Huang, "3d object detection with pointformer," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7463–7472, 2021.

[18] 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*, pp. 779–788, Las Vegas, United States, 2016.

[19] S. Zhang, Y. Wu, C. Men, and X. Li, "Tiny YOLO optimization oriented bus passenger object detection," *Chinese Journal of Electronics*, vol. 29, no. 1, pp. 132–138, 2020.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[21] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, HI, USA, 2017.

[22] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 8759–8768, Salt Lake City, United States, 2018.

[23] N. Ma, X. Zhang, and J. Sun, "Funnel Activation for Visual Recognition," in *Computer Vision – ECCV 2020*, pp. 351–368, Glasgow, United States, 2020.

[24] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-tiny: object detection and recognition using one stage improved model," in *2020 6th International Conference on Advanced Computing and Communication Systems*, pp. 687–694, Coimbatore, India, 2020.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[26] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision*, pp. 2961–2969, Venice, Italy, 2017.

[27] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 9025470, 28 pages
https://doi.org/10.1155/2021/9025470

Review Article

Deep Learning in Cancer Diagnosis and Prognosis Prediction: A Minireview on Challenges, Recent Trends, and Future Directions

Ahsan Bin Tufail,[1,2] Yong-Kui Ma,[1] Mohammed K. A. Kaabar,[3,4] Francisco Martínez,[5] A. R. Junejo,[6] Inam Ullah,[7] and Rahim Khan[1]

[1]*School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China*
[2]*Department of Electrical and Computer Engineering, COMSATS University Islamabad, Sahiwal Campus, Sahiwal 57000, Pakistan*
[3]*Gofa Camp, Near Gofa Industrial College and German Adebabay, Nifas Silk-Lafto, 26649 Addis Ababa, Ethiopia*
[4]*Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia*
[5]*Department of Applied Mathematics and Statistics, Technological University of Cartagena, Cartagena 30203, Spain*
[6]*School of Control Science and Control Engineering, Harbin Institute of Technology, Harbin 150001, China*
[7]*College of Internet of Things (IoT) Engineering, Hohai University (HHU), Changzhou Campus, 213022, China*

Correspondence should be addressed to Yong-Kui Ma; yk_ma@hit.edu.cn,
Mohammed K. A. Kaabar; mohammed.kaabar@wsu.edu, and Inam Ullah; inam.fragrance@gmail.com

Received 20 August 2021; Revised 30 September 2021; Accepted 5 October 2021; Published 31 October 2021

Academic Editor: Iman Yi Liao

Deep learning (DL) is a branch of machine learning and artificial intelligence that has been applied to many areas in different domains such as health care and drug design. Cancer prognosis estimates the ultimate fate of a cancer subject and provides survival estimation of the subjects. An accurate and timely diagnostic and prognostic decision will greatly benefit cancer subjects. DL has emerged as a technology of choice due to the availability of high computational resources. The main components in a standard computer-aided design (CAD) system are preprocessing, feature recognition, extraction and selection, categorization, and performance assessment. Reduction of costs associated with sequencing systems offers a myriad of opportunities for building precise models for cancer diagnosis and prognosis prediction. In this survey, we provided a summary of current works where DL has helped to determine the best models for the cancer diagnosis and prognosis prediction tasks. DL is a generic model requiring minimal data manipulations and achieves better results while working with enormous volumes of data. Aims are to scrutinize the influence of DL systems using histopathology images, present a summary of state-of-the-art DL methods, and give directions to future researchers to refine the existing methods.

1. Introduction

Cancer is defined as abnormal cell growth that arises from any body organ. In essence, further growth of the cells in these organs is saturated. These silent and saturated cells are increased at a rapid rate till either their removal through a physical procedure such as surgery, medication, use of hormonal therapy, or radiation therapy or their disappearance on their own naturally. The natural disappearance of cancer cells can happen in cancers related to kidney or melanomas. These cells can be screened using tools such as colonoscopy or pap smear examination or using mammograms. There are more than 150 different kinds of cancer, and there is a lack of strategies to cure them in their early stages. Cancer stem cells are an effective way to form stromal cells thus paving a way for the cure of cancers. Apart from stem cells, WNT16B protein also increases resistance against cancer along with chemotherapy. Therapies such as laser therapy and cryotherapy are some of the most vibrant approaches to treat cancer. Some of the most prevalent types of cancers worldwide include lip, oral cavity, breast and cervical, and thyroid cancers. On the other hand, rare cancers such as osteosarcoma, Ewing's sarcoma, male breast cancer, gastrointestinal stromal tumors, chondrosarcoma, mesothelioma,

adrenocortical carcinoma, cholangiocarcinoma, kidney chromophobe carcinoma, pheochromocytoma and paraganglioma, sarcoma, and ependymoma made up more than 20% of cancer cases and are rare types of cancers [1–4].

Cancer is a disease of genes. The process of replication, mitosis, and bombardment by oxygen cells bring continuous changes in normal and cancer cells. This process begins at the birth of a cancer cell and goes on till death. During this process, cancer cell gains mass using stromal support cells, immune cells, and endothelial cells. These cells become a part of cancer mass due to factors like stress ligands and antigens. Other emblems of cancer-based cellular stress are proteotoxicity, metabolic changes, and displaced acids of nucleotides. Another pattern of genes that drive them is chromosomes. They are drivers of a cell's nucleus. The human body has around 20,000 genes in somatic cells, and their study known as cytogenetics has seen large strides of progress over the past several decades where it is now possible to build a 3D model of chromosomes [5–7].

Sugar is an important constituent of tumor cells that fuels the rapid growth of these cells. They are an important part of the diet of cancer cells, and their growth ensures the formation of new clones. Bacteria and microbial cells colonize the human body. Microbial cells are estimated to be as abundant as human cells, yet their genome is roughly 100 times the human genome, providing significantly more genetic diversity. Helicobacter pylori, Chlamydia trachomatis, Salmonella enterica serovar typhi, Fusobacterium nucleatum, enterotoxigenic Bacteroides fragilis, Koribacteraceae, etc., are some of the most prominent bacteria that are associated with cancer. Apoptosis and necroptosis are two avenues of programmed cell death [8–10].

Cancer has long inspired fears. In the distant past, physicians related depression or melancholic humour to cancer's pathogenesis. It was believed that melancholy could give rise to a tumor as people attributed their cancer to sadness. Recently, inflammation and nonspecific immune activation are found to be key factors in the pathophysiology of depression related to cancer. Urban centers are at an increasing risk of cancer-related risks due to factors like nutrition; infections such as sea turtle fibropapillomatosis and feline immunodeficiency virus; urban chemical pollution such as carcinogens, polychlorinated biphenyls, glutathione, and urethane-induced adenomas; light and noise pollution such as suppression of pineal melatonin production; changes in survival; and life history strategies [11–13].

Deep learning (DL) has seen phenomenal growth in recent years in the use of artificial intelligence allowing complex computational models to represent abstractions gathered from data with wide applications in speech processing, visual processing, and other domains. These methods work by discovering fine structures in large and often complex datasets using a backpropagation algorithm. Compared to DL, conventional methods such as machine learning-based methods have limitations in processing natural data in its basic form without preprocessing [14].

Convolutional Neural Networks (CNNs) are DL systems equipped with the power to learn invariant features. CNNs have filter banks, feature pooling layers, batch normalization layers, dropout layers and dense layers that work in harmony to create patterns for different object recognition tasks such as detection, segmentation, and classification. CNNs have multilevel hierarchies where the distribution of inputs changes during the process of training. Preprocessed inputs, such as those obtained through the process of whitening, etc., are highly desirable to obtain better performances across tasks [15]. CNNs have many different variants such as those offering shorter connections, for example, DenseNet architecture, which offer advantages in terms of feature circulation and offer substantial reduction in hyperparameters to build efficient architectures [16]. The focal and nonfocal electroencephalogram signals in tunable Q-factor wavelet transform domain have been investigated and identified with the help of feature selection and neural network methods [17]. A recent study concerning the low-density parity-check (LDPC) codes for Internet of things networks has been conducted via a novel technique for obtaining the first two minima of check-node update operation of the min-sum-LDPC decoder [18]. In addition, a review of future robust networks including various scenario for 6G has been discussed in [19].

Other types of CNN architectures that have gained popularity in recent years are ResNets, Xception, and GoogLeNet architectures. The need for these networks arises due to degradation in performances across tasks when the network is getting deeper, the need for multiscale processing, and the search for better architectures with less number of parameters [20–23].

Another issue that holds considerable importance in DL is the ability of an architecture to store information over extended time intervals. One solution proposed for this problem is Long Short-Term Memory (LSTM). LSTM architecture works by enforcing consistent error flow that is nonglobal in space and time through states of specialized units [24].

Another idea worth mentioning in DL is the notion of transfer learning. In transfer learning, features extracted from deep CNNs are repurposed to new and novel tasks. The need arises because generic tasks may differ by a wide margin from the original tasks due to which there may be insufficient labelled or other data to train or adapt a DL architecture to new tasks. Using transfer learning, features can be adapted to have sufficient generalization expression using simple techniques reliably [25–27].

Finding better architecture design parameters for DL models is a problem worth considering. Reinforcement learning methods can help in this task. An inspirational example is NASNet architecture that uses a number of different network topologies to find repeated motifs that can be combined in series to handle inputs of varying spatial dimensions and depth [28, 29].

This paper presents an overview of DL methods for the task of cancer diagnosis, prognosis, and prediction. The aim is to highlight the differences between different model constructions and to provide limitations and future perspectives for further exploration of this exciting domain.

The rest of the paper is organized as follows. Section 2 presents the gist behind the selection of studies that are

made a part of this survey article. Section 3 presents an overview of publically available datasets for cancer research followed by the description of current applications of DL in cancer diagnosis, prognosis, and prediction tasks in Section 4. Section 5 presents the discussion covering limitations of the existing methods, perspectives, and some directions for future work. Finally, Section 6 concludes this work and proposes avenues for further research in this domain.

2. Methodology

The criterion used for the selection of articles for this mini-review was language and authenticity of electronic sources. Articles written only in English language are made a part of this survey due to wide recognition of English as the language of scientific and biomedical domains. Years of sources of articles considered for this study range between 1997 and 2020. We used PubMed, Web of Science, IEEE Xplore, and Science Direct platforms to conduct the research. The search terms used were diagnosis of cancer, prognosis of cancer, prediction of cancer using DL, and transfer learning models.

3. Publically Available Datasets for Cancer Research Using DL Methods

In this section, we will provide a brief description of publically available datasets for cancer studies. We will briefly describe The Cancer Genome Atlas (TCGA) database, Rotterdam tumor bank, Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), MITOS-ATYPIA-14 dataset, Tumor Proliferation Assessment Challenge (TUPAC) 2016 dataset, INbreast database, Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) datasets, Lung Nodule Analysis (LUNA16) dataset, Breast Cancer Histopathological Image Classification (BreakHis) dataset, 2015 Bioimaging breast histology classification challenge, Cancer Metastases in Lymph Nodes Challenge breast cancer metastasis detection (CAMELYON) dataset, PatchCamelyon dataset, 2018 International Conference on Image Analysis and Recognition (ICIAR) dataset, MITOS12 dataset, Leukemia microarray gene data, Gene Expression Omnibus repository, BioGPS data portal, The Cancer Imaging Archive (TCIA), Genomic Data Commons (GDC), Therapeutically Applicable Research to Generate Effective Treatments (TARGET), 1000 Genomes Project, Kvasir dataset, University of California Santa Barbara Bio Segmentation Benchmark (UCSB-BB) dataset, and the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) dataset.

(1) *TCGA Database*. Beginning in 2006, TCGA is a result of a joint partnership between the National Cancer Institute and the National Human Genome Research Institute characterizing over 20,000 primary cancer and matched normal samples spanning 33 cancer types such as acute myeloid leukemia, adrenocortical carcinoma, breast lobular carcinoma, and uveal melanoma. The total number of cases in this database is approximately 11,125. It also contains cases of rare types of cancers. This database is available at https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga (accessed on September 24, 2021).

(2) *Rotterdam Tumor Bank*. This dataset [30] is comprised of 2982 primary breast cancer patients of whom 1546 are positive cases, i.e., they had the disease. Different factors such as estrogen receptors, progesterone receptors, hormonal treatment, number of positive lymph nodes, differentiation grade, and tumor size characterize this dataset. An R package of this dataset can be accessed at https://stat.ethz.ch/R-manual/R-devel/library/survival/html/rotterdam.html (accessed on September 24, 2021).

(3) *SUPPORT Database*. This database [31], gathered with the support of five teaching hospitals in the United States, is comprised of 9105 adults with an overall 6-month mortality rate of 47%. Subjects are recruited in two phases. Phase I recruited 4301 patients while phase II recruited 4804 patients. In phase II, the intervention group has 2652 subjects while the control group has 2152 subjects. Patients are diagnosed with one of nine life-threatening diagnoses.

(4) *METABRIC Dataset*. This dataset [32] is comprised of 2509 primary breast tumor subjects with 548 matched normal subjects. There are 2506 breast cancer subjects and 3 breast sarcoma subjects. Subjects with breast invasive ductal carcinoma are the most frequently occurring in this dataset while subjects of metaplastic breast cancer and breast angiosarcoma are the least frequently occurring classes. In total, there are eight classes of cancer subjects. This dataset is available at https://www.cbioportal.org/study/summary?id=brca_metabric (accessed on September 24, 2021).

(5) *MITOS-ATYPIA-14 Dataset*. This dataset contains histological images of breast cancer for the detection of mitotic cells and for the evaluation of nuclear atypia score for the prognosis of breast cancer. These annotations are provided by two senior and three junior pathologists. The dataset provides samples of haematoxylin and eosin-stained slides with the size of 1539×1376 pixels at 20x and 40x magnification levels. For every slide, the pathologists selected several frames at ×10 magnification which are further subdivided into four frames at ×20 magnification which are further subdivided into four frames at ×40 magnification. Evaluation metrics for mitosis are the number of detected mitosis, number of true positives, number of false positives, number of false negatives, sensitivity, precision, and F_1-value. This dataset is available at https://mitos-atypia-14.grand-challenge.org/Home/ (accessed on September 24, 2021).

(6) *TUPAC 2016 Dataset*. This dataset [33] provides a way to predict tumor proliferation scores from

whole-slide images. The challenge dataset is made up of 500 training and 321 testing breast cancer histology whole-brain slides. This dataset is designed to fulfill three purposes. The first one is to predict mitotic scores while the second one is to predict gene expression-based proliferation scores. A third task was later added to the challenge for mitosis detection.

(7) *INbreast Dataset*. This breast cancer dataset [34] has a total of 115 cases and is made up of full-field digital mammograms. The number of images of these cases is 410. In these 115 cases, 90 cases are from women with both breasts affected while 25 cases represent mastectomy patients. Several types of lesions such as masses, calcifications, asymmetries, multiple findings, normal, and distortions are included. Eight cases also have images acquired at different timings.

(8) *LIDC-IDRI Database*. Initiated by the National Cancer Institute (NCI), this dataset [35] of Computed Tomography (CT) scans contains 1018 cases of three categories: nodule ≥ 3 mm, nodule < 3 mm, and nonnodule ≥ 3 mm. A two-phase image annotation process was performed by four experienced thoracic radiologists. The goal of this dataset is to identify as completely as possible all lung nodules in each CT scan. This dataset is available at https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI#1966254194132fe653e4a7db00715f6f775c012 (accessed on September 24, 2021).

(9) *LUNA16 Dataset*. Collected from the LIDC-IDRI dataset, the LUNA16 [36] dataset is designed for the detection of pulmonary nodules from CT scans. The scans where slices were thicker than 2.5 mm were excluded. It facilitates lung nodule segmentation by providing the option of multiple candidates per nodule. In total, this dataset includes 888 CT scans. Irrelevant findings that were not made a part of this dataset include nonnodules, nodules < 3 mm, and nodules annotated by only 1 or 2 radiologists. This dataset is available at https://luna16.grand-challenge.org/Data/ (accessed on September 24, 2021).

(10) *BreakHis Dataset*. This dataset [37] of breast cancer subjects contains 9109 microscopic images. These images of tumor tissue are collected from 82 subjects at four different magnification levels which are 40x, 100x, 200x, and 400x. It contains 2480 benign and 5429 malignant samples. These samples are stored in PNG format. The resolution of each sample is 700×460 pixels, 3-channel RGB with eight-bit depth in each channel. This database resulted from the collaboration of the P&D Laboratory Pathological Anatomy and Cytopathology, Parana, Brazil, and Laboratory of Vision, Robotics, and Imaging, Federal University of Parana, Brazil. Benign tumors are slow-growing and remain localized to a region while malignant tumors can spread to distant regions and possess the ability to destroy adjacent structures which may cause death. This dataset is available at https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/ (accessed on September 24, 2021).

(11) *2015 Bioimaging Breast Histology Classification Challenge Dataset*. This dataset [38] of breast cancer subjects has four classes which are normal, benign, in situ carcinoma, and invasive carcinoma. It has high-resolution, uncompressed, and annotated H&E stain slides. The images have a resolution of 2040×1536 pixels. It supports both image and patch-wise classification tasks. This dataset is available at https://rdm.inesctec.pt/dataset/nis-2017-003 (accessed on September 24, 2021).

(12) *CAMELYON Dataset*. This dataset [39] is designed for breast cancer metastasis detection and classification in whole-slide images of histological lymph nodes. It facilitates patient-level analysis by combining the assessment of several lymph node slides into one outcome for direct deployment in a clinical setting which will facilitate pathologists while at the same time reducing the subjectivity in diagnosis. It contains 1399 unique whole-slide images of lymph nodes which have a slide-level label indicating whether it contains no metastases, macrometastases, micrometastases, or isolated tumor cells. In addition, the dataset has detailed contours drawn by experts for all metastases in 209 whole-slide images. This dataset is available at https://camelyon17.grand-challenge.org (accessed on September 24, 2021).

(13) *PatchCamelyon Dataset*. This dataset [40] contains histopathologic scans of breast cancer lymph node sections. Each image in this dataset is annotated with a label to indicate the presence of metastatic tissue. It contains 327,680 color images with a resolution of 96×96 pixels. It is trainable on a single GPU. For size comparisons, it is bigger than CIFAR10 and smaller than the ImageNet dataset. This dataset is available at https://www.tensorflow.org/datasets/catalog/patch_camelyon (accessed on September 24, 2021).

(14) *2018 ICIAR Dataset*. This dataset is composed of haematoxylin and eosin-stained breast histology microscopy and whole-slide images. The images are labelled as normal, benign, in situ carcinoma, or invasive carcinoma. There are a total of 400 microscopy images with 100 images per class stored with .tiff extension. The microscopy images are color images with a dimension of 2048×1536 pixels. The whole-slide images are color images with a dimension of 42113×62625 pixels and are stored in .svs format with pixel-wise labels. This dataset is available at https://iciar2018-challenge.grand-challenge.org/Dataset/ (accessed on September 24, 2021).

(15) *MITOS12 Dataset*. This dataset [41] contains 50 breast cancer biopsy slides at a 40x magnification level with more than 300 mitoses in these slides. The dimensions of these images are 2084×2084 pixels and 2252×2250 pixels. This dataset is helpful with mitotic count to estimate the aggressiveness of the breast tumor. This dataset is available at http://ludo17.free.fr/mitos_2012/dataset.html (accessed on September 24, 2021).

(16) *Leukemia Microarray Gene Data*. This dataset [42] contains gene expression data from 60 bone marrow samples of patients belonging to acute lymphoblastic leukemia, acute myeloid leukemia, chronic lymphocytic leukemia, chronic myeloid leukemia, and healthy bone marrow. Microarray technology has been instrumental in genome-wide expression studies especially as the knowledge of metazoan genomes is improving. Further information about this dataset is available at https://www.bioconductor.org/packages/devel/data/experiment/manuals/leukemiasEset/man/leukemiasEset.pdf (accessed on September 24, 2021).

(17) *Gene Expression Omnibus Repository*. This repository [43] provides comprehensive sets of microarray, next-generation sequencing, and other genomic data to facilitate research in different types of cancers. Further information about this repository is available at https://www.ncbi.nlm.nih.gov/geo/ (accessed on September 24, 2021).

(18) *BioGPS Data Portal*. This portal [44] supports eight species which are humans, mouse, rat, fruitfly, nematode, zebrafish, thale-cress, frog, and pig to facilitate research in genes. Many different kinds of cancers are supported such as lung cancer, breast cancer, esophageal cancer, thyroid cancer, pancreatic cancer, colorectal cancer, and colon cancer. This portal is available at http://biogps.org/#goto=welcome (accessed on September 24, 2021).

(19) *TCIA*. This service [45] provides deidentification and hosting of a large archive of medical images of cancer for public access using different modalities such as Magnetic Resonance Imaging (MRI), CT, and digital histopathology. It also provides supporting data such as patient outcomes, treatment details, and genomics. This service is available at https://www.cancerimagingarchive.net (accessed on September 24, 2021).

(20) *GDC*. This portal provides genomic, clinical, and biospecimen data for different types of cancers such as bone marrow, breast, eye, skin, lung, liver, and nervous system. It supports cancer research initiatives such as TCGA and TARGET. This portal is available at https://gdc.cancer.gov (accessed on September 24, 2021).

(21) *TARGET*. This program provides vast amounts of genomic data to estimate molecular alterations in childhood cancers. The goal is to use data for the development of effective, less toxic therapies. It drives research in acute lymphoblastic leukemia, acute myeloid leukemia, kidney tumors, neuroblastoma, osteosarcoma, etc. Further information is available at https://ocg.cancer.gov/programs/target# (accessed on September 24, 2021).

(22) *1000 Genomes Project*. The aim of this project [46] is to find common genetic variants by taking advantage of developments in sequencing technology. It helps in sequencing a large number of people to provide a comprehensive resource on human genetic variation. Cell lines and DNA are available for all 1000 Genomes samples. These samples are completely anonymous with no associated medical data. Further information about this project is available at https://www.internationalgenome.org/1000-genomes-summary (accessed on September 24, 2021).

(23) *Kvasir Dataset*. This dataset [47] is accessible at https://dl.acm.org/do/10.1145/3193289/abs/ (accessed on September 24, 2021). It is designed to facilitate research in gastrointestinal (GI) tract cancer. The initial dataset consists of 4000 annotated images belonging to 8 classes with 500 images per class. The anatomical landmarks are Z-line, pylorus, and cecum, while the pathological finding includes esophagitis, polyps, and ulcerative colitis. Resolution of images ranges from 720×576 up to 1920×1072 pixels. This dataset continues to play an important role in deep learning research.

(24) *UCSB-BB Dataset*. This dataset [48] contains images of human, monkey, and cat species at subcellular, cellular, and tissue levels with resolutions ranging from 300×200 to 1024×1024 pixels. There are 58 images of breast cancer belonging to malignant/benign classes in humans with sizes of 896×768 and 768×512 at the cellular level associated with ground truth data. This dataset is available at https://bioimage.ucsb.edu/research/bio-segmentation (accessed on September 24, 2021).

(25) *BRATS Dataset*. This dataset [49] is composed of clinical and synthetic images. The clinical data has 65 MRI scans of low- and high-grade glioma patients. Four MRI contrasts which are T1, T1c, T2, and FLAIR are represented by clinical data. The synthetic data has 35 high-grade and 30 low-grade glioma subjects. It is aimed at facilitating segmentation of tumors and patient survival through prediction and differentiation between tumor recurrence and progression. This dataset is available at https://www.med.upenn.edu/cbica/brats2020/ (accessed on September 24, 2021).

Table 1 displays a summary of the databases/services/projects to facilitate cancer research covered in this section.

Table 1: Summary of the datasets for cancer research.

Dataset/service/project	Link	Type(s) of cancer(s)	Description
TCGA database	https://www.cancer.gov/aboutnci/organization/ccg/research/structural-genomics/tcga	Multiple	33 cancer types, total no. of cases is 11125
Rotterdam tumor bank	https://stat.ethz.ch/R-manual/R-devel/library/survival/html/rotterdam.html	Breast cancer	2982 primary breast cancer patients; 1546 are positive cases
SUPPORT database	[31]	Multiple	9105 adults, an overall 6-month mortality rate of 47%
METABRIC dataset	https://www.cbioportal.org/study/summary?id=brca_metabric	Breast cancer	2509 primary breast tumor subjects, 548 matched normal control subjects
MITOS-ATYPIA-14 dataset	https://mitos-atypia-14.grand-challenge.org/Home/	Breast cancer	Resolution of 1539×1376 pixels at 20x and 40x magnification levels
TUPAC 2016 dataset	[33]	Breast cancer	500 training and 321 testing breast cancer histology whole-brain slides
INbreast dataset	[34]	Breast cancer	Total of 115 cases and 410 images
LIDC-IDRI database	https://wiki.cancerimagingarchive.net/display/Public/LIDCIDRI#1966254194132fe653e4a7db00715f6f775c012	Lung cancer	CT scans of 1018 subjects, three categories (i) nodule ≥ 3 mm, (ii) nodule < 3 mm, and (iii) nonnodule ≥ 3 mm
LUNA16 dataset	https://luna16.grandchallenge.org/Data/	Lung cancer	888 CT scans, facilitates segmentation studies
BreakHis dataset	https://web.inf.ufpr.br/vri/databases/breast-cancerhistopathological-database-breakhis/	Breast cancer	9109 microscopic images; four different magnification levels which are 40x, 100x, 200x, and 400x collected from 82 subjects
2015 Bioimaging Breast Histology Classification Challenge	https://rdm.inesctec.pt/dataset/nis-2017-003	Breast cancer	Four classes which are normal, benign, in situ carcinoma, and invasive carcinoma; resolution of 2040×1536 pixels
CAMELYON dataset	https://camelyon17.grand-challenge.org	Breast cancer	Facilitates patient-level analysis; 1399 unique whole-slide images; no metastases, macrometastases, micrometastases, and isolated tumor cells
PatchCamelyon dataset	https://www.tensorflow.org/datasets/catalog/patch_camelyon	Breast cancer	327,680 color images with resolution of 96×96 pixels; bigger than CIFAR10 and smaller than ImageNet dataset
2018 ICIAR dataset	https://iciar2018-challenge.grand-challenge.org/Dataset/	Breast cancer	Represent normal, benign, in situ carcinoma, and invasive carcinoma; 400 microscopy images with 100 images per class
MITOS12 dataset	http://ludo17.free.fr/mitos_2012/dataset.html	Breast cancer	50 biopsy slides; 40x magnification level; more than 300 mitoses
Leukemia microarray gene data	https://www.bioconductor.org/packages/devel/data/experiment/manuals/leukemiasEset/man/leukemiasEset.pdf	Bone marrow cancer	60 bone marrow samples; acute lymphoblastic leukemia, acute myeloid leukemia, chronic lymphocytic leukemia, chronic myeloid leukemia, and healthy bone marrow

Table 1: Continued.

Dataset/service/project	Link	Type(s) of cancer(s)	Description
Gene Expression Omnibus repository	https://www.ncbi.nlm.nih.gov/geo/	Multiple	Provides comprehensive sets of microarray, next-generation sequencing, and other genomic data
BioGPS data portal	http://biogps.org/#goto=welcome	Multiple	Supports eight species including humans; supports different types of cancers
TCIA	https://www.cancerimagingarchive.net	Multiple	Supports a large number of modalities; supports data such as patient outcomes, treatment details, and genomics
GDC	https://gdc.cancer.gov	Multiple	Provides genomic, clinical, and biospecimen data
TARGET	https://ocg.cancer.gov/programs/target#	Multiple	Childhood cancers are supported; provides vast amounts of genomic data to estimate molecular alterations
1000 Genomes Project	https://www.internationalgenome.org/1000-genomes-summary	Multiple	Provides a comprehensive resource on human genetic variation
Kvasir dataset	https://dl.acm.org/do/10.1145/3193289/abs/	Gastrointestinal tract cancer	4000 annotated images belonging to 8 classes
UCSB-BB dataset	https://bioimage.ucsb.edu/research/bio-segmentation	Supports breast cancer research in human species	Contains images of human, monkey, and cat species at subcellular, cellular, and tissue levels
BRATS dataset	https://www.med.upenn.edu/cbica/brats2020/	Brain tumor	MRI scans of 65 subjects each in clinical and synthetic datasets, for brain tumor segmentation task

4. Current Applications of Deep Learning in Cancer Diagnosis, Prognosis, and Prediction

In this section, we will discuss some current research trends in the domain of DL for cancer diagnosis, prognosis, and prediction tasks. We will cover techniques for the prognosis/prediction of tumors, breast cancer, and other types of cancer. In addition, we will also cover techniques for the segmentation/detection of breast cancer and other types of cancer. Furthermore, we will cover different methods for the classification of breast cancer and other types of cancer. We will also cover techniques for the classification, segmentation, and detection of brain tumors. Figure 1 shows histopathological views of some of the cancer subtypes that will be covered in this review article.

4.1. Prognosis/Prediction of Tumors, Breast Cancer, Skin Cancer, Head and Neck Cancer, Brain Cancer, Liver Cancer, Colorectal Cancer, Ovarian Cancer, and Other Types of Cancer. Petalidis et al. [50] reported a gene expression dataset for astrocytic tumors. They employed an Artificial Neural Network (ANN) algorithm to combine signatures from histopathological subclasses of these tumors in order to address the need for proper grading of these tumors. In this study, they found 59 genes which belong to three classes, namely, angiogenesis, lower-grade astrocytic tumor discrimination, and cell differentiation. They further report that these tumor subtypes have very high prognostic value, and they are missing in other studies reported in the literature. Finally, they report 11 classifiers that used genes to differentiate among primary/secondary subtypes of glioblastomas. They used a custom as well as independent dataset reporting accuracy of 96.15% to identify correct classes for these subtypes. Chi et al. [51] use morphometric features to compare prediction outcomes on two different breast cancer datasets. They report successful predictions with good and bad prognostic values. Here, good means that prognosis stands valuable even after five years while bad suggests less than five years. The authors in [52] conducted experiments in female breast carcinoma patients using a DL approach. They did prediction using a Cox regression model and gene expression datasets. They called their approach Survival Analysis Learning with MultiOmics Neural Networks (SALMON). They report that performance of SALMON is improved when more data is used to combine and simplify cancer biomarkers and gene expressions to enable prognosis prediction. Shimizu and Nakayama [53] conducted experiments to identify breast cancer genes for prognosis prediction using The Cancer Genome Atlas (TCGA) database. They identified 184 genes using artificial intelligence (AI),

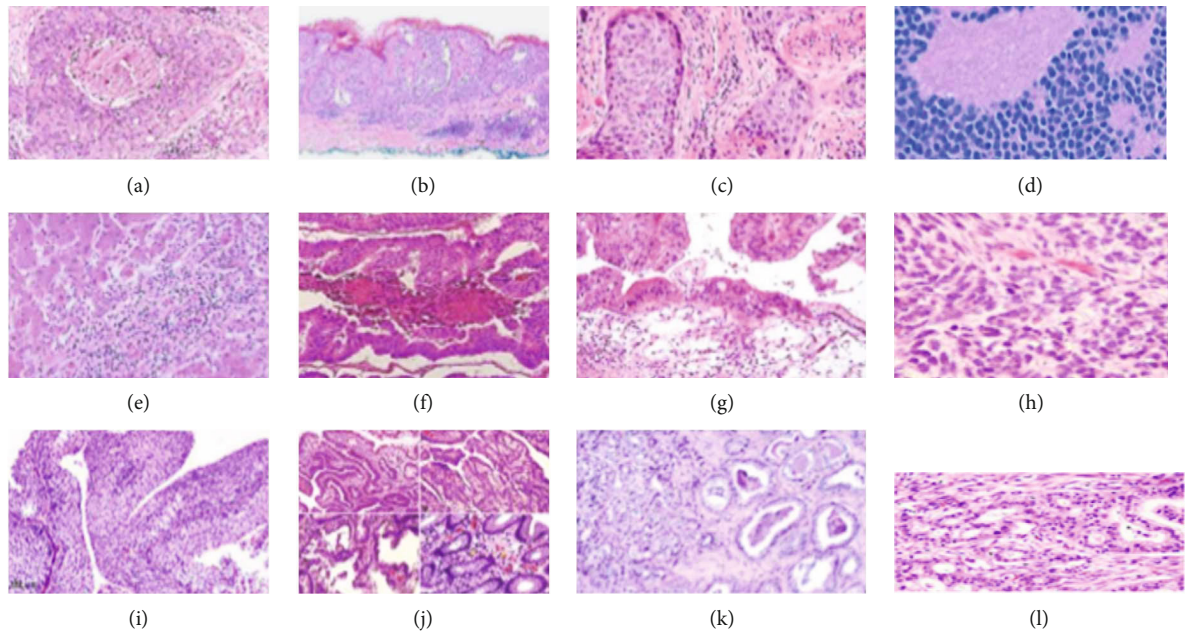


Figure 1: Histopathological images of different cancer subtypes: (a) breast cancer, (b) skin cancer, (c) head and neck cancer, (d) brain cancer, (e) liver cancer, (f) colorectal cancer, (g) ovarian cancer, (h) lung cancer, (i) bladder cancer, (j) gastric cancer, (k) prostate cancer, and (l) pancreatic cancer.

and for that purpose, they used random forest and neural network models. Furthermore, they used a molecular score for prognosis that uses only 23 of these genes. They confirmed that they have found potential drug targets in these genetic discoveries. The authors in [54] performed their experiments using malignant melanoma. They used a dataset with 1160 females and 786 males. They used an ANN architecture employing a flexible nonlinear structure for prognosis prediction of survival probabilities. They found the performance of their model to be at par with the Cox model with the advantage that it offers a flexible approach when analyzing data using a specified distributional form. Jing et al. [55] introduced a loss function combining a pairwise ranking loss and a mean squared error loss to optimize a DL model validated on four publically available datasets, such as the Worcester Heart Attack Study (WHAS), Rotterdam tumor bank, Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT), and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC). Their model achieved superior performance results than medical experts for nasopharyngeal carcinoma prognosis. Hao et al. [56] proposed a DL network predicting prognoses and describing complex biological pathways thus providing their model the power to interpret its outcomes. They conducted experiments for prediction in glioblastoma multiforme brain cancer from TCGA database. Their model achieves an Area under the Curve (AUC) of 0.6622 ± 0.013 and an F_1 score of 0.3978 ± 0.016 outperforming other models such as Logistic Least Absolute Shrinkage and Selection Operator (LASSO), Random LASSO, Support Vector Machines (SVMs), and a dropout neural network model which shows the superiority of their approach. The authors [57] put forward a multimodal DL network integrating multidimensional data. Their model combined gene expression data, alteration data, and other forms of clinical data achieving better performance than models with 1D data and other approaches. Chaudhary et al. [58] proposed the DL-based approach based on a combination of RNA and methylation data from TCGA to model hepatocellular carcinoma subjects. Their model achieves a concordance index of 0.68 and a p value of 7.13×10^{-6}. They found that TP53 mutations, KRT19 and EPCAM stemness markers, and Wnt and Akt signaling pathways are associated with more aggressive subtypes. The authors in [59] proposed a DL model combining CNN and Recurrent Neural Network (RNN) architectures for the prediction of colorectal cancer subjects using digitized haematoxylin-eosin-stained tumor tissue microarray samples. In the low- and high-risk patients, their model achieved a hazard ratio of 2.3 for visual assessment of histological tissues and a hazard ratio of 1.65 on the whole-slide level for both low- and high-risk subjects. Wang et al. [60] come up with a DL model to predict serous ovarian cancer subjects by extracting prognostic biomarkers from CT images. They further proposed a combined DL and Cox hazards model and achieved a concordance index of 0.713 and 0.694 for individual and three years of recurrence probability of subjects, respectively. The authors in [61] used a DL-based ANN model from transcriptomics data. They deployed TCGA datasets of RNA sequences belonging to ten different kinds of cancers. Their model achieved superior or equal level performance at both the pathway and gene levels. The authors in [62] came up with a DL model combining a Cox proportional hazards model with one of the best performing survival methods. They conducted experiments on WHAS, METABRIC, and SUPPORT datasets achieving good prediction performance levels for personalized treatment recommendations. Mobadersany et al. [63] predicted time-to-event results from histopathology images and gene-

based biomarkers using CNNs as DL models from glioma and glioblastoma cohorts of TCGA. They used a sampling- and filtering-based approach for the improvement of their predictions by not taking into account the intratumoral heterogeneity. Their model achieved a median concordance index of 0.754 surpassing other state-of-the-art approaches. The authors in [64] developed a DL-based approach using CNNs to predict the survival of mesothelioma cancer subjects. They used TCGA and a French source to test their approach. They achieved a concordance index of 0.656 on the TCGA cohort surpassing the performance of human experts and found key regions in the stroma that are associated with inflammation and cellular diversity. Liu et al. [65] modeled diagnostic prediction using DL models. The authors conducted their study on 27 diverse cancer types obtained from TCGA and Gene Expression Omnibus dataset. They successfully decoded 12 CpG and 13 promoter markers. The CpG markers that they identified achieved a sensitivity of 100% in the prediction of prostate cancer samples while promoter markers achieved 92% using cell-free deoxyribonucleic acid (DNA) methylation data.

Table 2 displays a summary of the studies for the task of prognosis and prediction of cancers covered in this subsection.

4.2. Segmentation/Detection of Breast Cancer, Lung Cancer, Bladder Cancer, and Other Types of Cancer. Yap et al. [66] used DL approaches for breast lesion detection using ultrasound images. They investigated the performance of LeNet, U-Net, and a pretrained AlexNet. They conducted their experiments on two custom datasets of 306 and 163 images termed dataset A and dataset B, respectively. Their pretrained AlexNet-based model achieved the best overall performance by achieving an F-measure of 0.91 and 0.89 on both datasets. The authors in [67] come up with different variants of fully convolutional networks (FCNs) for the segmentation of lesions of breast cancer subjects. They tried an AlexNet-based FCN, as well as 8-, 16-, and 32-layered FCN models. To overcome the problem of data deficiency, they used transfer learning and pretraining on the ImageNet dataset. Their dataset has two classes, benign and malignant. They reported an average dice score of 0.7626 using FCN with 16 layers on benign lesions. Their model correctly recognized 89.6% of benign lesions and 60.6% of the malignant lesions successfully. Liu et al. [68] used DL to detect breast cancer in lymph node biopsies. They used 399 slides from the Camelyon16 challenge dataset to achieve an AUC of 99% at the slide level. They used a second custom dataset that has 108 slides to achieve an AUC of 99.6%. As a preprocessing step, they used a color normalization procedure. The authors in [69] used different DL methods such as faster region CNN, ResNET-50, and DenseNet-201 architectures for breast cancer detection using histopathology images. They used three datasets to conduct their experiments which are International Conference on Pattern Recognition 2012, MITOS-ATYPIA-14, and Tumor Proliferation Assessment Challenge 2016 dataset. They achieved a precision of 0.876 on the International Conference on Pattern Recognition 2012 dataset, 0.848 on MITOS-ATYPIA-14, and a precision of 0.641 on the Tumor Proliferation Assessment Challenge 2016 dataset. As data augmentation methods, they employed horizontal and vertical flipping, translation, and resizing operations to artificially increase the size of datasets. Anuranjeeta et al. [70] used shape and morphological features derived from segmented images to detect cancer cells using a number of DL and machine learning-based models. They used J-Rip, logistic modal tree, rotation forest, multilayer perceptron, and other models trained by histopathological images. Rotation forest performed the best in cancerous/-noncancerous detection achieving an accuracy of 85.7%. The authors in [71] used a modified regional CNN method to efficiently determine mitosis in breast cancer using histopathological images. They employed subjects belonging to the 2014 International Conference on Pattern Recognition (ICPR) and TUPAC 2016 datasets in their study. They achieved 0.76 in precision on the TUPAC 2016 dataset. Zhou et al. [72] used a 3D deep CNN model to detect lesions in the breast cancer MRI dataset. They deployed a custom dataset with 1537 female patients and classify them as benign or malignant. They achieved an accuracy of 83.7% for the diagnostic task and a dice distance score of 0.501 for the detection task. The authors in [73] proposed a DL integrated architecture with the capability of performing classification, segmentation, and detection for the screening of breast masses as benign or malignant. They used digital X-ray mammograms from the INbreast database. Their model achieved a mass detection accuracy of 98.96%, while for mass segmentations, they achieved a dice score of 92.69%. To augment the dataset, the authors applied rotation 8 times to synthetically increase the size of the dataset. Nasrullah et al. [74] deployed DL-based architectures for the diagnosis of malignant nodules in lung cancer. They conducted studies on LUNA16 and LIDC-IDRI datasets. They used faster region CNN and U-Net styled architecture to achieve an accuracy of 94.17% on the classification task. The authors in [75] used a DL-based system for screening lung cancer using CT scans. They used LIDC-IDRI and Kaggle data science bowl challenge datasets for the experiments. Their system was based on 3D CNN architectures. The authors used heavy augmentations to artificially increase the size of the datasets using methods such as rotations, scaling, translation, and reflection. They achieved a dice coefficient of 0.4 on the LIDC-IDRI dataset. Shkolyar et al. [76] deployed DL-based models for the detection of papillary and flat bladder cancer. They used CNNs to construct an image analysis platform. They used two datasets of 100 and 54 subjects. Their model successfully detected 42 of 44 papillary and flat bladder cancers. They reported a per-tumor sensitivity of 90.9%. Fourcade et al. [77] used a combination of DL and superpixel segmentation-based methods to segment full body organs such as the brain and heart from Positron Emission Tomography (PET) images. To synthetically increase the size of the dataset, the authors deployed rotations, scaling, mirroring, and elastic deformations. Their best performing model achieved a dice score of 0.93. The authors in [78] deployed DL architectures to detect brain metastasis on MRI. They used data from 121 subjects in their proposed study. They used a faster region CNN model achieving an area under the ROC curve of 0.79. Ma et al.

Table 2: Summary of the studies for the prognosis/prediction of cancers.

Publication	Type(s) of cancer	Type of data	Methods	Performance
[50]	Astrocytic tumor	Microarray gene dataset	ANN	96.15% accuracy
[51]	Breast cancer	Nuclear morphometric features	ANNs	Good (>5 years) and bad (<5 years) prognoses
[52]	Breast invasive carcinoma	Gene expression data	Multiomics neural networks	Improved performance using more omics data
[53]	Breast cancer	TCGA	Random forest, neural network	Log-rank $p < 0.05$
[54]	Malignant melanoma	Custom dataset	Nonlinear ANN model	ANN model performs better than Cox model
[55]	Multiple	WHAS, SUPPORT, METABRIC, Rotterdam tumor bank	Deep feedforward neural network	Better prognostic accuracy than the clinical experts for the prognosis of nasopharyngeal carcinoma
[56]	Glioblastoma multiforme	TCGA	Pathway-associated sparse deep neural network	AUC = 0.6622 ± 0.013, $F_1 = 0.3978 \pm 0.016$
[57]	Breast cancer	Gene expression profile+copy number alteration profile+clinical data	Multimodal deep neural network	The proposed method achieves better performance than the prediction methods with single-dimensional data and other existing approaches
[58]	Hepatocellular carcinoma	TCGA	DL-based model	p value = 7.13×10^{-6} Concordance index = 0.68
[59]	Colorectal cancer	Images of tumor tissue samples	Combined convolutional and recurrent architectures	Prediction with only small tissue areas (hazard ratio 2.3), tissue microarray spot (hazard ratio 1.67), and whole-slide level (hazard ratio 1.65)
[60]	Ovarian cancer	CT images	Combined DL and Cox proportional hazards model	Concordance index was 0.713 and 0.694
[61]	Multiple	TCGA	ANN framework	Same or better predictive accuracy compared to other methods
[62]	Multiple	WHAS, SUPPORT, & METABRIC	Cox proportional hazards deep neural network	Superior in predicting personalized treatment recommendations
[63]	Lower-grade glioma and glioblastoma	TCGA	CNNs	Median concordance index = 0.754
[64]	Mesothelioma	TCGA+French source	CNNs	Concordance index of 0.656 on TCGA cohort
[65]	Multiple	TCGA+Gene Expression Omnibus dataset	DL-based model	For both marker types, the specificity of normal whole blood was 100%

[79] used you only look once v3 dense multireceptive fields CNN for thyroid cancer nodule detection. They used ultrasound images and deployed different data augmentation methods such as color jitter, change saturation, exposure, and hue on two datasets of 152 and 699 images. The number of images increased to 10845 after the application of data augmentation schemes. The values of mean average precision (mAP) reported by the authors were 90.05 and 95.23. Das et al. [80] proposed a system combining watershed segmentation, Gaussian mixture model (GMM), and deep neural network for the classification and segmentation of liver cancer using CT scans. Their model performed recognition of hemangioma, hepatocellular carcinoma, and metastatic carcinoma subjects. They employed 225 CT scans in their study achieving a dice score of 0.9743 on the testing set for the segmentation task and an accuracy of 99.38% for the multiclass classification task. The authors in [81] proposed a DL-based model for the segmentation of histopathology images of the liver organ. Their proposed DL model combined residual block, bottleneck block, and an attention decoder block. The authors further created a new dataset of 80 histopathology images which they named as the KMC liver dataset and proposed a joint loss function combining dice and Jaccard losses. They conducted their experiments on two datasets: KMC liver and multiorgan Kumar datasets. Each image in the Kumar dataset has a dimension of 1000×1000 while each image in the KMC liver dataset has a dimension of 1920×1440. Their model achieved a Jaccard index of 0.7206 on the KMC liver dataset and 0.6888 on the Kumar dataset. Wang and Chung [82] proposed a modified U-Net-based architecture for the segmentation and diagnosis of the colon gland. The authors employed two

datasets for the experiments: the gland segmentation dataset from Medical Image Computing and Computer-Assisted Intervention (MICCAI) challenge and an independent colorectal adenocarcinoma gland dataset. The authors conducted validation experiments on 378 images augmented using elastic transformation, cropping, rotation, flipping, blurring, and distortion operations. Their model achieved a dice score of 0.929 on the gland segmentation dataset from MICCAI and 0.89 on the independent colorectal adenocarcinoma gland dataset for the segmentation task. The authors in [83] proposed a CNN architecture combining three CNNs for the segmentation and classification of colorectal cancer using MRI images of 28 adenocarcinomas and 5 mucinous carcinomas. Their model achieved a dice score of 0.60, precision of 0.76, and recall of 0.55 on the testing set. The authors used cropping and min–max scaling as preprocessing methods. Juebin et al. [84] proposed segmentation algorithms based on U-Net models from ultrasound images of ovarian cancer. Image clipping was used as a preprocessing method. The methods are validated on 127 patients and a total of 469 images. Their best model achieved a dice score of 0.87, an average Pearson correlation of 0.90, and an average intraclass correlation of 0.89. Shibata et al. [85] used the mask R-CNN algorithm for the segmentation of 1208 healthy and 533 gastric cancer endoscopic images. The resolution of the images ranges from 640×480 to 1440×1080 pixels. Their model achieved an average dice index of 71%. Wang and Liu [86] proposed an architecture based on Deeplab v3+ for the segmentation of 1340 pathological slices of gastric cancer. The authors used image mirroring, random flip, scale, and rotation as augmentation techniques. Mean subtraction followed by division with variance was used as a preprocessing method. Their model achieved a dice score of 0.9166. Shrestha et al. [87] proposed a DL system combining four U-Net models. They used MRI images of prostate cancer from an online database. Each image has a resolution of $256 * 256$ pixels. They used a combination of modified dice and binary cross-entropy loss for the segmentation task. They preprocess the images using denoising and intensity normalization procedures achieving an overall average accuracy of 95.3%. Liu et al. [88] proposed a DL method integrating mask R-CNN and Inception version 3 models for the classification, segmentation, and detection tasks of prostate cancer. They used a dataset of 1200 ultrasound images. Their model achieved a dice score of 0.88 and a precision of 76% on malignant and 75% on benign classes for the classification task using an Inception v3 architecture. The authors in [89] proposed a 2D U-Net model deploying CT images of 556 cases of prostate cancer. They achieved a dice score of 0.85, 0.94, and 0.85 for three organs, namely, prostate, bladder, and rectum, respectively. Liang et al. [90] developed a DL-based model employing CNN architecture for the segmentation of pancreatic tumors. The authors deployed a dataset of T1w MRI images of 40 subjects. They achieved a dice score of 0.73 using rotation and flipping as data augmentation methods. The authors in [91] proposed a DL method using spiral transformation to perform segmentation of MRI images of pancreatic cancer. The authors used rotation and spiral transformation as data augmentation methods. They deployed a dataset of MRI images belonging to 73 patients. Their architecture is a combination of ResNet and U-Net architectures. They achieved a dice score of 0.656 ± 0.1021.

Table 3 displays a summary of the studies for the task of segmentation and detection of cancers covered in this subsection.

4.3. Classification of Breast Cancer. Huynh et al. [92] used DL methods to classify regions of interest taken from ultrasound images. Cystic, benign, or malignant labels were assigned to each region. Two binary classification tasks were performed using pretrained CNNs, nonmalignant (benign+cystic)/malignant and benign/malignant. They used SVM as a classifier on the CNN-derived features. On the nonmalignant/malignant classification task, they obtained an AUC of 0.9 while on the benign/malignant task, their method obtained an AUC of 0.88. The authors in [93] used CNNs as their DL approaches and introduced the concept of a matching layer to convert grayscale to red, green, and blue patterns. They used 882 ultrasound images obtained from two publicly available datasets. Using fine-tuning and matching layer, their method approached an AUC of 0.936 on a test set of 150 cases. Byra et al. [94] used DL transfer learning-based approaches such as Inception version 3 and VGG19 architectures on reconstructed B-mode images experiencing a decrease in classification performances. To counter this, they used data augmentation to reconstruct B-mode images achieving better performances on breast ultrasound images. The authors in [95] combined cross-modal and cross-domain transfer learning for the benign/-malignant classification task. In comparison to training from scratch and simple fine-tuning, their approach achieved better performance with 97% accuracy on ultrasound images. Hadad et al. [96] deployed cross-modal transfer learning using mammography images achieving an accuracy of 0.93 which is better than cross-domain transfer learning. The authors in [97] presented a study on the use of MRI in screening individuals younger than 40 years confirming the effectiveness of MRI as a modality of choice for such diagnoses. They reported a very high sensitivity around 93% to 100% and low specificity in the range of 37% to 97%. They found MRI to be effective especially after reconstructive surgery. Hu et al. [98] developed a transfer learning methodology using an MRI modality with multiple parameters. They used different sequences such as dynamic contrast-enhanced and a T2-weighted sequence to distinguish benign lesions from malignant. They used image, feature, and classifier fusion methods and achieved an AUC of 0.87 for the feature fusion scheme that statistically outperformed other methods. The authors in [99] proposed a methodology using Inception version 4 and the residual network transfer learning architectures as well as a recurrent CNN architecture on the 2015 Breast Cancer Classification Challenge and Break-His datasets for binary and multiclass classification tasks. They used rotation, translation, and other data augmentation methods to artificially increase the size of the datasets achieving an accuracy of $97.57 \pm 0.89\%$ on multiclass and

Table 3: Summary of the studies for the segmentation/detection of cancers.

Publication	Type(s) of cancer	Type of data	Methods	Performance
[66]	Breast cancer	Ultrasound images, 2 datasets (A & B)	LeNet, U-Net, AlexNet	F-measure = 0.91 (on dataset A) and F-measure = 0.89 (on dataset B)
[67]	Breast lesions	Two custom datasets	FCN-AlexNet, FCN-32s, FCN-16s, and FCN-8s	Dice score of 0.7626 (FCN-16s)
[68]	Breast cancer	Camelyon16	DL algorithm	Slide-level AUC of 99%
[69]	Breast cancer	International Conference on Pattern Recognition 2012, MITOS-ATYPIA-14, Tumor Proliferation Assessment Challenge 2016	Faster region CNN, ResNET-50, DenseNet-201	0.691 F_1-measure for the MITOS-ATYPIA-14 dataset
[70]	Multiple	Custom histopathology image dataset	Multilayer perceptron, logistic modal tree, sequential minimal optimization, Naïve Bayes, random forest, rotation forest, J-Rip, and PART algorithms	Rotation forest algorithm achieved an accuracy of 85.7% for binary classification between cancerous and noncancerous cells
[71]	Breast cancer	ICPR 2014 mitosis dataset, TUPAC 2016 mitotic cell dataset	Modified regional CNN	Precision = 0.76, recall = 0.72, F_1-score = 0.736 on TUPAC 2016 dataset
[72]	Breast cancer	Custom dynamic contrast-enhanced MRI dataset	3D deep CNN architecture	83.7% accuracy, 90.8% sensitivity, 69.3% specificity, AUC of 0.859, overall dice distance of 0.501 ± 0.274
[73]	Breast cancer	INbreast database	Different DL methods	Accuracy of 98.96%, MCC of 97.62%, F_1-score of 99.24%, Jaccard similarity coefficient of 86.37%
[74]	Lung nodules	LUNA16, LIDC-IDRI	Two deep 3D customized mixed link network encoder-decoder architectures	Accuracy of 94.17%
[75]	Lung cancer	LIDC-IDRI dataset, Kaggle data science bowl challenge dataset	3D CNN architectures	Dice coefficient for LIDC-IDRI of 0.40, with 0.25 precision and 0.93 recall
[76]	Bladder cancer	Custom datasets	DL algorithm	Per-frame sensitivity and specificity were 90.9% and 98.6%
[77]	Full body	PET images	DL-based approach	Dice score of 0.93 ± 0.05
[78]	Brain metastases	Custom MRI dataset	DL-based approach (faster region-based CNN model)	96% sensitivity, AUC = 0.79
[79]	Thyroid nodules	Two custom datasets of ultrasound images	You only look once v3 dense multireceptive field CNN	mAP = 90.05 and 95.23
[80]	Liver cancer	225 CT scans of hemangioma, hepatocellular carcinoma, and metastatic carcinoma	Watershed segmentation, Gaussian mixture model (GMM), and deep neural network	Dice score of 0.9743, accuracy of 99.38%
[81]	Liver cancer	KMC liver dataset, multiorgan Kumar dataset	DL model combining residual block, bottleneck block, and attention decoder	Jaccard index of 0.7206 on KMC liver dataset and 0.6888 on Kumar dataset
[82]	Colorectal cancer	MICCAI gland segmentation dataset, colorectal adenocarcinoma gland dataset	Modified U-Net-based architecture	Dice score of 0.929 on MICCAI gland segmentation dataset, 0.89 on the colorectal adenocarcinoma gland dataset
[83]	Colorectal cancer	Custom dataset of MRI images of 28 adenocarcinomas and 5 mucinous carcinomas	CNN architecture which is a combination of three CNN architectures	Dice score of 0.60, precision of 0.76, and recall of 0.55

Table 3: Continued.

Publication	Type(s) of cancer	Type of data	Methods	Performance
[84]	Ovarian cancer	Custom dataset of 127 patients and a total of 469 images	U-Net models	Dice score of 0.87, an average Pearson correlation of 0.90, and an average intraclass correlation of 0.89
[85]	Gastric cancer	Custom dataset of 1208 healthy and 533 endoscopic images	Mask R-CNN algorithm	Average dice index of 71%
[86]	Gastric cancer	Custom dataset of 1340 pathological slices	Deeplab v3+	Dice score of 0.9166
[87]	Prostate cancer	MRI images from an online database	DL system combining four U-Net models	Overall average accuracy of 95.3%
[88]	Prostate cancer	Custom dataset of 1200 ultrasound images	DL method integrating mask R-CNN and Inception version 3 models	Dice score of 0.88, a precision of 76% on malignant and 75% on benign classes for the classification task using an Inception v3 architecture
[89]	Prostate cancer	Custom dataset of CT images of 556 cases	2D U-Net model	Dice score of 0.85, 0.94, and 0.85 for prostate, bladder, and rectum, respectively
[90]	Pancreatic cancer	Custom dataset of T1w MRI images of 40 subjects	CNN architecture	Dice score of 0.73
[91]	Pancreatic cancer	Custom dataset of MRI images belonging to 73 patients	DL method using spiral transformation	Dice score of 0.656 ± 0.1021

an accuracy of $97.95 \pm 1.07\%$ on binary classification tasks. Bayramoglu et al. [100] used single and multitask CNN architectures to predict malignancy and image magnification levels. Cropping and rotation on the BreakHis dataset were deployed to augment the dataset. They achieved a classification rate of 83.72% for the benign/malignant binary classification task using a single task and an accuracy of 82.13% using multitask CNN architecture. The authors in [101] proposed an approach for progressive combining of weak DL classifiers into a stronger classifier for carcinomas/noncarcinomas binary and normal/benign/in situ/invasive carcinomas multiclass (4 classes) classification tasks. They used BreakHis and 2015 bioimaging breast histology classification challenge datasets. They deployed augmentation methods such as reflection, random cropping, rotation, and translation of an image. They achieved a classification accuracy of 99.5% and 96.9% for binary classification tasks using the 2015 bioimaging breast histology classification challenge database and BreakHis database while for multiclass classification, they achieved 96.4% classification accuracy on the 2015 bioimaging breast histology classification challenge database. Kassani et al. [102] used an ensemble of transfer learning architectures for binary classification of breast cancer. They used VGG19, MobileNet, and DenseNet architectures on four benchmark datasets: BreakHis, Patch-Camelyon, 2015 Bioimaging challenge, and 2018 ICIAR datasets. They used flipping, zoom, shear, rotation, etc., as data augmentation methods. They achieved accuracies of 98.13%, 94.64%, 95%, and 83.1% on these datasets. The authors in [103] proposed a DL method for multiclass (8 classes) classification of histopathological images on the BreakHis dataset. Data augmentation methods such as rotation, flipping, sharing, and their combinations were deployed to achieve a correct classification rate of 95.48% on the multiclass classification task. Toğaçar et al. [104] deployed a DL model for multiclass (8 classes) classification of breast histopathological images on the BreakHis dataset. They used convolutional, attention, residual, pooling, and dense blocks along with hypercolumn technique to build their architecture. As data augmentation methods, they used flipping, shifting, change of brightness, and rotation achieving 98.51% accuracy. The authors in [105] used a combination of DenseNet and Xception transfer learning architectures for benign/malignant binary and magnification-specific multiclass classification tasks. They used the BreakHis dataset achieving an accuracy of 99% and 92% on binary and multiclass classification tasks, respectively, while deploying stain normalization for preprocessing of images. Spanhol et al. [106] performed experiments for the binary (benign/malignant tumors) classification task using histology images. They report an accuracy of $90.0 \pm 6.7\%$ on images obtained from the BreakHis dataset. The authors in [107] proposed a DL model for the multiclass (8 classes) classification task using histopathological images. They conducted experiments on the BreakHis dataset. For data augmentation, the authors deployed rotation, level/vertical flipping, translation techniques, etc., and their combinations achieving a patient-level accuracy of $94.7 \pm 3.6\%$. Bardou et al. [108] compared CNNs and traditional machine learning techniques such as bag of words and linear coding using SVMs. They deployed BreakHis datasets for both binary and multiclass (8 classes) classification tasks to categorize images into benign/malignant classes and their subclasses. The authors achieved accuracies of 98.33% and 88.23% for binary and multiclass classifications, respectively, using the deployed approaches. The authors in [109] combined four residual networks for binary (benign/malignant)

and multiclass classification of histology images using the BreakHis dataset. They achieved an accuracy of 96.25% for the eight-class classification task. The authors deployed rotation, flipping, translation, and color variation augmentation as data enhancing methods while stain normalization as a preprocessing method. Budak et al. [110] used a DL model combining FCN and bi-LSTM architectures on the BreakHis dataset for binary (benign/malignant) classification achieving an accuracy of 96.32%. The authors in [111] used a DL-based model for binary (benign/malignant) classification of histopathological images. They performed experiments on the BreakHis dataset fusing ResNet-18, ResNet-50, and AlexNet architectures using belief theory. They achieved an image-level accuracy of 96.88%. Sudharshan et al. [112] deployed a weakly supervised scheme for the binary classification of benign and malignant tumors using histopathology images. They deployed the BreakHis dataset achieving an accuracy of 92.1% at 40x magnification. An important contribution of their approach is the absence of the need for labelling the images. The authors in [38] deployed CNNs for both binary (carcinoma/noncarcinoma) and multiclass (normal/-benign/in situ/invasive) classification tasks. They used the 2015 Bioimaging breast histology classification challenge dataset in their study. Their architectures were able to retrieve information at different scales. For the multiclass classification task, the authors achieved an accuracy of 77.8% while for the binary classification task, they achieved an accuracy of 83.3% using rotation and mirroring as data enhancement methods for both these tasks. Rakhlin et al. [113] deployed different transfer learning architectures using microscopic histological images from the ICIAR 2018 Grand Challenge dataset for binary (carcinomas/noncarcinomas) and multiclass (four classes) classification tasks. They used pretrained ResNet-50, Inception version 3, and VGG16 architectures. They deployed normalization, downscaling, cropping, and color variation as augmentation schemes achieving a correct classification rate of 87.2% for multiclass classification and 93.8% for the binary classification task. The authors in [114] extracted smaller/larger patches using a clustering approach and a CNN (ResNet-50 architecture) at cell and tissue levels deploying the 2015 Bioimaging breast histology classification challenge dataset. For the multiclass (4 classes) classification task, the authors reported accuracy of 88.89% using the proposed approach. The authors deployed stain normalization procedure as a preprocessing method. Shallu and Mehra [115] demonstrated the use of three different transfer learning architectures such as VGG16, VGG19, and ResNet-50 for the classification of histological images on the BreakHis dataset. They deployed rotation as the data enhancement scheme. They found the performance of a fine-tuned VGG16 with logistic regression classifier to be the best achieving an accuracy of 92.6% with this classifier. The authors in [116] deployed CNN, K nearest neighbour (KNN), Inception version 3, SVM, and ANN algorithms for the binary (benign/malignant) classification task. They used different schemes for preprocessing and data enhancement such as gray scaling, channel standardization, flipping, rotation, and cropping as well as image segmentation to reach an accuracy of 97% using ANN architecture. Bevilacqua et al. [117] evaluated two different frameworks for binary and multiclass classification of irregular/regular/-stellar/no opacity lesions from segmented high-resolution images. They used ANN classifiers with hand-crafted and morphological features for the first framework. For the second framework, they used different CNN models especially a VGG model. They reported accuracy of 84.19% for the first framework on binary and 74.84% on multiclass classification tasks while for the second framework, they obtained an accuracy of 92.02% for binary and multiclass classification tasks. The authors in [118] make a contrast between two machine learning approaches for the multiclass (8 classes) classification task using histopathological images on the BreakHis dataset. The first approach used handcrafted features while the second approach used CNN as a feature extractor. They used VGG16, VGG19, and ResNet-50 as their CNN models. They used rotation, translation, scaling, and flipping as data enhancement methods. The VGG16 model reaches an accuracy of 93.25% at the patient level for the multiclass classification task. Spanhol et al. [119] proposed a DL model that reused a previously trained CNN model on the BreakHis dataset achieving an F_1-score of 90.3 at the subject level. The authors in [120] exploited global covariance information using a matrix power normalization procedure into a simple CNN model. This arrangement can exploit second-order statistical information producing effective representations from histological images. On the BreakHis dataset for the binary (benign/malignant) classification task, they achieved an accuracy of 97.92% at the subject level while employing cropping and flipping operations to enhance the size of the dataset synthetically. Khan et al. [121] used different transfer learning (GoogLeNet, VGGNet, and ResNet) architectures for binary classification of benign/malignant tumor cells while deploying BreakHis and a custom dataset. For data augmentation, scaling, rotation, translation, and color augmentation methods were used by them to achieve a correct classification rate of 97.67%. The authors in [122] introduced an information-based architecture that is designed to exploit clinical information. There are six types of records in their dataset of 100 subjects, such as encounter notes, operation records, pathology notes, radiology notes, progress notes, and discharge summaries. They used fine-tuned transformer models from pretrained bidirectional encoder representations achieving a precision of 0.976 for relation recognition. Naik et al. [123] deployed a DL model to assess estrogen status from whole-slide histopathological images. They used the Australian Breast Cancer Tissue Bank as well as TCGA datasets in their study and further deployed flipping, rotation, color jitter, and cutout regularization as augmentation methods. Their model achieved an AUC of 0.861 on TCGA and an AUC of 0.905 on Australian Breast Cancer Tissue Bank datasets. The authors in [124] compared different DL techniques for the classification of mammograms. They used single as well as 4-model averaging to conduct their experiments on INbreast as well as an independent database. They used different data enhancement techniques such as flipping, rotation, intensity shifting, and zoom. The single model achieved an AUC of 0.88 while 4-model averaging achieved an AUC of 0.91 on the independent database. On the INbreast dataset, the single model achieved an AUC of

0.95 while 4-model averaging achieved an AUC of 0.98. Their study shows the superiority of combining models over a single model.

Table 4 displays a summary of the studies for the classification of breast cancer covered in this subsection.

4.4. Classification of Colorectal Cancer, Gastric Cancer, Bladder Cancer, Lung Cancer, Prostate Cancer, Skin Cancer, Liver Cancer, Head and Neck Cancer, Pancreatic Cancer, and Other Types of Cancers. Kather et al. [125] deployed different transfer learning architectures for multiclass (9 classes) classification of colorectal cancer subjects. They used VGG19, AlexNet, SqueezeNet, GoogLeNet, and ResNet-50 models on two datasets of 86 and 25 subjects reaching an accuracy of 98.7% and greater than 94% on them. The authors in [126] deployed a CNN architecture to extract features from Optical Coherence Tomography (OCT) images of colorectal cancer subjects. Their network was trained using 26000 OCT images representing 42 areas achieving an AUC of 0.998. Dong et al. [127] deployed a DL method to exploit information in multiphase CT nomograms in gastric cancer subjects. They used three cohorts to test the effectiveness of their model achieving a discrimination rate of 0.821, 0.797, and 0.822 in the primary, external validation, and international validation cohorts. Woerl et al. [128] deployed a DL method to identify bladder cancer from histomorphological images. They used 2 datasets of 407 and 16 subjects each from TCGA and custom cohorts, respectively, achieving accuracies of 69.91% and 75% on TCGA and custom subsets, respectively. Wang et al. [129] used the idea of weakly supervised learning exploiting image-level labels for the classification of lung cancer images. They used two datasets, one from TCGA and the other is a custom dataset. To enhance the training set, color jittering, translation, flipping, and rotation were used.

Their model successfully achieves an accuracy of 97.3% and an AUC of 85.6% on custom and TCGA datasets. Karimi et al. [130] used a DL method combining three separate CNNs that used different patch sizes for the classification of histopathological images with limited data. They used new data enhancement methods such as elastic deformation and augmentation in the space of learned features for binary classification of cancerous/benign and low-grade–high-grade patches achieving an accuracy of 92% and 86%, respectively, on both binary classification tasks. Dascalu and David [131] used DL architectures for binary classification of benign/malignant cases of skin cancer subjects. They used a skin magnifier with polarized light and an advanced dermoscope to construct their datasets. The authors achieved an F2-score sensitivity of 91.7% and 89.5% respectively for skin magnifier with polarized light and advanced dermoscope images. The authors in [132] used DL techniques to build a skin cancer classification model for binary and multiclass classification of malignant and benign skin tumors. They used Kaohsiung Chang Gung Memorial Hospital and HAM10000 datasets in their study. Their model achieved an accuracy of 85.8% on the HAM10000 dataset for 7-class classification tasks. On the Kaohsiung Chang Gung Memorial Hospital dataset, their model achieved an accuracy of 72.1% for 5-class classification and 89.5% for binary classification tasks. Thomas et al. [133] applied interpretable DL models to classify skin cancers in a histopathological setting. They studied three types of cancers basal cell carcinoma, squamous cell carcinoma, and intraepidermal carcinoma. They deployed a multiclass (12 classes) classification model to achieve accuracies between 93.6% and 97.9%. To solve the class imbalance problem and to increase the size of the dataset, they used flipping and rotation as data augmentation methods to increase the size of the dataset 8 times. The authors in [134] developed a CNN model for the classification of melanoma and nevi. They used a dataset of 11444 images belonging to five categories. They deployed novel DL techniques to train a single CNN model. In addition, they also asked 112 dermatologists to grade the images. Then, they used a gradient boosting method to develop a new classifier for binary (benign/malignant) and multiclass (5 classes) classification tasks achieving accuracies of 86.5% and 82.95% on both these tasks. Sun et al. [135] developed a DL method to classify liver cancer subjects as abnormal/normal on publically available TCGA datasets. Transfer learning and multiple instance learning were combined for the classification of patch features. The authors used tissue extraction, color normalization, and patch extraction for preprocessing of histopathological images. Diao et al. [136] used a transfer learning-based CNN architecture named Inception version 3 to classify nasopharyngeal carcinoma subjects into three classes. They used a total of 1970 images of 731 subjects. The three classes considered in their study were chronic nasopharyngeal inflammation, lymphoid hyperplasia, and nasopharyngeal carcinoma. Their model achieved a mean AUC of 0.936. Liu et al. [137] used a CNN classifier to diagnose subjects with pancreatic cancer using contrast-enhanced CT images. They used three different datasets to test the effectiveness of their approach. The first dataset named local test set 1 has 295 patients with pancreatic cancer and 256 controls for training and 75 patients with pancreatic cancer and 64 controls for validation. The second dataset named local test set 2 has 101 patients with pancreatic cancers and 88 controls while the third dataset named the US dataset has 281 pancreatic cancer subjects and 82 controls. In local test set 1, local test set 2, and US datasets, their model achieved an accuracy of 98.6%, 98.9%, and 83.2%, respectively. To augment the datasets, the authors used moving window and flipping operations. Korfiatis et al. [138] compared the performances of ResNet-18, ResNet-34, and ResNet-50 architectures for the classification of MRI scans of 155 subjects for multiclass (3 classes) classification of no tumor, methylated methylguanine methyltransferase methylation, or nonmethylated classes. ResNet-50 architecture achieved the best performance with an accuracy of 94.9%; ResNet-34 architecture achieved an accuracy of 80.72% while ResNet-18 architecture achieved an accuracy of 76.75%. The authors in [139] used a two-phase training to study and mitigate class biasedness using a DL-based CNN model for the classification of breast cancer histological images. They conducted their experiments using MITOS12 and 2016 Tumor Proliferation

Table 4: Summary of the studies for the classification of breast cancer.

Publication	Type of data	Methods	Performance
[92]	Custom dataset of ultrasound images	Pretrained CNNs	AUC of 0.90 (nonmalignant vs. malignant), AUC of 0.88 (benign vs. malignant)
[93]	1 custom+2 publically available datasets	DL-based approach using a matching layer	AUC = 0.936 (custom dataset), AUCs around 0.89 (publically available datasets)
[94]	Custom dataset of ultrasound images	Inception v3, VGG19	Robust and efficient classification performances
[95]	Custom dataset of ultrasound images	Training from scratch, pretrained VGG16, fine-tuning approach	0.97 accuracy, 0.98 AUC using fine-tuning approach
[96]	Custom dataset of breast MRI images	Cross-modal transfer learning approach	Overall accuracy of 0.93 using cross-modal approach
[97]	Custom dataset of breast MRI images	DL-based method	High sensitivity in the range of 93-100%
[98]	Custom dataset of multiparametric MRI images	Pretrained CNN architectures	$AUC_{\text{feature fusion}} = 0.87$
[99]	BreakHis, Breast Cancer Classification Challenge 2015	Inception recurrent residual CNN model	100% for the binary and multiclass (Breast Cancer Classification Challenge 2015 dataset)
[100]	BreakHis	Single-task CNN, multitask CNN	Patient score of 83.72% for binary classification using single-task CNN
[101]	2015 bioimaging breast histology classification challenge, BreakHis dataset	Progressive DL-based models	Recognition rate of 96.4% and 99.5% on multiclass and binary classification tasks on 2015 bioimaging breast histology classification challenge
[102]	BreakHis dataset, PatchCamelyon dataset, 2015 Bioimaging challenge dataset, 2018 ICIAR dataset	VGG19, MobileNet, DenseNet	Accuracy of 98.13% on BreakHis dataset
[103]	BreakHis dataset	DL and hierarchical classification approach	Accuracy of 95.48% on the multiclass classification task
[104]	BreakHis dataset	Integrated DL model	98.51% classification success on the multiclass classification task
[105]	BreakHis dataset	DenseNet and Xception architectures	99% and 92% accuracy on binary and multiclass classification tasks
[106]	BreakHis dataset	DL-based model	Mean recognition rate of 90.0 ± 6.7 for binary classification
[107]	BreakHis dataset	DL-based model	Accuracy of 94.7 ± 3.6 for multiclass classification
[108]	BreakHis dataset	Bag of words, locality-constrained linear coding, CNNs	For CNN model accuracies between 96.15% and 98.33% for the binary classification and 83.31% and 88.23% for the multiclass classification
[109]	BreakHis dataset	Combination of 4 residual networks	Correct classification rate of 96.25% for 8-class categorization
[110]	BreakHis dataset	End-to-end model based on FCN and bidirectional LSTM	Accuracy of 96.32 ± 0.51 on binary classification task
[111]	BreakHis dataset	ResNet-18, ResNet-50, and AlexNet	Image-level accuracy of 96.88% for binary classification
[112]	BreakHis dataset	Weakly supervised learning framework	Classification rate of up to 92.1% for binary classification
[38]	2015 Bioimaging challenge dataset	CNN models	Accuracies of 77.8% for four classes and 83.3% for carcinoma/noncarcinoma were achieved
[113]	ICIAR 2018 Grand Challenge	Pretrained ResNet-50, Inception v3, and VGG16 architectures	Accuracies of 87.2% for multiclass, 93.8% for binary classification tasks
[114]	2015 Bioimaging challenge database	Clustering algorithm and ResNet-50 architecture	88.89% accuracy on the overall test set for multiclass classification
[115]	BreakHis dataset		92.60% accuracy

Table 4: Continued.

Publication	Type of data	Methods	Performance
		VGG16, VGG19, and ResNet-50 architectures	
[116]	Custom dataset	CNN, KNN, Inception v3, SVM, and ANN models	Accuracy of 97% using ANN algorithm for binary classification
[117]	Custom dataset	CNN and ANN models	Accuracy of $92.02 \pm 0.51\%$ for the binary classification task using a VGG model, accuracy of $92.02 \pm 0.48\%$ for the multiclass classification task
[118]	BreakHis dataset	VGG16, VGG19, ResNet-50 architectures	Accuracy of 93.25% for multiclass classification task
[119]	BreakHis dataset	DL-based model	F_1-score of 90.3
[120]	BreakHis dataset	Deep second-order pooling network	Accuracy of 97.92% for binary classification
[121]	BreakHis+custom datasets	Pretrained CNN architectures (GoogLeNet, VGGNet, and ResNet)	Accuracy of 97.67% for binary classification
[122]	Custom dataset	Transformer models	Precision of 0.976 for relation recognition
[123]	Australian Breast Cancer Tissue Bank, TCGA dataset	Deep neural network	AUC on TCGA of 0.861, AUC on Australian Breast Cancer Tissue Bank was 0.905
[124]	INbreast database	DL models	AUC = 0.98

Assessment Challenge datasets. Prior to phase 1 of training, segmentation using the global binary thresholding method was applied. In phase 1, a CNN was trained on the segmented patches using rotation and flipping data augmentation methods as well as the blue ratio histogram-based k-means clustering approach. In phase 2, the dataset was again modified to reduce the effects of class imbalance yielding an F-measure of 0.79. Campanella et al. [140] proposed a DL-based system utilizing information from multiple instances in order to help the pathologists exclude information without compromising performance metrics. They used 44732 whole-slide images belonging to 15187 patients. They achieved AUC above 0.98 and 100% sensitivity for prostate cancer, basal cell carcinoma, and breast cancer metastases to axillary lymph nodes. The authors in [141] proposed two DL-based systems to detect myeloid leukemia from the leukemia microarray genetic dataset. The first DL system is a single-layered neural network while the second one has 3 hidden layers. They used information of 22283 genes extracted from the Gene Expression Omnibus repository. Their models achieved accuracies of 63.33% and 96.67% for single and multilayered DL architectures with a significant normalization test ($p > 0.05$). Jeyaraj and Samuel Nadar [142] used a regression-based DL algorithm to investigate hyperspectral images to diagnose oral cancer. Their system extracted patches for classification into normal, benign, and malignant classes using BioGPS, TCIA, and GDC datasets. For 100 malignant image patch training, they achieved an accuracy of 91.4% while for 500 malignant image patch training, they achieved an accuracy of 94.5%. The authors in [143] proposed a DL method to study the relationship between genomic variations and traits. They analyzed 6083 sample exon sequencing files belonging to 12 cancer types. They used TCGA and 1000 Genomes Project. They performed both binary (cancer/healthy) and multiclass (12 classes) classification tasks using specific, total, and mixture models to achieve an accuracy of 97.47%, 70.08%, and 94.7% for specific, mixture, and total specific models for the identification of cancer. Owais et al. [144] deployed a DL-based classification framework for the diagnosis of gastrointestinal diseases from endoscopic images. They deployed two datasets that are publicly available: Kvasir dataset and Gastrolab dataset. They followed a 2-step process. The classification network predicts the disease type in the first step, and then in the second step, the retrieval part shows the relevant cases. They performed multiclass (37 classes) classification using DenseNet transfer learning architecture, LSTM architecture, PCA, and KNN methods to achieve a correct recognition rate of 96.19% on this task. The authors in [145] proposed a CNN-based DL architecture for the multiclass (4 classes) classification of acute lymphoblastic leukemia. They used stained bone marrow images achieving an accuracy of 97.78%. Kann et al. [146] deployed a 3D CNN model to identify nodal metastasis and tumor extranodal extension. Their dataset has 2875 CT samples, 124 samples for validation and 131 samples for testing. They used a series of rotations and flipping technique to augment the datasets while achieving an AUC of 0.91. The authors in [147] proposed a DL approach to study the limited sample training problem from holographic images. They studied the classification of healthy and cancer cell lines. They used Generative Adversarial Networks (GANs) as the data augmentation method to train a large number of unclassified samples from sperm cells. Their model achieved an accuracy of 99% for healthy/primary cancer/metastatic cancer multiclass classification problems.

Table 5 displays a summary of the studies for the task of classification of cancers covered in this subsection.

4.5. Classification, Segmentation, Prediction, and Detection of Brain Tumors. Sun et al. [148] proposed a 3D fully

Table 5: Summary of the studies for the classification of other types of cancer.

Publication	Type(s) of cancer	Type of data	Methods	Performance
[125]	Colorectal cancer	Custom dataset	VGG19, AlexNet, SqueezeNet version 1.1, GoogLeNet, ResNet-50	98.7% accuracy using VGG19 model
[126]	Colon cancer	Custom dataset	CNN model	AUC = 0.998, specificity = 99.7%, sensitivity = 100%
[127]	Gastric cancer	Custom datasets	DL models	Accuracy of 0.822 in the international validation cohort
[128]	Bladder cancer	TCGA+custom dataset	DL models	Accuracy (custom) = 75%
[129]	Lung cancer	TCGA+custom dataset	Weakly supervised DL algorithm	Accuracy of 97.3% on the custom dataset
[130]	Prostate cancer	Custom dataset	DL methods	Accuracy of 92% in cancerous/benign classification
[131]	Skin cancer	Custom dataset	DL algorithms	Positive predictive value of 59.9%
[132]	Skin cancer	HAM10000, Kaohsiung Chang Gung Memorial Hospital	Lightweight DL algorithm	Accuracy = 85.8% (HAM10000, multiclass classification)
[133]	Skin cancer	Custom dataset	Interpretable DL methods	Accuracies between 93.6% and 97.9%
[134]	Skin cancer	Custom dataset	CNN model	Accuracy of 82.95% for multiclass classification
[135]	Liver cancer	TCGA dataset	DL model	High accuracy (abnormal/normal classification)
[136]	Head and neck cancer	Custom dataset	Inception version 3	Mean AUC was 0.936 based on the testing set
[137]	Pancreatic cancer	Three custom datasets	CNN architectures	Accuracy of 0.986 for test set 2
[138]	Multiple	Custom datasets	ResNet-18, ResNet-34, ResNet-50	Accuracy of 94.90% for ResNet-50 architecture
[139]	Breast cancer	MITOS12, 2016 Tumor Proliferation Assessment Challenge	CNN architecture	F-measure of 0.79
[140]	Multiple	Custom dataset	Multiple instance learning-based DL system	Sensitivity = 100%
[141]	Blood and bone marrow cancer	Leukemia microarray gene data, Gene Expression Omnibus repository	Single-layer neural network, 3-layered deep network	96.67% for 3 layered model
[142]	Multiple	BioGPS data portal, TCIA, GDC dataset	Regression-based partitioned DL algorithm	Accuracy = 94.5%
[143]	Multiple	TCGA, 1000 Genomes Project	DL algorithms	Accuracy = 97.47%
[144]	Multiple	Kvasir dataset, Gastrolab	DL-based classification network	Accuracy = 96.19%
[145]	Multiple	Custom datasets	CNN model	Accuracy = 97.78%
[146]	Multiple	Custom dataset	3D CNN model	AUC = 0.91
[147]	Multiple	Custom dataset	GAN-based model	90–99% accuracies

convolutional network-based multipathway architecture to extract features from MRI images from the BRATS 2019 challenge for the segmentation of brain tumor regions. They used the concept of dilated convolutions in each pathway to achieve a dice score of 0.89, 0.78, and 0.76 for whole tumor (WT), tumor core (TC), and enhancing tumor (ET) on the BRATS 2019 challenge, respectively. They used cropping, random slicing, and z-score normalization as the preprocessing methods. The authors in [149] proposed a novel architecture combining U-Net encoding and decoding sub-architecture, dilated convolutional feature extracting layers, and a residual module. Their proposed architecture achieved a dice score of 0.843, 0.897, and 0.906 and 0.798, 0.902, and 0.845 on ET, WT, and TC brain tumor subregions on BRATS 2018 and BRATS 2019 challenges, respectively. They used normalization and cropping techniques to preprocess the images. Khan et al. [150] utilized VGG16 and VGG19 transfer learning-based CNN models, partial least square covariance matrix, discrete cosine transform, and extreme learning machine to extract and classify features on BRATS

2015, BRATS 2017, and BRATS 2018 challenge datasets to achieve an accuracy of 97.8%, 96.9%, and 92.5% for BRATS 2015, BRATS 2017, and BRATS 2018 datasets, respectively. To preprocess the images before feeding them to the classification model, the authors used the histogram equalization approach. Pei et al. [151] proposed a joint deep and machine learning-based model for classification, segmentation, and prediction of brain tumors. Using a context-aware CNN architecture for segmentation, 3D CNN architecture for classification, and LASSO for prediction, the authors achieved a dice score of 0.821, 0.895, and 0.835 for ET, WT, and TC regions, respectively, on BRATS 2019 for the segmentation task, an accuracy of 58.6% for the survival prediction task on the BRATS 2019 dataset, and balanced accuracy of 63.9% on the test set for the 2019 Computational Precision Medicine Radiology-Pathology (2019 CPM-RadPath) challenge. The authors in [152] proposed a resource-efficient CNN model integrating memory connections and an adaptive dense block for the segmentation of brain tumors. They used the BRATS 2015 challenge dataset for the validation of their model and z-score normalization as a preprocessing method achieving a dice coefficient score of 0.858, 0.816, and 0.818 for WT, TC, and ET subregions. Badža and Barjaktarović [153] present a 22-layered CNN architecture for brain tumor classification of T1-weighted MRI images belonging to three categories: meningioma, glioma, and pituitary tumor. They normalize and resize the scans to 256×256 pixels followed by 90° rotation and vertical flipping augmentation methods to synthetically increase the size of the dataset. The authors achieved an accuracy of 96.56% for the multiclass classification task on a custom dataset. The authors in [154] proposed a transfer learning-based approach for segmentation and classification of brain tumors using Inception version 3-based features. They concatenated the CNN-based features with local binary pattern- (LBP-) based features. Contrast improvement is used as a preprocessing method. The authors achieved a dice score of 0.8373, 0.937, and 0.7994 for TC, WT, and ET subregions on the BRATS 2017 dataset and a dice score of 0.8834, 0.912, and 0.8184 for TC, WT, and ET subregions on the BRATS 2018 dataset. For the classification task, the authors achieved an average accuracy upward of 92% on BRATS 2013, BRATS 2014, BRATS 2017, and BRATS 2018 datasets. Rai et al. [155] proposed a U-Net-based DL model using skip connections for the classification, segmentation, and detection of tumors in brain MRI scans. They conducted their experiments on 120 patients of lower-grade glioma in TCGA database with 1373 scans for patients and 2556 scans for normal controls. The authors deployed cropping, resizing, global pixel normalization, horizontal flipping, flipping and rotation, random rotation, shift scale rotate, transposition, blurring, Gaussian blurring, random gamma, random brightness, and normalization as preprocessing and data augmentation methods. Their model achieved an accuracy of 99.7% on the classification task, a dice score of 0.9573 on the segmentation task, and a Jaccard index of 0.86 on the detection task. The authors in [156] compared and contrasted the performance of different transfer learning architectures for the binary classification of brain tumors into benign and malignant categories. They chose AlexNet, GoogLeNet, ResNet-50, ResNet-101, and SqueezeNet architectures for comparison. They employed a dataset of 224 benign category and 472 malignant category T1-weighted MRI images acquired from the TCIA public access repository. They used resizing, flipping, mirroring, salt noise addition, and rotation as preprocessing and data augmentation methods to achieve an accuracy of 99.04% using an AlexNet-type architecture. Feng et al. [157] developed a 3D U-Net model for brain tumor segmentation. They picked up an ensemble of models to extract features from brain MRI images on the BRATS 2018 challenge for segmentation and survival prediction. The authors achieved a dice score of (0.7946, 0.9114, and 0.8304) on (ET, WT, and TC) subregions for the segmentation task and an accuracy of 32.1% on the survival prediction task. The authors in [158] proposed an ensemble of deep CNN architectures integrating two and three paths of parallel models in a single model. They used 2D slices of brain MRI images from the BRATS 2013 dataset achieving a dice score of (0.86, 0.86, and 0.88) on (WT, TC, and ET) subregions. As a preprocessing step, they standardized the slices using the zero mean and unit variance normalization procedure. Naser and Deen [159] proposed a DL approach combining U-Net architecture, VGG16 transfer learning architecture, and a fully connected architecture for classification and segmentation of brain MRI images into lower-grade gliomas belonging to 110 patients. They used normalization, cropping, resizing, padding, rescaling, rotation, zooming, shifting, shearing, and flipping as preprocessing and data augmentation methods. Their approach achieved a dice score of 0.84 on the segmentation task and accuracy, sensitivity, and specificity of 92% on the binary classification (grade II/grade III) task. The authors in [160] proposed a multiscale 3D CNN architecture for the recognition and segmentation of 220 high- and 54 low-grade glioma MRI scans from the BRATS 2015 challenge dataset. As a preprocessing method, the authors used histogram matching to ensure consistency among gray levels. Their model achieved a dice score of 0.89 on the segmentation task, a sensitivity of 0.89, and a specificity of 0.90 on the recognition task. Chang et al. [161] proposed a DL model combining average pooling and max pooling layers along with 1×1 kernels. They further combined this model with conditional random fields to optimize prediction results. The authors used the BRATS 2013 dataset to achieve a dice score of (0.80, 0.75, and 0.71) on (WT, TC, and ET) subregions. As a preprocessing method, the authors used an intensity normalization method. The authors in [162] proposed a multiscale CNN model for the categorization of an MRI scan into healthy, meningioma, glioma, and pituitary tumor categories. The authors used 2D MRI images acquired from local hospitals in China to conduct their experiments. They achieved a dice score of (0.894, 0.779, 0.813, and 0.828) on (meningioma, glioma, pituitary tumor, and average), respectively, and accuracy of 97.3% on the classification task. As preprocessing and data augmentation methods, the authors used pixel standardization and elastic transformation methods.

Table 6 displays a summary of the studies for the classification, segmentation, prediction, and detection of brain tumors covered in this subsection.

Table 6: Summary of the studies for the classification, segmentation, prediction, and detection of brain tumors.

Publication	Dataset(s)	Task(s)	Method	Performance
[148]	BRATS 2019	Segmentation	3D fully convolutional network-based multipathway architecture	Dice score of 0.89, 0.78, and 0.76 for WT, TC, and ET subregions, respectively
[149]	BRATS 2018 and BRATS 2019	Segmentation	Combination of U-Net encoding and decoding subarchitecture, dilated convolutional feature extracting layers, and a residual module	Dice score of 0.843, 0.897, and 0.906 and 0.798, 0.902, and 0.845 on ET, WT, and TC brain tumor subregions on BRATS 2018 and BRATS 2019 challenges, respectively
[150]	BRATS 2015, BRATS 2017, and BRATS 2018	Classification	VGG16 and VGG19 transfer learning-based CNN models, partial least square covariance matrix, discrete cosine transform, and extreme learning machine	Accuracy of 97.8%, 96.9%, and 92.5% for BRATS 2015, BRATS 2017, and BRATS 2018 datasets, respectively
[151]	BRATS 2019 and 2019 CPM-RadPath	Classification, segmentation, and prediction	Context-aware CNN architecture for segmentation, 3D CNN architecture for classification, and LASSO for prediction	Dice score of 0.821, 0.895, and 0.835 for ET, WT, and TC regions, respectively, on BRATS 2019 for segmentation task, accuracy of 58.6% for survival prediction task on BRATS 2019 dataset, and balanced accuracy of 63.9% on 2019 CPM-RadPath challenge
[152]	BRATS 2015	Segmentation	Resource-efficient CNN model with memory connections and an adaptive dense block	Dice coefficient score of 0.858, 0.816, and 0.818 for WT, TC, and ET subregions
[153]	Custom	Classification	22-layered CNN architecture	Accuracy of 96.56%
[154]	BRATS 2013, BRATS 2014, BRATS 2017, and BRATS 2018	Segmentation and classification	Inception version 3+LBP	Dice score of 0.8373, 0.937, and 0.7994 for TC, WT, and ET subregions on BRATS 2017; dice score of 0.8834, 0.912, and 0.8184 for TC, WT, and ET on BRATS 2018; average accuracy upward of 92% on BRATS 2013, BRATS 2014, BRATS 2017, and BRATS 2018 datasets
[155]	TCGA database	Classification, segmentation, and detection	U-Net-based DL model using skip connections	Accuracy of 99.7% on the classification task, dice score of 0.9573 on the segmentation task, and Jaccard index of 0.86 on the detection task
[156]	TCIA public access repository	Classification	AlexNet, GoogLeNet, ResNet-50, ResNet-101, and SqueezeNet	An accuracy of 99.04% using AlexNet-type architecture
[157]	BRATS 2018	Segmentation and prediction	3D U-Net model	Dice score of 0.7946, 0.9114, and 0.8304 on ET, WT, and TC, accuracy of 32.1%
[158]	BRATS 2013	Segmentation	Ensemble of deep CNN architectures	Dice score of 0.86, 0.86, and 0.88 on WT, TC, and ET
[159]	Custom	Segmentation and classification	U-Net architecture, VGG16 transfer learning architecture, and a fully connected architecture	Dice score of 0.84; accuracy, sensitivity, and specificity of 92% on the binary classification task
[160]	BRATS 2015	Segmentation and classification	Multiscale 3D CNN architecture	Dice score of 0.89, sensitivity of 0.89, and a specificity of 0.90
[161]	BRATS 2013	Segmentation	DL model combining average pooling and max pooling layers along with 1×1 kernels	Dice score of 0.80, 0.75, and 0.71 on WT, TC, and ET
[162]	Custom	Segmentation and classification	Multiscale Convolutional Neural Network	Dice score of 0.894, 0.779, 0.813, and 0.828 on meningioma, glioma, pituitary tumor, and average and an accuracy of 97.3%

5. Discussion

The dynamics of cancer growth with respect to time are difficult to estimate. Precise measures can be made largely at the end of the cycle in cancer's evolution, when it is detached from the body. Ongoing mutations provide a rich history of clonal lineages which lead to changes in both genotype and phenotype.

Psycho-oncology is a branch of oncology that deals directly with psychological and social issues. It deals with both emotional and psychobiological dimensions of cancer. However, there are still a number of obstacles in its wide adoption such as the dearth of medical practitioners as well as assessment tools and supporting instruments. It is important that both psychological and psychobiological factors influence the way cancers are treated. This domain must fulfill the demands for the availability of resources, support for caregivers and patients, and carving out new research directions for enthusiastic researchers [163, 164].

Research in AI has proven its worth in the support of medical decision-making. Due to the unknown nature of these algorithms, their widespread adoption is still limited. Explanatory AI algorithms provide a solution to this problem. However, performance issues might hinder their adoption as well. Robustness, local attribution, and completeness are three key properties of an explainable AI system. One way to get around this problem is to find strategies that optimally merge explainable and nonexplainable AI models. Some solutions that point in this direction are winning the confidence of clinicians by marking the regions in an image that are involved in AI predictions; another way is to attack or deceive the DL models through adversarial augmentations as it could potentially reveal the important features and discard the unimportant ones. There is a close link between interpretability and explainability. An explainable model is interpretable, but the reverse connection may not hold. A prediction relying on thousands of parameters is neither interpretable nor explainable [165, 166].

Precise DL model predictions are dependent on the availability of a large corpus of data (labelled or unlabelled), and it is a challenge to train it on a relatively small dataset. One way to look at this problem is through understanding the genetic evolution process. Gene transfer is the transfer of genetic information from a parent to its offspring. Genes encode genetic instructions (knowledge) from ancestors to descendants. The ancestors do not necessarily have better knowledge; yet, the evolution of knowledge across generations promotes a better learning curve for the descendants. There is a need for methods that can mimic this behaviour and use a limited number of examples to achieve their desired performance on different tasks [167]. Catastrophic forgetting is another problem limiting the performance of modern networks as they lack the ability to learn from continuous streams of data. The quality of the feature representation considerably determines the amount of forgetting. Boosting secondary information is the key to improving the transferability of features from old to new tasks without forgetting and is a promising direction for future work [168] especially for cancer diagnosis, prognosis, and prediction.

Despite the claims made by researchers, multiclass classification is an immensely difficult problem requiring a deeper understanding of human visual perception that moves beyond large datasets, and DL is perhaps necessary to solve many domain problems [169] including cancer diagnosis, prognosis, and prediction.

Another challenge that is worth mentioning is to find intricate hierarchical patterns from all forms of data such as labelled and unlabelled in a way that integrates information to perform visual inference. Unsupervised and semisupervised learning can help in this direction by offering potential solutions that help us in delving deeper into cancer pathogenesis and prediction tasks [170].

Can we use real-world images from another domain for calibration? Bridging the gap between cross-domain calibration and in-domain calibration is required to get optimal performance from neural networks. Techniques such as gram matrix similarity can be used as a criterion to select calibration datasets from a candidate pool to further improve performance [171]. This process can be used for effective feature construction in cancer diagnosis, prognosis, and prediction.

Modern DL object detection networks rely heavily on region proposal calculating algorithms to identify object locations. However, region proposal computation is a slow task. Faster region CNNs solve this problem by sharing convolutional layers with object detection subsystems. This process requires further research, and there is a need for improved computationally lightweight methods [25, 26]. Cancer lesion detection can be improved by doing thorough research in this domain.

Modern DL networks rely heavily on global image statistics. This reliance can cause problems for these systems as shape and texture recognition is often better done at the local rather than the global level. Research in this domain can lead to better network generalization [172] holding the potential to improve cancer diagnosis, prognosis, and prediction.

Mitigating gradient explosion or decay in RNN training based on pondering over informative inputs to strengthen their contribution in the hidden state and finding computationally efficient ways for this purpose by suppressing noise in inputs or imposing novel constraints is a problem worth investigating [173].

Image recognition and image generation are two cornerstones of computer vision. While both are burgeoning fields, specialized techniques from both subareas can sometimes form a dichotomy. Historically, the field of DL was widely popularized in discriminative image classification with AlexNet architecture and image generation through GANs and Variational Autoencoders (VAEs). Novel data augmentation methods that force a network to pay attention to the moments extracted by layers of a deep network are a need of time [174] and can improve the performance of models in cancer diagnosis, prognosis, and prediction.

Further research should also target the discovery of novel objects (such as those having an aberrant organization, rare tumor, and foreign bodies), interpretable DL models (using influence functions or an attention mechanism), intraoperative decision-making, and tumor-infiltrating immune cell analysis. Some problems such as the appearance of whole-slide image as orderless texture-like image and color variation and artefacts are potentially hindering the performance of DL techniques [175] for cancer diagnosis, prognosis, and prediction.

Different types of imaging modalities like mammography, CT, MRI, and ultrasound have helped in the staging of cancer especially breast cancer. These systems have helped medical practitioners in the early identification of breast

cancer [176]. For breast cancer, varying types of breast densities make masses very difficult to detect and classify in comparison to calcifications providing room for further research in this domain [177].

Other areas for potential research are scarcity of data, imbalanced datasets, missing data, and high dimensionality of patient data. Future work should be focused on testing and improving methods to achieve better performing DL models for cancer diagnosis, prognosis, and prediction tasks.

6. Conclusion

DL models have revolutionized the diagnosis and predictions of cancers. Data have been accepted in various forms and multiple sources. These models are excellent feature extractors, and their characteristics can improve cancer prognosis and prediction. Data augmentation is important for cancer diagnosis and prediction tasks to improve the final performance of systems. These methods will play a key role in making predictions about the cancer diagnosis and prediction tasks. However, further testing and validation are required on larger datasets for clinical applications. More research on data augmentation methods, learning in different domains such as frequency domain, and deploying novel architectures such as graph convolutional networks will likely improve their performance further.

Data Availability

No data were used to support this study.

Conflicts of Interest

No conflicts of interest exist between authors for the present study.

Authors' Contributions

A.B.T. conceived and designed the study. Y.K.M. and M.K.A.K. performed the analysis. All authors wrote and revised the draft manuscript.

Acknowledgments

We would like to acknowledge the group effort made in this research.

References

[1] P. S. Roy and B. J. Saikia, "Cancer and cure: a critical analysis," *Indian Journal of Cancer*, vol. 53, no. 3, pp. 441-442, 2016.

[2] L. A. Torre, R. L. Siegel, E. M. Ward, and A. Jemal, "Global cancer incidence and mortality rates and trends–an update," *Cancer Epidemiology Biomarkers & Prevention*, vol. 25, no. 1, pp. 16–27, 2016.

[3] F. Abbas-Aghababazadeh, Q. Mo, and B. L. Fridley, "Statistical genomics in rare cancer," *Seminars in Cancer Biology*, vol. 61, pp. 1–10, 2020.

[4] J. M. Baust, Y. Rabin, T. J. Polascik et al., "Defeating cancers' adaptive defensive strategies using thermal therapies: examining cancer's therapeutic resistance, ablative, and computational modeling strategies as a means for improving therapeutic outcome," *Technology in Cancer Research & Treatment*, vol. 17, 2018.

[5] R. Seelige, S. Searles, and J. D. Bui, "Innate sensing of cancer's non-immunologic hallmarks," *Current Opinion in Immunology*, vol. 50, pp. 1–8, 2018.

[6] A. V. Lichtenstein, "Genetic mosaicism and cancer: cause and effect," *Cancer Research*, vol. 78, no. 6, pp. 1375–1378, 2018.

[7] A. S. Balajee and M. P. Hande, "History and evolution of cytogenetic techniques: current and future applications in basic and clinical research," *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, vol. 836, Part A, pp. 3–12, 2018.

[8] S. Parida and D. Sharma, "The microbiome and cancer: creating friendly neighborhoods and removing the foes within," *Cancer Research*, vol. 81, no. 4, pp. 790–800, 2021.

[9] A. Dasgupta, M. Nomura, R. Shuck, and J. Yustein, "Cancer's Achilles' heel: Apoptosis and necroptosis to the rescue," *International Journal of Molecular Sciences*, vol. 18, no. 1, p. 23, 2017.

[10] V. Marx, "How to follow metabolic clues to find cancer's Achilles heel," *Nature Methods*, vol. 16, no. 3, pp. 221–224, 2019.

[11] T. Sepp, B. Ujvari, P. W. Ewald, F. Thomas, and M. Giraudeau, "Urban environment and cancer in wildlife: available evidence and future research avenues," *Proceedings of the Royal Society B: Biological Sciences*, vol. 286, no. 1894, article 20182434, 2019.

[12] C. Vrinten, L. M. McGregor, M. Heinrich et al., "What do people fear about cancer? A systematic review and meta-synthesis of cancer fears in the general population," *Psychooncology*, vol. 26, no. 8, pp. 1070–1079, 2017.

[13] M. Karamanou, E. Tzavellas, K. Laios, M. Koutsilieris, and G. Androutsos, "Melancholy as a risk factor for cancer: a historical overview," *JBUON*, vol. 21, no. 3, pp. 756–759, 2016.

[14] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, pp. 448–456, Lille, France, 2015.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017*, pp. 4700–4708, Honolulu, HI, USA, July 2017.

[17] M. T. Sadiq, H. Akbari, A. U. Rehman et al., "Exploiting feature selection and neural network techniques for identification of focal and nonfocal EEG signals in TQWT domain," *Journal of Healthcare Engineering*, vol. 2021, Article ID 6283900, 24 pages, 2021.

[18] M. Asif, W. U. Khan, H. M. R. Afzal et al., "Reduced-complexity LDPC decoding for next-generation IoT networks," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 2029560, 10 pages, 2021.

[19] A. R. Junejo, M. K. A. Kaabar, and S. Mohamed, "Future robust networks: current scenario and beyond for 6G," *IMCC Journal of Science*, vol. 11, no. 1, pp. 67–81, 2021.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*, pp. 1–14, San Diego, CA, USA, May 2015.

[21] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9, IEEE, Piscataway, NJ, USA, Boston, MA, USA, June 2015.

[22] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 1251–1258, Honolulu, HI, USA, July 2017.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[26] Z. Zhong, L. Sun, and Q. Huo, "An anchor-free region proposal network for Faster R-CNN-based text detection approaches," *International Journal on Document Analysis and Recognition*, vol. 22, no. 3, pp. 315–327, 2019.

[27] J. Donahue, Y. Jia, O. Vinyals et al., "DeCAF: a deep convolutional activation feature for generic visual recognition," in *Proceedings of the 31st International Conference on Machine Learning, ICML 2014*, pp. 647–655, Beijing, China, June 2014.

[28] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," 2018, https://arxiv.org/abs/1707.07012v4.

[29] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," 2020, https://arxiv.org/abs/1905.11946.

[30] P. Royston and D. G. Altman, "External validation of a Cox prognostic model: principles and methods," *BMC medical research methodology*, vol. 13, no. 1, p. 33, 2013.

[31] A. F. Connors, "A controlled trial to improve care for seriously III hospitalized patients," *JAMA*, vol. 274, no. 20, pp. 1591–1598, 1995.

[32] C. Curtis, S. P. Shah, S. F. Chin et al., "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups," *Nature*, vol. 486, no. 7403, pp. 346–352, 2012.

[33] M. Veta, Y. J. Heng, N. Stathonikos et al., "Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge," *Medical Image Analysis*, vol. 54, pp. 111–121, 2019.

[34] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[35] S. G. Armato, G. McLennan, L. Bidaut et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

[36] A. A. Setio, A. Traverso, T. de Bel et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.

[37] F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016.

[38] T. Araújo, G. Aresta, E. Castro et al., "Classification of breast cancer histology images using convolutional neural networks," *PLoS One*, vol. 12, no. 6, article e0177544, 2017.

[39] G. Litjens, P. Bandi, B. Ehteshami Bejnordi et al., "1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset," *GigaScience*, vol. 7, no. 6, 2018.

[40] B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.

[41] L. Roux, D. Racoceanu, N. Loménie et al., "Mitosis detection in breast cancer histological images an ICPR 2012 contest," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 8, 2013.

[42] A. Risueno, C. Fontanillo, M. E. Dinger, and J. De Las Rivas, "GATExplorer: genomic and transcriptomic explorer; mapping expression probes to gene loci, transcripts, exons and ncRNAs," *BMC Bioinformatics*, vol. 11, no. 1, p. 221, 2010.

[43] T. Barrett, D. B. Troup, S. E. Wilhite et al., "NCBI GEO: archive for functional genomics data sets–10 years on," *Nucleic Acids Research*, vol. 39, pp. D1005–D1010, 2011.

[44] C. Wu, X. Jin, G. Tsueng, C. Afrasiabi, and A. I. Su, "BioGPS: building your own mash-up of gene annotations and expression profiles," *Nucleic Acids Research*, vol. 44, no. D1, pp. D313–D316, 2016.

[45] O. Diaz, K. Kushibar, R. Osuala et al., "Data preparation for artificial intelligence in medical imaging: a comprehensive guide to open-access platforms and tools," *Physica Medica*, vol. 83, pp. 25–37, 2021.

[46] P. Sudmant, T. Rausch, E. Gardner et al., "An integrated map of structural variation in 2,504 human genomes," *Nature*, vol. 526, no. 7571, pp. 75–81, 2015.

[47] K. Pogorelov, K. Randel, C. Griwodz et al., "Kvasir: a multi-class image-dataset for computer aided gastrointestinal disease detection," in *ACM Multimedia Systems (MMSys'17)*, pp. 164–169, Taipei, Taiwan, 2017.

[48] E. Drelie Gelasca, J. Byun, B. Obara, and B. S. Manjunath, "Evaluation and benchmark for biological image segmentation," in *2008 15th IEEE International Conference on Image Processing*, pp. 1816–1819, San Diego, CA, USA, 2008.

[49] B. H. Menze, A. Jakab, S. Bauer et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[50] L. P. Petalidis, A. Oulas, M. Backlund et al., "Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data," *Molecular Cancer Therapeutics*, vol. 7, no. 5, pp. 1013–1024, 2008.

[51] C. L. Chi, W. N. Street, and W. H. Wolberg, "Application of artificial neural network-based survival analysis on two breast cancer datasets," in *AMIA Annual Symposium Proceedings 2007*, pp. 130–134, Chicago, IL, USA, 2007.

[52] Z. Huang, X. Zhan, S. Xiang et al., "SALMON: Survival Analysis Learning with Multi-Omics Neural Networks on breast cancer," *Frontiers in Genetics*, vol. 10, p. 166, 2019.

[53] H. Shimizu and K. I. Nakayama, "A 23 gene-based molecular prognostic score precisely predicts overall survival of breast cancer patients," *EBioMedicine*, vol. 46, pp. 150–159, 2019.

[54] R. Joshi and C. Reeves, "Beyond the Cox model: artificial neural networks for survival analysis part II," in *Proceedings of the Eighteenth International Conference on Systems Engineering*, pp. 179–184, Coventry, UK, May 2003.

[55] B. Jing, T. Zhang, Z. Wang et al., "A deep survival analysis method based on ranking," *Artificial Intelligence in Medicine*, vol. 98, pp. 1–9, 2019.

[56] J. Hao, Y. Kim, T.-K. Kim, and M. Kang, "PASNet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data," *BMC Bioinformatics*, vol. 19, no. 1, p. 510, 2018.

[57] D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 3, pp. 841–850, 2019.

[58] K. Chaudhary, O. B. Poirion, L. Lu, and L. X. Garmire, "Deep learning-based multi-omics integration robustly predicts survival in liver cancer," *Clinical Cancer Research*, vol. 24, 2018.

[59] D. Bychkov, N. Linder, R. Turkki et al., "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Scientific Reports*, vol. 8, no. 1, 2018.

[60] 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.

[61] T. Ching, X. Zhu, and L. X. Garmire, "Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data," *PLoS Computational Biology*, vol. 14, no. 4, 2018.

[62] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Medical Research Methodology*, vol. 18, no. 1, p. 24, 2018.

[63] P. Mobadersany, S. Yousefi, M. Amgad et al., "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences of the United Statetess of America*, vol. 115, no. 13, pp. E2970–E2979, 2018.

[64] 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.

[65] B. Liu, Y. Liu, X. Pan, M. Li, S. Yang, and S. C. Li, "DNA methylation markers for pan-cancer prediction by deep learning," *Genes*, vol. 10, no. 10, p. 778, 2019.

[66] M. H. Yap, G. Pons, J. Marti et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2018.

[67] M. H. Yap, M. Goyal, F. M. Osman et al., "Breast ultrasound lesions recognition: end-to-end deep learning approaches," *Journal of Medical Imaging*, vol. 6, pp. 1–7, 2019.

[68] Y. Liu, T. Kohlberger, M. Norouzi et al., "Artificial intelligence–based breast cancer nodal metastasis detection insights into the black box for pathologists," *Archives of Pathology & Laboratory Medicine*, vol. 143, no. 7, pp. 859–868, 2019.

[69] T. Mahmood, M. Arsalan, M. Owais, M. B. Lee, and K. R. Park, "Artificial intelligence-based mitosis detection in breast cancer histopathology images using Faster R-CNN and deep CNNs," *Journal of Clinical Medicine*, vol. 9, no. 3, p. 749, 2020.

[70] A. Anuranjeeta, K. K. Shukla, A. Tiwari, and S. Sharma, "Classification of histopathological images of breast cancerous and non cancerous cells based on morphological features," *Biomedical and Pharmacology Journal*, vol. 10, no. 1, pp. 353–366, 2017.

[71] D. Cai, X. Sun, N. Zhou, X. Han, and J. Yao, "Efficient mitosis detection in breast cancer histology images by RCNN," in *Proceedings of the International Symposium on Biomedical Imaging*, pp. 919–922, Venice, Italy, April 2019.

[72] J. Zhou, L. Y. Luo, Q. Dou et al., "Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images," *Journal of Magnetic Resonance Imaging*, vol. 50, no. 4, pp. 1144–1151, 2019.

[73] M. A. Al-Antari, M. A. Al-Masni, M. T. Choi, S. M. Han, and T. S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *International Journal of Medical Informatics*, vol. 117, pp. 44–54, 2018.

[74] N. Nasrullah, J. Sang, M. S. Alam, M. Mateen, B. Cai, and H. Hu, "Automated lung nodule detection and classification using deep learning combined with multiple strategies," *Sensors*, vol. 19, no. 17, p. 3722, 2019.

[75] O. Ozdemir, R. L. Russell, and A. A. Berlin, "A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1419–1429, 2020.

[76] E. Shkolyar, X. Jia, T. C. Chang et al., "Augmented bladder tumor detection using deep learning," *European Urology*, vol. 76, no. 6, pp. 714–718, 2019.

[77] C. Fourcade, L. Ferrer, G. Santini et al., "Combining superpixels and deep learning approaches to segment active organs in metastatic breast cancer PET images," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1536–1539, Montreal, QC, Canada, 2020.

[78] M. Zhang, G. S. Young, H. Chen et al., "Deep-learning detection of cancer metastases to the brain on MRI," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 4, pp. 1227–1236, 2020.

[79] J. Ma, S. Duan, Y. Zhang et al., "Efficient deep learning architecture for detection and recognition of thyroid nodules," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 1242781, 15 pages, 2020.

[80] A. Das, R. Acharya, S. S. Panda, and S. Sabut, "Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques," *Cognitive Systems Research*, vol. 54, pp. 165–175, 2019.

[81] S. Lal, D. Das, K. Alabhya, A. Kanfade, A. Kumar, and J. Kini, "NucleiSegNet: robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images," *Computers in Biology and Medicine*, vol. 128, article 104075, 2021.

[82] P. Wang and A. C. S. Chung, "DoubleU-Net: colorectal cancer diagnosis and gland instance segmentation with text-guided feature control," in *Computer Vision – ECCV 2020 Workshops*, pp. 338–354, Springer, 2020.

[83] J. Panic, A. Defeudis, S. Mazzetti et al., "A convolutional neural network based system for colorectal cancer segmentation on MRI images," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1675–1678, Montreal, QC, Canada, 2020.

[84] J. Juebin, H. Zhu, Z. Jindi et al., "Multiple U-Net-based automatic segmentations and radiomics feature stability on ultrasound images for patients with ovarian cancer," *Frontiers in Oncology*, vol. 10, p. 3428, 2021.

[85] T. Shibata, A. Teramoto, H. Yamada, N. Ohmiya, K. Saito, and H. Fujita, "Automated detection and segmentation of early gastric cancer from endoscopic images using mask R-CNN," *Applied Sciences*, vol. 10, no. 11, p. 3842, 2020.

[86] J. Wang and X. Liu, "Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network," *Computer Methods and Programs in Biomedicine*, vol. 207, article 106210, 2021.

[87] S. Shrestha, A. Alsadoon, P. W. C. Prasad, I. Seher, and O. H. Alsadoon, "A novel solution of using deep learning for prostate cancer segmentation: enhanced batch normalization," *Multimedia Tools and Applications*, vol. 80, no. 14, pp. 21293–21313, 2021.

[88] Z. Liu, C. Yang, J. Huang, S. Liu, Y. Zhuo, and X. Lu, "Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer," *Future Generation Computer Systems*, vol. 114, pp. 358–367, 2021.

[89] T. Nemoto, N. Futakami, M. Yagi et al., "Simple low-cost approaches to semantic segmentation in radiation therapy planning for prostate cancer using deep learning with non-contrast planning CT images," *Physica Medica*, vol. 78, pp. 93–100, 2020.

[90] Y. Liang, D. Schott, Y. Zhang et al., "Auto-segmentation of pancreatic tumor in multi-parametric MRI using deep convolutional neural networks," *Radiotherapy and Oncology*, vol. 145, pp. 193–200, 2020.

[91] X. Chen, Z. Chen, J. Li, Y.-D. Zhang, X. Lin, and X. Qian, "Model-driven deep learning method for pancreatic cancer segmentation based on spiral-transformation," *IEEE Transactions on Medical Imaging*, p. 1, 2021.

[92] B. Huynh, K. Drukker, and M. Giger, "MO-DE-207B-06: computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks," *Medical Physics*, vol. 43, pp. 3705–3705, 2016.

[93] M. Byra, M. Galperin, H. Ojeda-Fournier et al., "Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion," *Medical Physics*, vol. 46, no. 2, pp. 746–755, 2019.

[94] M. Byra, T. Sznajder, D. Korzinek et al., "Impact of ultrasound image reconstruction method on breast lesion classification with deep learning," 2018, https://arxiv.org/abs/1804.02119.

[95] A. Hijab, M. A. Rushdi, M. M. Gomaa, and A. Eldeib, "Breast cancer classification in ultrasound images using transfer learning," in *Proceedings of the 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME)*, pp. 1–4, Tripoli, Lebanon, October 2019.

[96] O. Hadad, R. Bakalo, R. Ben-Ari, S. Hashoul, and G. Amit, "Classification of breast lesions using cross-modal deep learning," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 109–112, Melbourne, VIC, Australia, April 2017.

[97] D. S. Salem, R. M. Kamal, S. M. Mansour, L. A. Salah, and R. Wessam, "Breast imaging in the young: the role of magnetic resonance imaging in breast cancer screening, diagnosis and follow-up," *Journal of Thoracic Disease*, vol. 5, pp. 9–18, 2013.

[98] Q. Hu, H. M. Whitney, and M. L. Giger, "A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI," *Scientific Reports*, vol. 10, 2020.

[99] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 605–617, 2019.

[100] N. Bayramoglu, J. Kannala, and J. Heikkila, "Deep learning for magnification independent breast cancer histopathology image classification," in *Proceedings of the International Conference on Pattern Recognition*, pp. 2440–2445, Cancun, Mexico, December 2016.

[101] D. M. Vo, N. Q. Nguyen, and S. W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Information Sciences*, vol. 482, pp. 123–138, 2019.

[102] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Classification of histopathological biopsy images using ensemble of deep learning networks," 2019, https://arxiv.org/abs/1909.11870.

[103] G. Murtaza, L. Shuib, G. Mujtaba, and G. Raza, "Breast cancer multi-classification through deep neural network and hierarchical classification approach," *Multimedia Tools and Applications*, vol. 79, no. 21-22, pp. 15481–15511, 2020.

[104] M. Toğaçar, K. B. Özkurt, B. Ergen, and Z. Cömert, "BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer," *Physica A: Statistical Mechanics and its Applications*, vol. 545, article 123592, 2020.

[105] S. Alkassar, B. A. Jebur, M. A. M. Abdullah, J. H. Al-Khalidy, and J. A. Chambers, "Going deeper: magnification-invariant approach for breast cancer classification using histopathological images," *IET Computer Vision*, vol. 15, no. 2, pp. 151–164, 2021.

[106] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 2560–2567, Vancouver, BC, Canada, July 2016.

[107] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Scientific Reports*, vol. 7, no. 1, 2017.

[108] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018.

[109] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, "MuDeRN: multi-category classification of breast histopathological image using deep residual networks," *Artificial Intelligence in Medicine*, vol. 88, pp. 14–24, 2018.

[110] Ü. Budak, Z. Cömert, Z. N. Rashid, A. Sengür, and M. Çıbuk, "Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images," *Applied Soft Computing*, vol. 85, article 105765, 2019.

[111] K. George, S. Faziludeen, P. Sankaran, and J. K. Paul, "Deep learned nucleus features for breast cancer histopathological image analysis based on belief theoretical classifier fusion," in *Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON*, pp. 344–349, Kochi, India, October 2019.

[112] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Systems with Applications*, vol. 117, pp. 103–111, 2019.

[113] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *Proceedings of the 15th International Conference, ICIAR 2018*, pp. 737–744, Póvoa de Varzim, Portugal,, June 2018.

[114] Y. Li, J. Wu, and Q. Wu, "Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning," *IEEE Access*, vol. 7, pp. 21400–21408, 2019.

[115] Shallu and R. Mehra, "Breast cancer histology images classification: training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, 2018.

[116] K. Wadkar, P. Pathak, and N. Wagh, "Breast cancer detection using ANN network and performance analysis with SVM," *International Journal of Computer Engineering and Technology*, vol. 10, no. 3, pp. 75–86, 2019.

[117] V. Bevilacqua, A. Brunetti, A. Guerriero, G. F. Trotta, M. Telegrafo, and M. Moschetta, "A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images," *Cognitive Systems Research*, vol. 53, pp. 3–19, 2019.

[118] S. Sharma and R. Mehra, "Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight," *Journal of Digital Imaging*, vol. 33, no. 3, pp. 632–654, 2020.

[119] F. A. Spanhol, P. R. Cavalin, L. S. Oliveira, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, pp. 1868–1873, Banff, AB, Canada, October 2017.

[120] J. Li, J. Zhang, Q. Sun et al., "Breast cancer histopathological image classification based on deep second-order pooling network," in *Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, Glasgow, UK, July 2020.

[121] S. U. Khan, N. Islam, Z. Jan, I. Ud Din, and J. J. P. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, vol. 125, pp. 1–6, 2019.

[122] X. Zhang, Y. Zhang, Q. Zhang et al., "Extracting comprehensive clinical information for breast cancer using deep learning methods," *International Journal of Medical Informatics*, vol. 132, article 103985, 2019.

[123] N. Naik, A. Madani, A. Esteva et al., "Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains," *Nature Communications*, vol. 11, no. 1, p. 5727, 2020.

[124] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, article 12495, 2019.

[125] 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.

[126] Y. Zeng, S. Xu, W. C. Chapman Jr. et al., "Real-time colorectal cancer diagnosis using PR-OCT with deep learning," *Theranostics*, vol. 10, no. 6, pp. 2587–2596, 2020.

[127] D. Dong, M. J. Fang, L. Tang et al., "Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study," *Annals of Oncology*, vol. 31, no. 7, pp. 912–920, 2020.

[128] A. C. Woerl, M. Eckstein, J. Geiger et al., "Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides," *European Urology*, vol. 78, no. 2, pp. 256–264, 2020.

[129] X. Wang, H. Chen, C. Gan et al., "Weakly supervised deep learning for whole slide lung cancer image analysis," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3950–3962, 2020.

[130] D. Karimi, G. Nir, L. Fazli, P. C. Black, L. Goldenberg, and S. E. Salcudean, "Deep learning-based gleason grading of prostate cancer from histopathology images—role of multiscale decision aggregation and data augmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1413–1426, 2020.

[131] A. Dascalu and E. O. David, "Skin cancer detection by deep learning and sound analysis algorithms: a prospective clinical study of an elementary dermoscope," *EBioMedicine*, vol. 43, pp. 107–113, 2019.

[132] H. W. Huang, B. W. Hsu, C. H. Lee, and V. S. Tseng, "Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers," *The Journal of Dermatology*, vol. 48, no. 3, pp. 310–316, 2021.

[133] S. M. Thomas, J. G. Lefevre, G. Baxter, and N. A. Hamilton, "Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer," *Medical Image Analysis*, vol. 68, article 101915, 2021.

[134] A. Hekler, J. S. Utikal, A. H. Enk et al., "Superior skin cancer classification by the combination of human and artificial intelligence," *European Journal of Cancer*, vol. 120, pp. 114–121, 2019.

[135] C. Sun, A. Xu, D. Liu, Z. Xiong, F. Zhao, and W. Ding, "Deep learning-based classification of liver cancer histopathology images using only global labels," *IEEE journal of biomedical and health informatics*, vol. 24, no. 6, pp. 1643–1651, 2020.

[136] S. Diao, J. Hou, H. Yu et al., "Computer-aided pathologic diagnosis of nasopharyngeal carcinoma based on deep learning," *American Journal of Pathology*, vol. 190, no. 8, pp. 1691–1700, 2020.

[137] K. L. Liu, T. Wu, P. T. Chen et al., "Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation," *The Lancet Digital Health*, vol. 2, no. 6, pp. e303–e313, 2020.

[138] P. Korfiatis, T. L. Kline, D. H. Lachance, I. F. Parney, J. C. Buckner, and B. J. Erickson, "Residual deep convolutional

neural network predicts mgmt methylation status," *Journal of Digital Imaging*, vol. 30, no. 5, pp. 622–628, 2017.

[139] N. Wahab, A. Khan, and Y. S. Lee, "Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection," *Computers in Biology and Medicine*, vol. 85, pp. 86–97, 2017.

[140] G. Campanella, M. G. Hanna, L. Geneslaw et al., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.

[141] E. Nazari, A. H. Farzin, M. Aghemiri, A. Avan, M. Tara, and H. Tabesh, "Deep learning for acute myeloid leukemia diagnosis," *Journal of Medicine and Life*, vol. 13, no. 3, pp. 382–387, 2020.

[142] P. R. Jeyaraj and E. R. Samuel Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *Journal of Cancer Research and Clinical Oncology*, vol. 145, no. 4, pp. 829–837, 2019.

[143] Y. Sun, S. Zhu, K. Ma et al., "Identification of 12 cancer types through genome deep learning," *Scientific Reports*, vol. 9, no. 1, article 17256, 2019.

[144] M. Owais, M. Arsalan, T. Mahmood, J. K. Kang, and K. R. Park, "Automated diagnosis of various gastrointestinal lesions using a deep learning-based classification and retrieval framework with a large endoscopic database: model development and validation," *Journal of Medical Internet Research*, vol. 22, no. 11, article e18563, 2020.

[145] A. Rehman, N. Abbas, T. Saba, S. I. U. Rahman, Z. Mehmood, and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," *Microscopy Research and Technique*, vol. 81, no. 11, pp. 1310–1317, 2018.

[146] B. H. Kann, S. Aneja, G. V. Loganadane et al., "Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks," *Scientific Reports*, vol. 8, no. 1, article 14036, 2018.

[147] M. Rubin, O. Stein, N. A. Turko et al., "TOP-GAN: stain-free cancer cell classification using deep learning with a small training set," *Medical Image Analysis*, vol. 57, pp. 176–185, 2019.

[148] J. Sun, Y. Peng, Y. Guo, and D. Li, "Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN," *Neurocomputing*, vol. 423, pp. 34–45, 2021.

[149] J. Wang, J. Gao, J. Ren et al., "DFP-ResUNet:convolutional neural network with a dilated convolutional feature pyramid for multimodal brain tumor segmentation," *Computer Methods and Programs in Biomedicine*, vol. 208, p. 106208, 2021.

[150] M. A. Khan, I. Ashraf, M. Alhaisoni et al., "Multimodal brain tumor classification using deep learning and robust feature selection: a machine learning application for radiologists," *Diagnostics*, vol. 10, no. 8, p. 565, 2020.

[151] L. Pei, L. Vidyaratne, M. M. Rahman, and K. M. Iftekharuddin, "Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images," *Scientific Reports*, vol. 10, no. 1, article 19726, 2020.

[152] 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.

[153] M. M. Badža and M. Č. Barjaktarović, "Classification of brain tumors from MRI images using a convolutional neural network," *Applied Sciences*, vol. 10, no. 6, p. 1999, 2020.

[154] M. I. Sharif, J. P. Li, M. A. Khan, and M. A. Saleem, "Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images," *Pattern Recognition Letters*, vol. 129, pp. 181–189, 2020.

[155] H. M. Rai, K. Chatterjee, and S. Dashkevich, "Automatic and accurate abnormality detection from brain MR images using a novel hybrid UnetResNext-50 deep CNN model," *Biomedical Signal Processing and Control*, vol. 66, article 102477, 2021.

[156] R. Mehrotra, R. A. Ansari, R. Agrawal, and R. S. Anand, "A transfer learning approach for AI-based classification of brain tumors," *Machine Learning with Applications*, vol. 2, article 100003, 2020.

[157] X. Feng, N. J. Tustison, S. H. Patel, and C. H. Meyer, "Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features," *Frontiers in Computational Neuroscience*, vol. 14, 2020.

[158] S. Sajid, S. Hussain, and A. Sarwar, "Brain tumor detection and segmentation in MR images using deep learning," *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9249–9261, 2019.

[159] M. A. Naser and M. J. Deen, "Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images," *Computers in Biology and Medicine*, vol. 121, p. 103758, 2020.

[160] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "Computer-aided grading of gliomas combining automatic segmentation and radiomics," *International Journal of Biomedical Imaging*, vol. 2018, Article ID 2512037, 11 pages, 2018.

[161] 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.

[162] F. J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, and D. González-Ortega, "A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network," *Healthcare*, vol. 9, no. 2, p. 153, 2021.

[163] T. A. Graham and A. Sottoriva, "Measuring cancer evolution from the genome," *The Journal of Pathology*, vol. 241, no. 2, pp. 183–191, 2017.

[164] J. C. Holland, "Psycho-oncology: overview, obstacles and opportunities," *Psycho-oncology*, vol. 27, no. 5, pp. 1364–1376, 2018.

[165] G. Quellec, H. Al Hajj, M. Lamard, P. H. Conze, P. Massin, and B. Cochener, "ExplAIn: explanatory artificial intelligence for diabetic retinopathy diagnosis," *Medical Image Analysis*, vol. 72, article 102118, 2021.

[166] A. Rahnama and A. Tseng, "An adversarial approach for explaining the predictions of deep neural networks," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3253–3262, Nashville, TN, USA, 2021.

[167] A. Taha, A. Shrivastava, and L. Davis, "Knowledge evolution in neural networks," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 12843–12852, Nashville, TN, USA, 2021.

[168] S. Mittal, S. Galesso, and T. Brox, "Essentials for class incremental learning," in *IEEE International Conference on*

Computer Vision and Pattern Recognition, pp. 3513–3522, Nashville, TN, USA, 2021.

[169] H. Ahmed, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, "Object classification from randomized EEG trials," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3845–3854, Nashville, TN, USA, 2021.

[170] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of the ISCAS 2010—2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems*, pp. 253–256, Paris, France, May 2010.

[171] H. Yu, L. Yang, and H. Shi, "Is in-domain data really needed? A pilot study on cross-domain calibration for network quantization," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3043–3052, Nashville, TN, USA, 2021.

[172] O. Nuriel, S. Benaim, and L. Wolf, "Permuted AdaIN: reducing the bias towards global statistics in image classification," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 9482–9491, Nashville, TN, USA, 2021.

[173] A. Kag and V. Saligrama, "Time adaptive recurrent neural network," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 15149–15158, Nashville, TN, USA, 2021.

[174] B. Li, F. Wu, S.-N. Lim et al., "On feature normalization and data augmentation," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 12383–12392, Nashville, TN, USA, 2021.

[175] D. Komura and S. Ishikawa, "Machine learning methods for histopathological image analysis," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 34–42, 2018.

[176] C. Kaushal, S. Bhat, D. Koundal, and A. Singla, "Recent trends in computer assisted diagnosis (cad) system for breast cancer diagnosis using histopathological images," *IRBM*, vol. 40, no. 4, pp. 211–227, 2019.

[177] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: systematic review," *Journal of Medical Internet Research*, vol. 21, no. 7, article e14464, 2019.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 3440778, 9 pages
https://doi.org/10.1155/2021/3440778

Research Article

Sentiment Analysis Based on the Nursing Notes on In-Hospital 28-Day Mortality of Sepsis Patients Utilizing the MIMIC-III Database

Qiaoyan Gao,[1] **Dandan Wang**,[1] **Pingping Sun**,[1] **Xiaorong Luan**,[2] **and Wenfeng Wang**[3,4,5]

[1]*Nursing Department, Weihai Central Hospital, Weihai, 264400 Shandong, China*
[2]*Nursing Department, Qilu Hospital of Shandong University, Jinan, 250012 Shandong, China*
[3]*School of Science, Shanghai Institute of Technology, Shanghai 201418, China*
[4]*International Academy of Visual Art and Engineering, London E16 1AH, UK*
[5]*Interscience Institute of Management and Technology, Bhubaneswar 752054, India*

Correspondence should be addressed to Xiaorong Luan; luanxrdoctor@outlook.com

Received 8 June 2021; Accepted 24 September 2021; Published 13 October 2021

Academic Editor: Pan Zheng

In medical visualization, nursing notes contain rich information about a patient's pathological condition. However, they are not widely used in the prediction of clinical outcomes. With advances in the processing of natural language, information begins to be extracted from large-scale unstructured data like nursing notes. This study extracted sentiment information in nursing notes and explored its association with in-hospital 28-day mortality in sepsis patients. The data of patients and nursing notes were extracted from the MIMIC-III database. A COX proportional hazard model was used to analyze the relationship between sentiment scores in nursing notes and in-hospital 28-day mortality. Based on the COX model, the individual prognostic index (PI) was calculated, and then, survival was analyzed. Among eligible 1851 sepsis patients, 580 cases suffered from in-hospital 28-day mortality (dead group), while 1271 survived (survived group). Significant differences were shown between two groups in sentiment polarity, Simplified Acute Physiology Score II (SAPS-II) score, age, and intensive care unit (ICU) type (all $P < 0.001$). Multivariate COX analysis exhibited that sentiment polarity (HR: 0.499, 95% CI: 0.409-0.610, $P < 0.001$) and sentiment subjectivity (HR: 0.710, 95% CI: 0.559-0.902, $P = 0.005$) were inversely associated with in-hospital 28-day mortality, while the SAPS-II score (HR: 1.034, 95% CI: 1.029-1.040, $P < 0.001$) was positively correlated with in-hospital 28-day mortality. The median death time of patients with $PI \geq 0.561$ was significantly earlier than that of patients with $PI < 0.561$ (13.5 *vs.* 49.8 days, $P < 0.001$). In conclusion, sentiments in nursing notes are associated with the in-hospital 28-day mortality and survival of sepsis patients.

1. Introduction

Sepsis, a syndrome of life-threatening physiologic, pathologic, and biochemical dysfunction due to uncontrolled responses to infection, is one of the leading causes of deaths in intensive care units (ICUs) [1]. Despite advances in care, sepsis remains among the costliest diseases, approximately accounting for over 20 billion (5.2%) of total United States (US) hospital costs [2]. In the US, admission for sepsis has overtaken that for stroke and myocardial infarction [3]. According to statistics, the prevalence of sepsis is up to 535 cases per 100 100,000 person-years and on the rise [4]. Population-level epidemiological data show that there are 31.5 million cases of sepsis and 19.4 million cases of severe sepsis worldwide, with 5.3 million potential deaths each year [5], and the in-hospital mortality reaches up to 25%-30% [6].

Currently, severity of illness scores (SOI) is usually used to predict mortality in ICUs. The SOI system is established

according to the coded data of patients' demographics, vital signs, and laboratory results usually accessed from the electronic health records, but there also exist unstructured data in the electronic health records, such as clinical notes written by clinicians which are not frequently used for predicting mortality [7]. Studies have demonstrated that clinicians can properly predict mortality in ICUs [8, 9]. Thus, their notes may provide some important information for patients' health status assessment. A previous study showed that the sentiment of clinicians towards patients could be evaluated by sentiment analysis, a method to classify the subjective properties of written text [10]. Sentiments measured in clinical notes are different according to demographic features and clinical outcomes [10]. There are studies suggesting that sentiments measured in clinical notes are associated with hospital readmission and mortality [11, 12].

In this study, we investigated the association of sentiments in nursing notes with the in-hospital 28-day mortality of sepsis patients based on the Medical Information Mart for Intensive Care (MIMIC-III) database, a freely accessible critical care database, aimed at providing some evidence for the improvement of patients' outcomes in ICUs.

2. Methods

2.1. Study Population. The data of patients and nursing notes were accessed from the MIMIC-III database developed by the MIT Lab for computational physiology. As an openly available dataset, MIMIC-III contains deidentified health data related to approximately 60,000 ICU admissions, including demographics, laboratory tests, medications, vital signs, transcribed nursing notes, diagnostic and procedure codes, fluid balance, length of stay, survival data, and others [13]. The inclusion criteria of this study were as follows: (1) patients diagnosed with sepsis, severe sepsis, and septic shock (International Classification of Diseases 9 (ICD-9) codes: 99591, 99592, and 78552) in the MIMIC-III database and (2) 15 years old or above at hospital admission. The exclusion criteria were as follows: (1) notes identified by physicians as errors, (2) notes written less than 12 hours before the time of death, and (3) patients without any data of nursing notes.

The data used in this study were obtained from the MIMIC-III database (https://mimic.physionet.org/), an openly available dataset. The data collection in the MIMIC-III was approved by the Ethics Review Board of the Beth Israel Deaconess Medical Center, and all private information has been desensitized.

2.2. Sentiment Analysis. Two techniques (syntactic and sematic) are mainly used to classify and compute the sentiment polarity in text [14]. A semantic approach means that the sentiment is extracted based on text meaning and is commonly obtained using a classifier [14]. To make inferences based on text structural features, this study employed a syntactic technique to extract sentiments.

Both the Python programming language and TextBlob natural language processing library were adopted to compute sentiment scores for the nursing notes [15]. The sentiment of text strings was computed using the pattern module in TextBlob. The pattern comprised a lexicon for various English language adverbs and adjectives able to be mapped to three dimensions of sentiment scores: polarity, subjectivity, and intensity [16]. The sentiment polarity was returned using TextBlob with a score from -1 to 1, and the sentiment subjectivity was returned with a score from 0 to 1. Higher scores showed more positive, subjective sentiments. In this study, both the polarity score and subjectivity score were assigned for each nursing note, and the scores were computed through establishment of a TextBlob object initialized with nursing note strings and extraction of sentiment attributes from the object [7]. The mean scores of sentiment polarity and subjectivity in nursing notes written during hospitalization were calculated for the first hospital admission of each patient and then used as predictors in the model of this study. For an example of sentiment polarity scores using TextBlob, see Table 1.

2.3. Mortality and Survival Assessment. As a common predictor of ICU mortality, Simplified Acute Physiology Score II (SAPS-II) is a composite score, including 17 variables (age, 12 physiology variables, type of admission, and 3 underlying disease variables) [17]. In this study, the SAPS-II score was calculated by the data from the MIMIC-III database and SQL scripts in the MIT Lab for computational physiology git repository. Additionally, gender and ICU type were also enrolled as variables because they were freely accessed from the MIMIC-III database, but not involved in SAPS-II. Survival was defined as the number of days from hospital admission to death or right-censoring time.

2.4. Statistical Analysis. Statistical analysis was performed using SPSS 22.0 software (IBM Corp., Armonk, NY, USA) and Python text analysis (version 3.7). Normally distributed data were compared by the t-test and manifested as mean $\pm$ standard deviation ($\bar{x} \pm s$); abnormally distributed data were compared with the Mann-Whitney U rank-sum test and presented as median and quartile (M (Q1, Q3)). Enumeration data were compared by the χ^2 test, with n (%) as the manifestation. The COX proportional hazard model was used to analyze the relationship between sentiment scores in nursing notes and the in-hospital 28-day mortality of sepsis patients. The size power of our study was 0.858.

The common type of the COX model was $h(t) = h0(t)$ $\exp(X'\beta)$, in which $h0(t)$ and $h(t)$ represented the datum risk function and the risk function at t time point, respectively, X was the covariate vector quantity, and β was the unknown vector quantity of the regression coefficient. The formula of the individual prognostic index (PI) was $PI = X1\beta1 + X2\beta2 + \cdots + Xk\beta k$. Based on the COX model, the individual PI was calculated. The greater the individual PI, the worse the prognosis. The survival curves were compared using a log-rank test. Box plot, histogram, and forest plot in our study were plotted with Python software. The power analysis was carried out to assess the statistical power $(1 - \beta)$ using PASS 15.0 software (NCSS, LLC). The results showed that the power values of the sentiment polarity score and sentiment subjectivity score were all 1.000. It was indicated that our findings performed well reliability. A significant difference was shown at $P < 0.05$.

Table 1: An example of sentiment polarity scores using TextBlob.

Excerpt text with deidentified patient information	Polarity score	Subjectivity score
New pt from ED,reffered from [∗∗Hospital3 1589∗∗],kc/o ESRD with HD, chronic diarrhea,c-diff,MRSA,presented with low BP treated with antibiotics,levophed,and dopamine,and transffered to [∗∗Hospital1 54∗∗] ED for further management,received 3lit FB IN ED,off dopa,on levophed and transffered to [∗∗Hospital Unit Name 44∗∗] for further management.VSS. ON NRBM sats 99%.having multiple sores on both legs and arms,amputated most of the fingers and toes. Pt not following with commands,not talking. Unable to assess orientation.	0.017	0.357
STATUS D: INTUBATED ON PROPOFOL GTT FOR COMFORT..FOLLOWS COMMANDS MOVES ALL EXTREM'S..K+'S DOWN A: PROPOFOL WEANED OFF & PT EXTUBATED WITHOUT PROB..ORAL NG DC'D.. SAT'S GOOD ON OFM @ 40%..STRONG COUGH..OOB TO CHAIR WITH 2 ASSISTS TOL WELL..A-LINE DC'D..K+ REPLETED..GOOD HUO..ABD SOFT WITH + BS'S R: STABLE P: [∗∗Month (only) 83∗∗] BE TRANSFERED TO FLOOR IF NEED BED PER DR [∗∗First Name (STitle) 349∗∗]..MONITOR K+'S & REPLETE AS NEEDED..LABS PER HO..CONTINUE WITH GOOD PULMONARY TOILET START CL LIQ'S AS TOL	0.101	0.522
See careview data and transfer note. Neuro: A&O x3. CV: Afib with controlled ventricular response, sbp 100-130's/50's. Pulm: RA sats 94-97%, lungs clear, pt occ productive of thick blood tinged sputum. GU: Uo 25-120cc hr/clr yellow, foley catheter dc'd at noon. GI: Taking house diet without difficulty. P: Chest CT today/?mass, pt to transfer to medical floor with goal to discharge home.	-0.050	0.215

Table 2: Baseline characteristics of survived and dead groups (n (%)).

Variables	Survived group ($n = 1271$)	Dead group ($n = 580$)	Total	t/χ^2	P
Sentiment polarity score ($\bar{x} \pm s$)	0.051 ± 0.038	0.026 ± 0.040	0.043 ± 0.040	12.47	<0.001
Sentiment subjectivity score ($\bar{x} \pm s$)	0.377 ± 0.036	0.376 ± 0.040	0.377 ± 0.037	0.95	0.340
SAPS-II score ($\bar{x} \pm s$)	43.63 ± 15.51	57.27 ± 15.77	47.906 ± 16.825	17.44	<0.001
Age (years)				24.45	<0.001
<40	92 (7.20)	23 (4.00)	115 (6.20)		
40-59	330 (26.00)	143 (24.70)	473 (25.60)		
60-69	222 (17.50)	126 (21.70)	348 (18.80)		
70-74	152 (12.00)	53 (9.10)	205 (11.10)		
75-79	172 (13.50)	57 (9.80)	229 (12.40)		
≥80	303 (23.80)	178 (30.70)	481 (26.00)		
Gender				0.096	0.757
Female	573 (45.10)	257 (44.30)	830 (44.80)		
Male	698 (54.90)	323 (31.60)	1021 (55.20)		
ICU type				25.36	<0.001
CCU	51 (4.00)	48 (8.30)	99 (5.30)		
CSRU	36 (2.80)	25 (4.30)	61 (3.30)		
MICU	738 (58.10)	352 (60.70)	1 090 (58.90)		
SICU	284 (22.30)	96 (16.60)	380 (20.50)		
TSICU	162 (12.70)	59 (10.20)	221 (11.90)		

SAPS-II: Simplified Acute Physiology Score II; ICU: intensive care unit; CCU: coronary care unit; CSRU: cardiac surgery recovery unit; MICU: medical intensive care unit; SICU surgical intensive care unit; TSICU: trauma/surgical intensive care unit.

3. Results

3.1. Baseline Characteristics of the Study Population. In the MIMIC-III database, there were a total of 3567 patients admitted to the ICU. Of these patients, 1128 patients without sepsis, 356 cases lacking sentiment polarity and subjectivity scores, 172 with missing SAPS-II, and 60 with missing survival data were excluded. Totally, 1851 sepsis

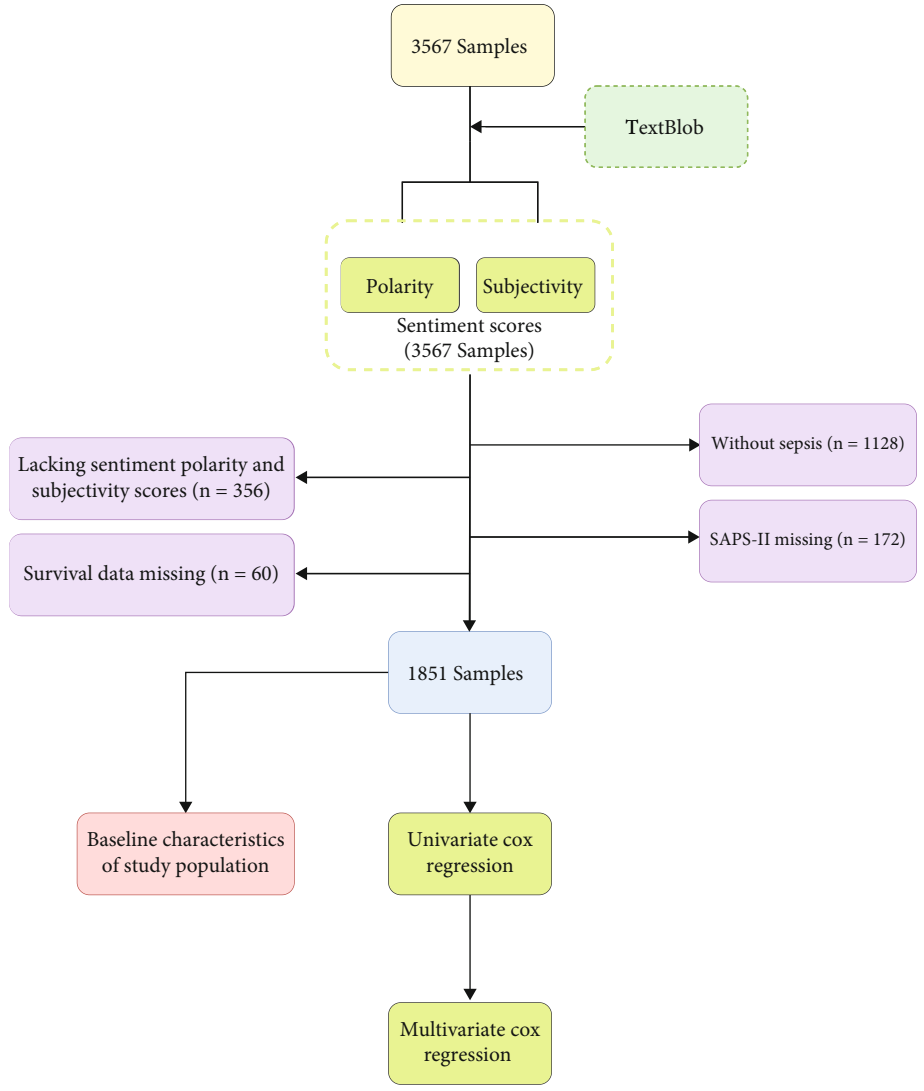


Figure 1: Flow diagram of patient screening from the MIMIC-III database.

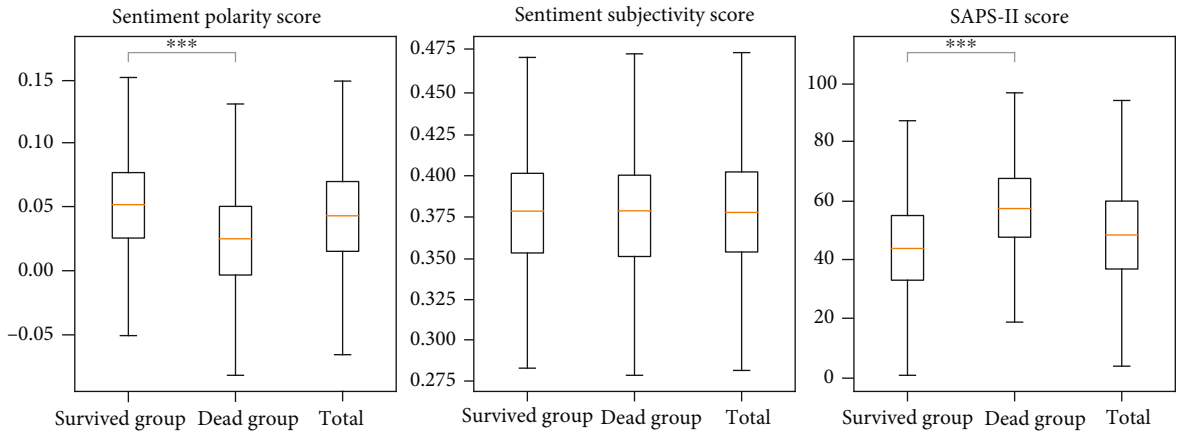


Figure 2: Box plot of the score of sentiment polarity, sentiment subjectivity, and SAPS-II between two groups.

patients were eligible for the study, among whom 580 patients suffered from in-hospital 28-day mortality from the date of ICU admission (dead group), while 1271 patients survived (survived group). The baseline characteristics of the two groups were compared as shown in Table 2, and the flowchart is presented in Figure 1.

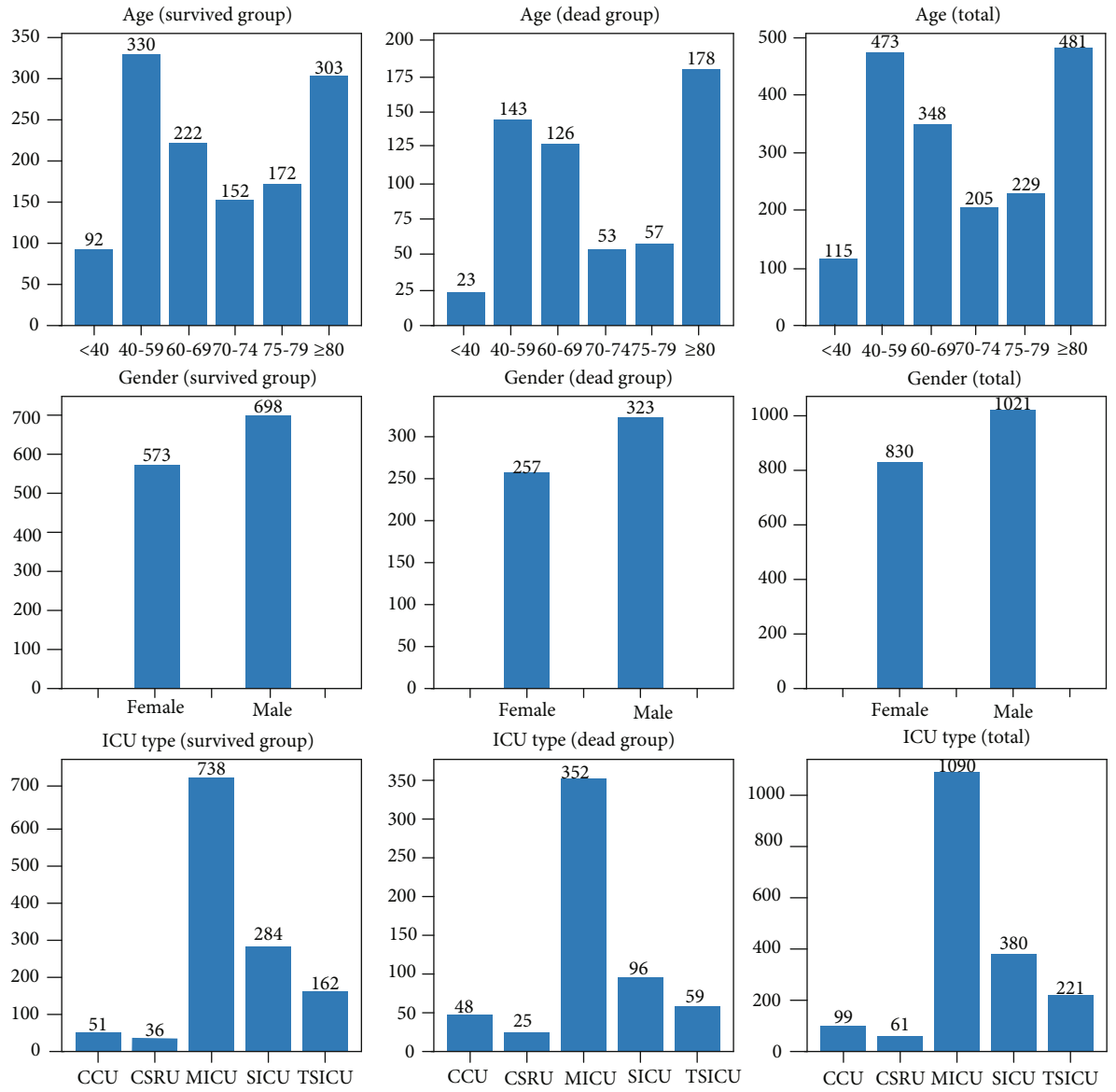


Figure 3: Histogram of the age, gender, and ICU type between survived and dead groups.

The sentiment polarity score of patients in the survived group was significantly higher than that in the dead group ($P < 0.001$), while the SAPS-II score was notably lower than that in the dead group ($P < 0.001$) (Table 2, Figure 2). The differences were significant between the two groups in age ($P < 0.001$) and ICU type ($P < 0.001$), but not in the sentiment subjectivity score ($P = 0.340$) and gender ($P = 0.757$) (Table 2, Figure 3).

3.2. COX Regression Analysis of the Association between Sentiments and 28-Day Mortality. As shown in Table 3, univariate analysis showed an inverse association between sentiment polarity and 28-day mortality (hazard ratio (HR): 0.458, 95% confidence interval (95% CI): 0.401-0.524, $P < 0.001$) and no association between sentiment subjectivity and 28-day mortality (HR: 0.863, 95% CI: 0.657-1.133, $P = 0.289$). The risk of 28-day mortality in sepsis patients would increase 0.04 times when 1 point in the SAPS-II score was increased each time (HR: 1.040, 95% CI: 1.036-1.045, $P < 0.001$). There was no association between gender and 28-day mortality (HR: 1.104, 95% CI: 0.936-1.301, $P = 0.240$).

In multivariate analysis, it was observed that both sentiment polarity (HR: 0.499, 95% CI: 0.409-0.610, $P < 0.001$) and sentiment subjectivity (HR: 0.710, 95% CI: 0.559-0.902, $P = 0.005$) were inversely associated with in-hospital 28-day mortality, while the SAPS-II score (HR: 1.034, 95% CI: 1.029-1.040, $P < 0.001$) was positively correlated with in-hospital 28-day mortality. The patients aged ≥80 years had an increased risk of in-hospital 28-day mortality compared with those aged <40 years (HR: 1.612, 95% CI: 1.032-2.520, $P = 0.036$). There were no differences in in-hospital 28-day mortality between the age of 40-59 (HR: 1.217, 95% CI: 0.781-1.886, $P = 0.385$), 60-69 (HR: 1.479, 95% CI: 0.943-2.321, $P = 0.089$), 70-74 (HR: 1.048, 95% CI:

Table 3: COX regression analysis of the association between sentiments and 28-day mortality.

Variables	Univariate analysis		Multivariate analysis		
	HR (95% CI)	*P*	HR (95% CI)	*P*	*β*
Sentiment polarity score	0.458 (0.401, 0.524)	<0.001	0.499 (0.409, 0.610)	<0.001	-0.694
Sentiment subjectivity score	0.863 (0.657, 1.133)	0.289	0.710 (0.559, 0.902)	0.005	-0.342
SAPS-II score	1.040 (1.036, 1.045)	<0.001	1.034 (1.029, 1.040)	<0.001	0.034
Age (<40) (years)					
40-59	3.027 (1.955, 4.687)	<0.001	1.217 (0.781, 1.886)	0.385	0.196
60-69	1.462 (0.941, 2.272)	0.091	1.479 (0.943, 2.321)	0.089	0.391
70-74	2.074 (1.328, 3.238)	0.001	1.048 (0.637, 1.723)	0.854	0.047
75-79	1.584 (0.969, 2.591)	0.067	1.030 (0.629, 1.687)	0.906	0.030
≥80	1.721 (1.058, 2.799)	0.029	1.612 (1.032, 2.520)	0.036	0.478
Gender (female)					
Male	1.104 (0.936, 1.301)	0.240	1.104 (0.934, 1.306)	0.245	0.099
ICU type (CCU)					
CSRU	0.335 (0.206, 0.545)	<0.001	0.397 (0.243, 0.650)	<0.001	-0.923
MICU	0.475 (0.351, 0.643)	<0.001	0.527 (0.388, 0.715)	<0.001	-0.641
SICU	0.254 (0.179, 0.361)	<0.001	0.299 (0.210, 0.426)	<0.001	-1.208
TSICU	0.240 (0.163, 0.354)	<0.001	0.280 (0.190, 0.414)	<0.001	-1.272

SAPS-II: Simplified Acute Physiology Score II; ICU: intensive care unit; CCU: coronary care unit; CSRU: cardiac surgery recovery unit; MICU: medical intensive care unit; SICU: surgical intensive care unit; TSICU: trauma/surgical intensive care unit; HR: hazard ratio; 95% CI: 95% confidence interval.

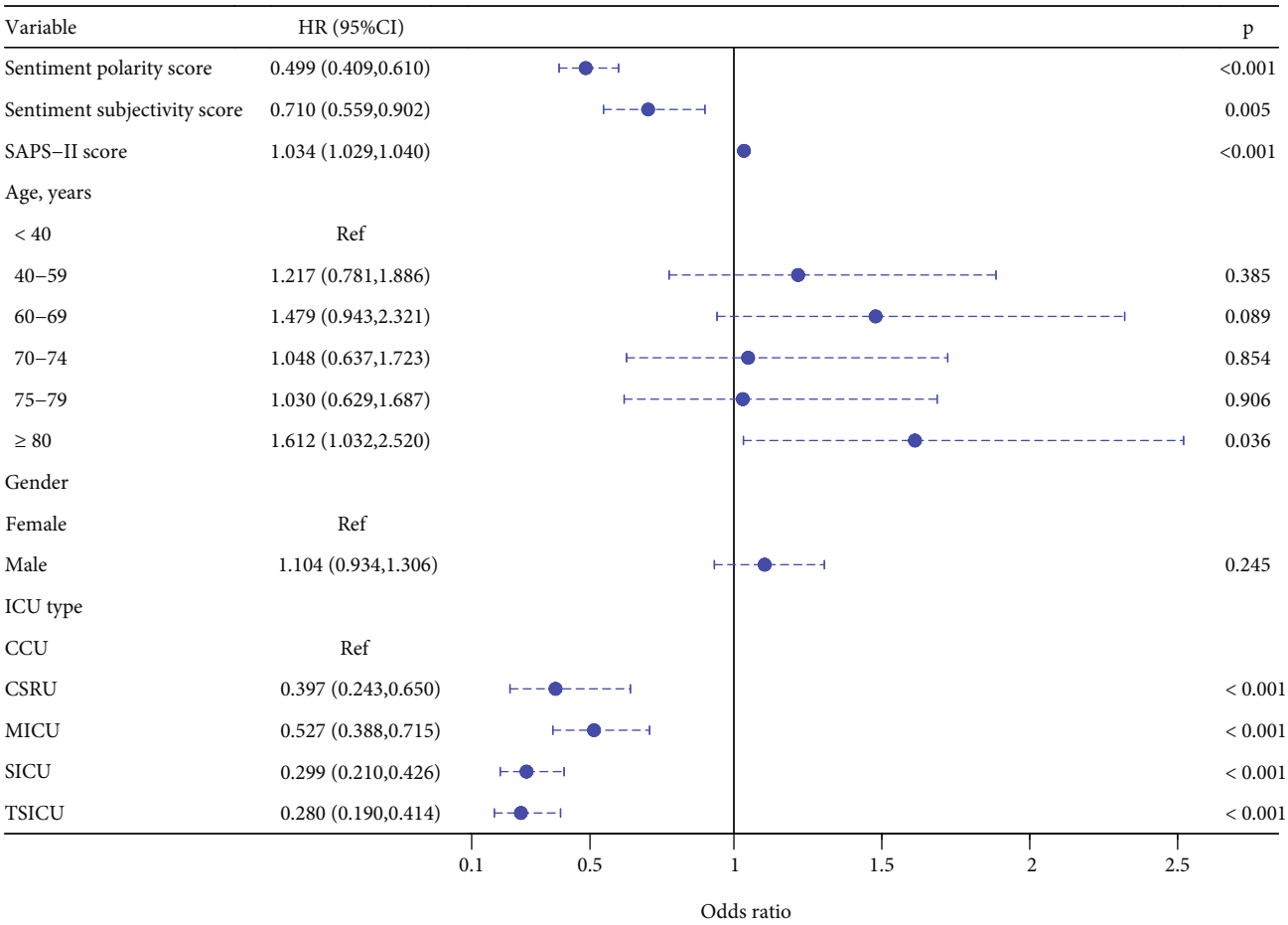


Figure 4: Forest plot of multivariate analysis of the association between sentiments and 28-day mortality.

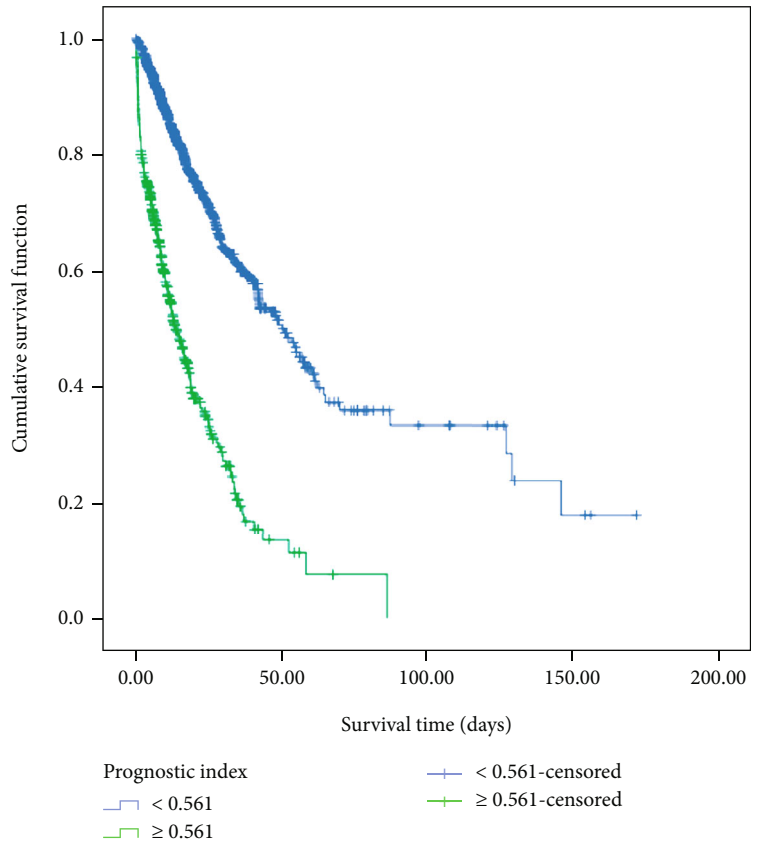


Figure 5: Survival curve comparison of two groups according to the individual prognostic index.

0.637-1.723, $P = 0.854$), 75-79 (HR: 1.030, 95% CI: 0.629-1.687, $P = 0.906$), and <40 years. In addition, no significant difference was found between gender and 28-day mortality (HR: 1.104, 95% CI: 0.934-1.306, $P = 0.245$). Patients that stayed in the trauma/surgical intensive care unit (TSICU) were least likely to die within 28 days after admission (HR: 0.280, 95% CI: 0.190-0.414, $P < 0.001$) (Table 3, Figure 4).

3.3. Survival Analysis. According to the individual PI, patients were assigned into the high-risk group (PI ≥ 0.561) and the low- and middle-risk group (PI < 0.561), and the survival curves are illustrated in Figure 5. It could be observed that the median death time of the high-risk group was significantly earlier than that of the low- and middle-risk group (13.5 *vs.* 49.8 days, $P < 0.001$).

4. Discussion

In the present study, a total of 1851 sepsis patients were eligible according to inclusion and exclusion criteria, among whom 580 cases suffered from in-hospital 28-day mortality, while 1271 cases survived. Multivariate COX analysis showed that sentiment polarity and sentiment subjectivity were inversely associated with in-hospital 28-day mortality. Based on the quartiles of the individual PI, patients were assigned into the high-risk group and the low- and middle-risk group. Survival analysis indicated that the high-risk group had earlier median death time compared with the low- and middle-risk group. These all suggested that the quantitative measurement of sentiments in nursing notes was associated with the in-hospital 28-day mortality and survival of sepsis patients; nursing notes containing rich information may serve as a potential predictor of clinical outcomes in the ICU.

To the best of our knowledge, brief fragments of the text are conducive to reflecting the author's feelings about a given topic. Recently, language processing tools have been developed and allow the characterization of feelings, such as the sentiment in text documents [18]. Sentiment is usually described as the relative positivity or polarity of a text string and is measured by a number from -1 (very negative) to 1 (very positive) [14]. It can also be interpreted as the estimated probability of "positive" or "negative" through a classifier. Sentiment analysis permits us to gain insights into the clinicians' emotions and attitudes towards patients through the subjective expressions made by clinicians in the text of clinical notes, thus contributing to the prediction of patients' outcomes [19–22]. In health-related fields, sentiment analysis has been widely applied to Cancer Survivors Network (CSN) breast and colorectal cancer discussion posts [23], health reforms on Twitter [24], encounter notes of patients with critical illness [25], etc. This study was aimed at identifying the association between sentiments in nursing notes and the in-hospital 28-day mortality of sepsis patients. The results exhibited that both sentiment polarity and sentiment subjectivity were inversely associated with in-hospital 28-

day mortality, supported by the results of McCoy et al. that the sentiment measured in hospital discharge notes was related to hospital readmissions and mortality risk [11]. Based on the COX model, the patients with PI ≥ 0.561 were found to have a higher risk of death than those with PI < 0.561, highlighting the potential value of sentiments in survival analysis. A previous study has shown a strong association between sentiments and the risk of death even after adjustment for severity of illness and baseline information [25].

The superiority of the present study was that it was the first study to investigate the association between sentiments in nursing notes and the in-hospital 28-day mortality of sepsis patients. The nursing notes written less than 12 hours before the time of death were excluded, which made the results more reliable. However, the present study also had several limitations that should be cautiously interpreted. First, nursing notes from the MIMIC-III database with single-center samples may manifest different characteristics because of variations in clinicians, experience, training, or working environment, easily causing the results to be nongeneralizable. Second, the approach used to measure the sentiment in the present study was not the only approach available. Other techniques could produce different results, such as those based on the machine learning model to make semantic inferences. Third, the mean sentiment scores could only characterize the variations at the level of patients, but not at the levels of sentences, paragraphs, or documents. Forth, the nursing notes were recorded by caregivers who are research nurses, medical doctors, or so on (available at https://mimic.mit.edu/docs/iii/tab les/caregivers/). It cannot be determined whether the sentiments based on nursing notes are based on past or personal experiences. Moreover, the subtle difference in sentiments was not obtained over time. In the future, the temporal mode of nursing notes will be examined to gain more insights.

5. Conclusions

Sentiments in nursing notes are associated with the in-hospital 28-day mortality and survival of sepsis patients, suggesting the importance of sentiments in nursing notes for the prediction of clinical outcomes in the ICU. Although predicting clinical outcomes is still a complex problem, the information extracted from unstructured data like nursing notes may contribute to further improving prediction performance.

Abbreviations

PI:	Prognostic index
SAPS-II:	Simplified Acute Physiology Score II
ICU:	Intensive care unit
SOI:	Severity of illness scores
MIMIC-III:	Medical Information Mart for Intensive Care II
ICD-9:	International Classification of Diseases 9
HR:	Hazard ratio
TSICU:	Trauma/surgical intensive care unit
CSN:	Cancer Survivors Network.

Data Availability

The data utilized to support the findings are available from the corresponding authors upon request. The data applied in the present study were from the MIMIC-III database (https://mimic.physionet.org/), a freely accessible database.

Ethical Approval

The data collected in the MIMIC-III was approved by the Ethics Review Board of the Beth Israel Deaconess Medical Center, and all private information has been desensitized.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

QYG contributed to the study design and manuscript writing. DW, PPS, and WFW contributed to the data collection and analysis. XRL contributed to the study design and study supervision. QYG, DW, PPS, WFW, and XRL contributed to the critical revisions of important content. All authors revised and approved the final manuscript.

Acknowledgments

This research was funded by the Shandong Social Science Planning Research Project-2019 (19CCXJ05).

References

[1] S. Li, X. Hu, J. Xu et al., "Increased body mass index linked to greater short- and long-term survival in sepsis patients: a retrospective analysis of a large clinical database," *International Journal of Infectious Diseases*, vol. 87, pp. 109–116, 2019.

[2] M. Singer, C. S. Deutschman, C. W. Seymour et al., "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *Journal of the American Medical Association*, vol. 315, no. 8, pp. 801–810, 2016.

[3] C. W. Seymour, T. D. Rea, J. M. Kahn, A. J. Walkey, D. M. Yealy, and D. C. Angus, "Severe sepsis in pre-hospital emergency care," *American Journal of Respiratory and Critical Care Medicine*, vol. 186, no. 12, pp. 1264–1271, 2012.

[4] A. J. Walkey, T. Lagu, and P. K. Lindenauer, "Trends in sepsis and infection sources in the United States. A population-based study," *Annals of the American Thoracic Society*, vol. 12, no. 2, pp. 216–220, 2015.

[5] C. Fleischmann, A. Scherag, N. K. Adhikari et al., "Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations," *American Journal of Respiratory and Critical Care Medicine*, vol. 193, no. 3, pp. 259–272, 2016.

[6] J. Cohen, J. L. Vincent, N. K. Adhikari et al., "Sepsis: a roadmap for future research," *The Lancet Infectious Diseases*, vol. 15, no. 5, pp. 581–614, 2015.

[7] I. E. R. Waudby-Smith, N. Tran, J. A. Dubin, and J. Lee, "Sentiment in nursing notes as an indicator of out-of-hospital

mortality in intensive care patients," *PLoS One*, vol. 13, no. 6, article e0198687, 2018.

[8] G. Rocker, D. Cook, P. Sjokvist et al., "Clinician predictions of intensive care unit mortality," *Critical Care Medicine*, vol. 32, no. 5, pp. 1149–1154, 2004.

[9] T. Sinuff, N. K. Adhikari, D. J. Cook et al., "Mortality predictions in the intensive care unit: comparing physicians with scoring systems," *Critical Care Medicine*, vol. 34, no. 3, pp. 878–885, 2006.

[10] M. M. Ghassemi, R. G. Mark, and S. Nemati, "A visualization of evolving clinical sentiment using vector representations of clinical notes," in *2015 Computing in cardiology conference (CinC)*, pp. 629–632, Nice, France, 2015.

[11] T. H. McCoy, V. M. Castro, A. Cagan, A. M. Roberson, I. S. Kohane, and R. H. Perlis, "Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: an electronic health record study," *PLoS One*, vol. 10, no. 8, article e0136341, 2015.

[12] N. Tran and J. Lee, "Using multiple sentiment dimensions of nursing notes to predict mortality in the intensive care unit," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 283–286, Las Vegas, NV, USA, 2018.

[13] A. E. Johnson, T. J. Pollard, L. Shen et al., "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, article 160035, 2016.

[14] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, Springer, Boston, MA, 2012.

[15] S. Loria, P. Keen, and M. Honnibal, *TextBlob: Simplified Text Processing*, 2018, 2018, https://media.readthedocs.org/pdf/textblob/dev/textblob.pdf.

[16] J. Zhang and H. Jin, "Method of subjective lexicon creation for Chinese sentiment analysis," *Applied Mechanics & Materials*, vol. 34-35, pp. 801–805, 2010.

[17] J. R. Le Gall, S. Lemeshow, and F. Saulnier, "A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study," *Journal of the American Medical Association*, vol. 270, no. 24, pp. 2957–2963, 1993.

[18] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth, "Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter," *PLoS One*, vol. 6, no. 12, article e26752, 2011.

[19] J. Cioffi, "Recognition of patients who require emergency assistance: a descriptive study," *Heart & Lung*, vol. 29, no. 4, pp. 262–268, 2000.

[20] M. Brabrand, J. Hallas, and T. Knudsen, "Nurses and physicians in a medical admission unit can accurately predict mortality of acutely admitted patients: a prospective cohort study," *PLoS One*, vol. 9, no. 7, article e101739, 2014.

[21] S. A. Collins and D. K. Vawdrey, ""Reading between the lines" of flow sheet data: nurses' optional documentation associated with cardiac arrest outcomes," *Applied Nursing Research*, vol. 25, no. 4, pp. 251–257, 2012.

[22] S. A. Collins, K. Cato, D. Albers et al., "Relationship between nursing documentation and patients' mortality," *American Journal of Critical Care*, vol. 22, no. 4, pp. 306–313, 2013.

[23] K. Portier, G. E. Greer, L. Rokach et al., "Understanding topics and sentiment in an online cancer survivor community," *Journal of the National Cancer Institute. Monographs*, vol. 2013, no. 47, pp. 195–198, 2013.

[24] D. King, D. Ramirez-Cano, F. Greaves, I. Vlaev, S. Beales, and A. Darzi, "Twitter and the health reforms in the English National Health Service," *Health Policy*, vol. 110, no. 2-3, pp. 291–297, 2013.

[25] G. E. Weissman, L. H. Ungar, M. O. Harhay, K. R. Courtright, and S. D. Halpern, "Construct validity of six sentiment analysis methods in the text of encounter notes of patients with critical illness," *Journal of Biomedical Informatics*, vol. 89, pp. 114–121, 2019.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 2747274, 10 pages
https://doi.org/10.1155/2021/2747274

Research Article

Segmentation and Automatic Identification of Vasculature in Coronary Angiograms

Yaofang Liu,[1] Wenlong Wan,[2] Xinyue Zhang,[1] Shaoyu Liu,[2] Yingdi Liu,[1] Hu Liu,[3] Xueying Zeng,[1] Weiguo Wang,[1] and Qing Zhang[4]

[1]*School of Mathematical Sciences, Ocean University of China, 238 Songling Road, Qingdao, Shandong 266100, China*
[2]*School of Computer Science and Technology, Ocean University of China, 238 Songling Road, Qingdao, Shandong 266100, China*
[3]*School of Materials Science and Engineering, Ocean University of China, 238 Songling Road, Qingdao, Shandong 266100, China*
[4]*Department of Cardiology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong 266035, China*

Correspondence should be addressed to Weiguo Wang; wgwang@ouc.edu.cn and Qing Zhang; qingzhang2019@foxmail.com

Received 1 August 2021; Revised 28 August 2021; Accepted 3 September 2021; Published 7 October 2021

Academic Editor: Tao Song

Coronary angiography is the "gold standard" for the diagnosis of coronary heart disease, of which vessel segmentation and identification technologies are paid much attention to. However, because of the characteristics of coronary angiograms, such as the complex and variable morphology of coronary artery structure and the noise caused by various factors, there are many difficulties in these studies. To conquer these problems, we design a preprocessing scheme including block-matching and 3D filtering, unsharp masking, contrast-limited adaptive histogram equalization, and multiscale image enhancement to improve the quality of the image and enhance the vascular structure. To achieve vessel segmentation, we use the C-V model to extract the vascular contour. Finally, we propose an improved adaptive tracking algorithm to realize automatic identification of the vascular skeleton. According to our experiments, the vascular structures can be successfully highlighted and the background is restrained by the preprocessing scheme, the continuous contour of the vessel is extracted accurately by the C-V model, and it is verified that the proposed tracking method has higher accuracy and stronger robustness compared with the existing adaptive tracking method.

1. Introduction

Cardiovascular disease is currently recognized as one of the most important chronic diseases leading to human death in the world. In recent years, morbidity and mortality from cardiovascular diseases continue to increase, ranking first among various diseases. Coronary angiography (CA) is a common and effective method for diagnosing coronary heart disease. It is regarded as the "gold standard" for the diagnosis of coronary heart disease and is widely used in clinical diagnosis [1].

Normally, human arteries and vessels are invisible under X-rays. However, by injecting X-ray impervious substances into the coronary arteries and then irradiating the coronary artery area with X-rays, the arteries and vessels can be visualized. To decide the treatment plan, doctors need to find the location and degree of coronary artery stenosis based on the image by themselves. Nevertheless, in this way, a large amount of repetitive work and subjective errors are inevitable. Thus, it is of great benefit to invent technologies to segment and identify vessels in angiograms. For this reason, many scholars have proposed various methods.

For many years, image segmentation is one of the focuses of image processing. Up to now, many segmentation technologies for vessels have been proposed. Based on the two characteristics of discontinuity between regions and similarity within regions, we can divide vessel segmentation technologies into three categories: boundary-based segmentation technologies [2–8], region-based segmentation technologies [9–11], and technologies combined with specific

theories and tool segmentation [12–15]. Sahoo et al. [16] adopted the maximum entropy method and the gray threshold that maximizes entropy corresponded to the optimal segmentation threshold. Sato et al. [17] constructed a multiparameter similarity function for enhancing vessels by analyzing the properties of the eigenvalues of the Hessian matrix of spherical, tubular, and sheet-like structures at a certain scale. Based on the simplified Mumford-Shah model and the level set idea, Chan and Vese [18] proposed a new active contour C-V model to evolve the curve through the minimization of the energy function. Most recently, deep learning methods have also been widely used in the field of vessel segmentation. For example, Chen et al. [19] trained the 3D U-Net to perform three-dimensional vessel segmentation and achieved high segmentation accuracy.

Moreover, people have studied a variety of methods for vascular identification, such as multiscale-based methods [20–24] and tracking-based methods [25–29]. Among these methods, the tracking-based method has been proved to be very effective. It can detect coronary information based on the local response of angiogram without scanning the entire image. In the process of coronary artery extraction, the extraction result is unstable due to the manual setting of seed points. Aiming at this problem, Xiao et al. [30] proposed an automatic seed point acquisition method based on ridge point detection. These ridge points serve as seed points for adaptive tracking of the centerline of the coronary artery. Aylward and Bullitt [26] proposed a multiscale spatial centerline tracking algorithm based on ridge detection, which uses the eigenvalue decomposition of the Hessian matrix to extract the ridge. However, due to limitations in algorithm design and the effects of low image quality, noise, etc., the accuracy and robustness of these methods still have room for improvement.

Our main work and contributions are as follows: first, we designed a preprocessing scheme to increase the quality of the image and enhance the vascular structure. Then, we used the C-V model to achieve vessel segmentation. Finally, we proposed an improved adaptive tracking algorithm to realize automatic identification of the vascular skeleton, which achieved better effects than former methods according to our experiments.

This paper is organized as follows. In Section 2, we introduce our scheme of image preprocessing. Section 3 describes the active contour model to extract the vascular contour. Section 4 describes the details of our proposed improved adaptive tracking method. Section 5 presents the analysis and experimental results of testing the robustness and accuracy of our methods. Finally, conclusions are drawn in Section 6.

2. Image Preprocessing

The complex and varied configuration of the coronary artery structure, noise caused by various factors, artifact caused by the beating of the heart, and low contrast of terminal vessels make precise segmentation very challenging. Therefore, before the extraction of coronary artery structure, coronary angiograms should be preprocessed to enhance the vascular structure and suppress the background noise. In this paper, block-matching and 3D filtering (BM3D) [31] is used to effectively filter out noise. Unsharp masking (UM) [32], contrast-limited adaptive histogram equalization (CLAHE), [33] and multiscale image enhancement [34] are used to improve image contrast and highlight the vascular structure.

2.1. Block-Matching and 3D Filtering. BM3D is a 3D block-matching algorithm used primarily for noise reduction in images. Firstly, by the grouping technique of block-matching, image fragments are grouped based on similarity and are integrated into a three-dimensional matrix. Then, filtering is done on every fragment group. At last, the image is transformed back into its two-dimensional form and all overlapping image fragments are weight-averaged to ensure that they are filtered for noise yet retain their distinct signal. This algorithm can effectively remove image noise.

2.2. Unsharp Masking. The main procedures of UM algorithm are as follows: first, a passivated fuzzy image is generated after low-pass filtering of the original image. Then, the image with high-frequency components is obtained by subtraction of the original image and the fuzzy image. Finally, the high-frequency image is enlarged with a parameter and superimposed with the original image; that is, an image with enhanced edges is generated. The specific algorithm steps are as follows:

(1) Generate the smoothing result:

$$g_{\text{mask}}(x, y) = I(x, y) - \bar{I}(x, y), \tag{1}$$

where $I(x, y)$ represents the gray of the pixel (x, y), $\bar{I}(x, y)$ represents the gray of the pixel (x, y) after low-pass filtering, and $g_{\text{mask}}(\bullet)$ generates the high-frequency component of the image

(2) Add the passivation template to the original image with a certain proportion:

$$\mathrm{g}(x, y) = I(x, y) + k * g_{\text{mask}}(x, y), \tag{2}$$

where k is the enlarge coefficient and $g(\bullet)$ generates the image with enhanced edges

2.3. Contrast-Limited Adaptive Histogram Equalization. As a variant of adaptive histogram equalization, the CLAHE method limits the contrast amplification to reduce excessive amplification of noise. In CLAHE, the contrast amplification in the vicinity of a given pixel is given by the slope of the transformation function, which is proportional to the slope of the neighborhood cumulative distribution function (CDF) and therefore to the value of the histogram. CLAHE limits the amplification by clipping the histogram at a predefined value before computing the CDF. This limits the slope of the CDF and therefore of the transformation function. It is advantageous not to discard the part of the histogram that exceeds the clip limit but to redistribute it equally among all

histogram bins. The process of clipping the histogram is shown in Figure 1.

2.4. Multiscale Image Enhancement. Frangi et al. [34] proposed the multiscale enhancement method based on the Hessian matrix of the image. In this method, the relationship among the eigenvalues, eigenvectors of the Hessian matrix, and the orientation of vascular structure are utilized, combined with the multiscale theory. Then, the vascular structure in the coronary angiogram is detected by constructing an appropriate vascular similarity function. At present, the method has become one of the most commonly used methods of multiscale enhancement.

The vascular similarity function is established as follows:

$$V(P;\sigma) = \begin{cases} 0, \text{if } \lambda_2 > 0, \\ \exp\left(-\frac{R_B^2}{2\beta^2}\right) \exp\left(-\frac{2m^2}{\lambda_2^2}\right)\left(1 - \exp\left(-\frac{S_H^2}{2c^2}\right)\right), \text{otherwise}, \end{cases} \tag{3}$$

where λ_1 and λ_2 are two eigenvalues of the Hessian matrix, $|\lambda_1| \leq |\lambda_2|$, P is an arbitrary point in the image, σ is the scale parameter, $R_B = |\lambda_1|/|\lambda_2|, S_H = \sqrt{\lambda_1^2 + \lambda_2^2}$, and R_B is the vascular structure enhancement factor, which is used to distinguish the globular structures from the tubular structures; S_H is the norm of the Hessian matrix; and β, c, and m control the overall smoothness of linear objects.

When the scale factor σ is consistent with the width of the tubular structure, the filtering result $V(P;\sigma)$ gets the maximum value. By iterating the scale parameters σ, the values $V(P;\sigma)$ under different scales are obtained, and the maximum value is taken as the actual output of the point P:

$$V(P) = \max_{\sigma_{\min} \leq \sigma \leq \sigma_{\max}} V(P;\sigma), \tag{4}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are the minimum and maximum sizes of the vascular structure, respectively.

3. Vessel Segmentation

In this section, we will introduce the active contour model to extract the vascular contour of coronary angiograms.

Kass et al. [35] proposed the active contour model (ACM). This method converts the image segmentation problem into solving an energy minimization problem. The contour curve is the edge of the blood vessel when the energy function reaches the minimum. The active contour model is mainly divided into edge-based and region-based according to the different construction methods of the energy function. The most prominent advantage of the ACM is its resistance against strong noise.

The C-V model [18, 36, 37] is a representative region-based active contour model. The specific algorithm steps are as follows:

(1) Put forward the energy function:

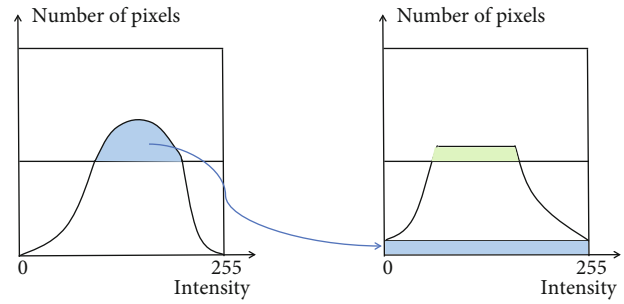


Figure 1: The clipping process of CLAHE.

$$F(C, c_0, c_b) = u \cdot L(C) + v \cdot S_b(C) + \lambda_0 \int_{\text{outside}} |I(x,y) - c_0|^2 dxdy + \lambda_b \int_{\text{inside}} |I(x,y) - c_b|^2 dxdy, \tag{5}$$

where c_0, c_b represent the average gray levels of the outside and inside areas of the curve C, respectively; $L(C)$ represents the length of the closed curve C; $S_b(C)$ represents the area of the inner area of C; and u, v, λ_0, and λ_b represent the weights of items in the energy function.

(2) Introduce the level set method, set $w(x,y)$ as a sign distance function with positive, negative, zero representing inside, outside, and right on the curve C, respectively:

$$\begin{cases} C = \{(x,y): w(x,y) = 0\}, \\ \text{inside}(C) = \{(x,y): w(x,y) > 0\}, \\ \text{outside}(C) = \{(x,y): w(x,y) < 0\}. \end{cases} \tag{6}$$

Introduce the following H and δ functions:

$$H(w) = \begin{cases} 1, w \geq 0, \\ 0, w < 0, \end{cases} \tag{7}$$

$$\delta(w) = \frac{d}{dw} H(w). \tag{8}$$

Rewrite the energy function as a level set equation:

$$F(C, c_0, c_b) = u \cdot \int \delta(w)|\nabla w| dxdy + v \cdot \int H(w) dxdy + \lambda_0 \int_{\text{outside}} |I(x,y) - c_0|^2 (1 - H(w)) dxdy + \lambda_b \int_{\text{inside}} |I(x,y) - c_b|^2 H(w) dxdy, \tag{9}$$

where λ_b and λ_0 are the iterative parameters in the C-V model, and their values affect the evolution rate of the curve C. When the curve C contains the segmentation target, the

internal homogeneity of the curve C is low; thus, it is necessary to enlarge λ_0 to accelerate the evolution of the curve C to the target, and vice versa.

(3) The energy minimization problem can be solved by minimizing the level set equation iteratively

The C-V model minimizes the energy function to obtain the evolution curve that approaches the edge of the blood vessels and finally segments the target. Compared with other methods, it has better effects on the continuous gradient.

4. Improved Adaptive Tracking

In this section, we will propose an improved adaptive tracking method, which is more robust and has fewer misjudgments in the tracking process, to automatically extract the skeleton of the coronary blood vessels.

4.1. Ridge Point Detection. Ridge point detection is important for seed point selection, blood vessel tracking, and bifurcation point detection. Ridge point is the local gray maximum point of the two-dimensional image. After multiscale enhancement, a ridge point of the blood vessel is usually located at the maximum point perpendicular to the direction of the blood vessel. The gradient of the local maximum point in the image is zero, and its Hessian matrix is negative [38]. Since the coordinates of image pixels are all integers, according to the principle of linear interpolation, if the point (ε, η) $(x < \varepsilon < x + 1, y < \eta < y + 1)$ satisfies the following conditions:

$$\begin{aligned} &\nabla(x, y)\nabla(x + 1, y + 1) < 0 \text{ or} \\ &\nabla(x + 1, y)\nabla(x, y + 1) < 0, \\ &\lambda_i(x + m, y + n) < 0, (i = 1, 2, m = 0, 1, n = 0, 1), \end{aligned} \tag{10}$$

where $\nabla(x, y)$ is the gray gradient of point (x, y) and $\lambda_i(x, y)$ are the eigenvalues of the Hessian matrix of point (x, y); then, (ε, η) can be considered as a local maximum point, and the pixel (x, y), as its approximate solution, is defined as a ridge point.

The ridge points may be misjudged due to the image noise caused by the uneven distribution of contrast agents and other factors. Thus, a gray threshold is used to screen out those misjudged ridge points. This method can effectively remove most of the ridge points outside the blood vessel.

4.2. Tracking Process. The tracking algorithm starts from a seed point and gradually tracks to the end of the vessel, extracting the blood vessel skeleton. We randomly select the seed point from the detected ridge points.

The initial tracking direction can be calculated from the gray information around the seed point. According to [38], take the seed point as the center and search for the gray maximum point P^+ on the circle with radius d. P^+ is the first point of forward tracking, the forward initial tracking direction u^+ and angle θ^+ can be expressed as

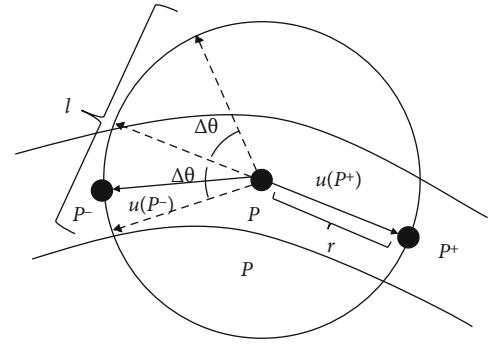


Figure 2: Initial direction detection.

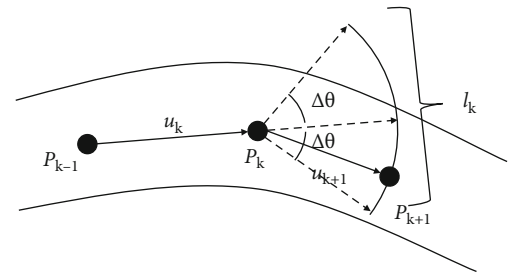


Figure 3: Forward tracking.

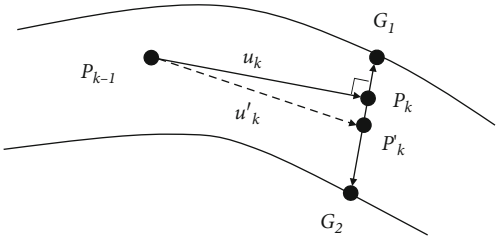


Figure 4: Centerline adjustment.

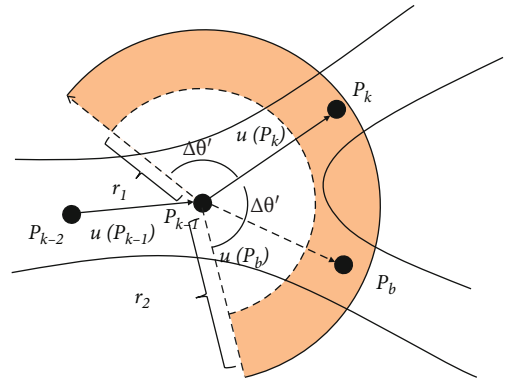


Figure 5: Vascular branch detection.

$$u(P^+) = \frac{P^+ - P}{\|P^+ - P\|} = \left(\cos\theta^+, \sin\theta^+\right). \tag{11}$$

After obtaining the forward tracking direction, we search for the local maximum point P^- on arc $l(2\pi - \theta^+ - \Delta\theta, 2\pi - \theta^+ + \Delta\theta)$ centered in the opposite direction $(2\pi - \theta^+)$

(a) (b) (c)

Figure 6: Three selected original images.

of the forward tracking angle θ^+. The search area is shown in Figure 2.

The backward direction of the initial trace u^- can be calculated as

$$u(P^-) = \frac{P^- - P}{\|P^- - P\|}. \tag{12}$$

Tracking from the current point forward to the next point is the main step of this algorithm. The current tracking direction is determined by the direction from the previous point P_{k-1} to the current point P_k:

$$u_k = \frac{P_{k-1} - P_k}{\|P_{k-1} - P_k\|}. \tag{13}$$

After determining the tracking direction, we search for the local maximum point P_{k+1} on arc $l_k(\theta_k - \Delta\theta, \theta_k + \Delta\theta)$ and the following conditions should be met:

$$\begin{cases} I(P_{k+1}) > I_0, \\ N_P(P_{k+1}) < \tau_P, \end{cases} \tag{14}$$

where $I(P_{k+1})$ is the gray of P_{k+1}, $N_P(P_{k+1})$ is the number of tracking points around P_{k+1}, and I_0 and τ_P are two thresholds. The first condition is to prevent the overtracking beyond the vessel area, while the second condition can avoid repeatedly tracking the vessel and being trapped in a local endless loop. If both conditions are satisfied, we continue to track from P_{k+1}. Otherwise, P_{k+1} is the endpoint of the vessel. We illustrate the tracking process in Figure 3.

Due to the noise and other issues mentioned above, a few tracking points may deviate from the center of the vessel. The tracking point can be adjusted to the center by center adjustment, which combines the blood vessel contour and tracking direction information. The specific steps are as follows: get the normal line of the vessel through the vertical direction of the current tracking direction and find the intersection points G_1, G_2 of the normal line and the blood vessel contour;, then tracking point P_k can be adjusted to

$$P'_k = \frac{G_1 + G_2}{2}, \tag{15}$$

meanwhile, change the tracking direction u_k to

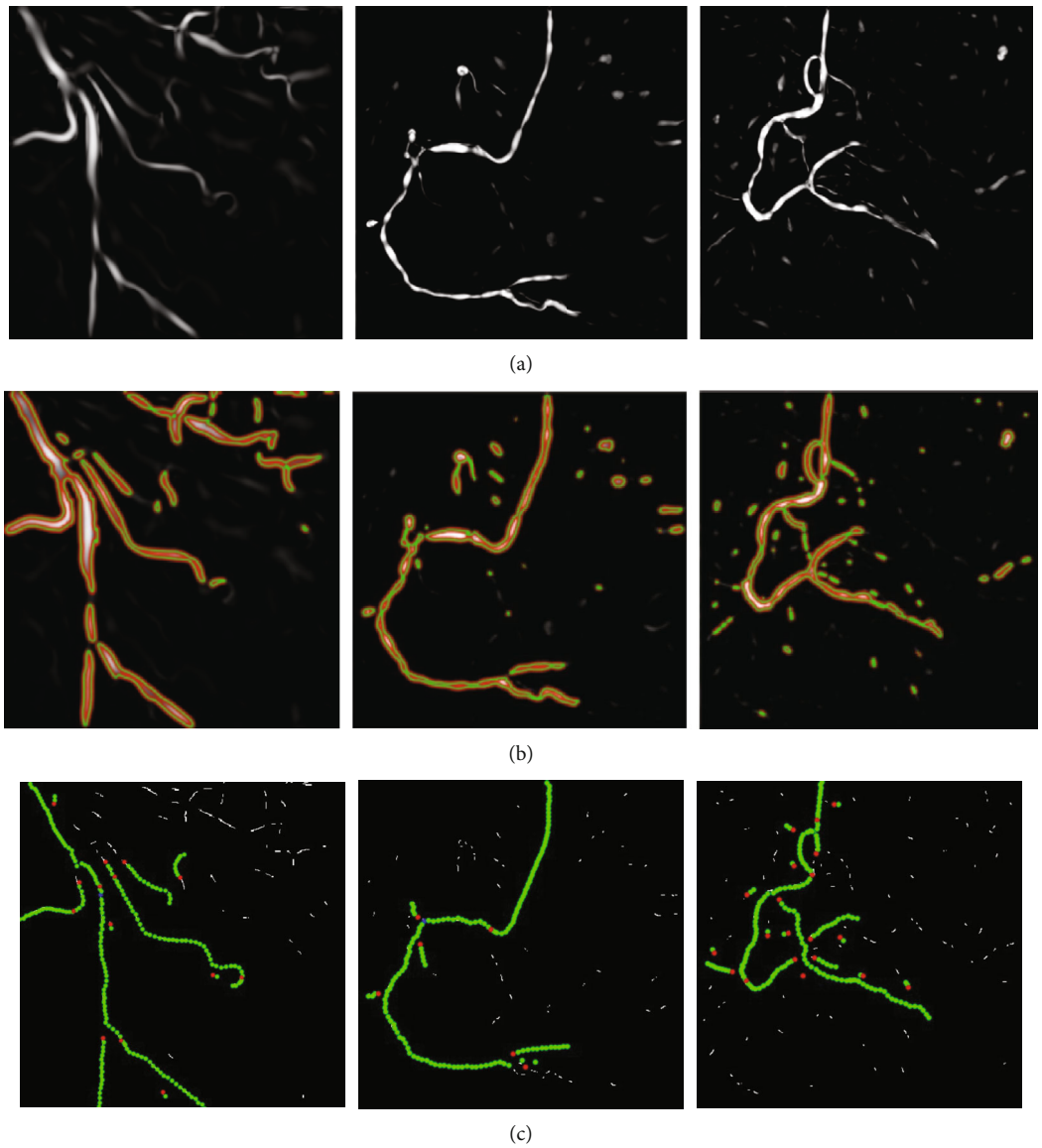
$$u'_k = \frac{P'_k - P_{k-1}}{\|P'_k - P_{k-1}\|}. \tag{16}$$

The adjustment process is illustrated in Figure 4.

Bifurcation detection is another important process of the tracking algorithm. Ideally, we only need to distinguish two different vessel branches at the vessel bifurcation. However, in the actual tracking process, the accurate positions of vessel bifurcations are usually unknown. Thus, bifurcation detection is required at each point of the tracking process. We propose a robust bifurcation detection method. It includes two main steps: first, obtain one branch point (tracking point) P_k by the tracking method, and second, search in the fan ring area between angle $(\theta_k - \Delta\theta', \theta_k + \Delta\theta')$ and radius (r_1, r_2) to find a ridge point that satisfies the following conditions:

$$\begin{cases} |\theta_b - \theta_k| > \tau_1, \\ |\theta_b - \theta_{k-1}| > \tau_2, \\ \|P_b - P_k\| > d, \\ N_B(P_b) < \tau_B, \end{cases} \tag{17}$$

where P_b and $u(P_b)$ represent the detected ridge point of the new branch (the branch point) and its direction, respectively; $N_B(P_b)$ is the number of bifurcations around P_b; and τ_B is a threshold. The first three conditions mean that when $u(P_b)$ significantly differs from $u(P_k)$ and $u(P_{k-1})$ and the distance between P_b and P_k is large enough, the new branch has a large gap with the former branch. The last condition indicates that $N_B(P_b)$ should be smaller than τ_B to avoid duplication with existing tracking. If all the conditions are satisfied, P_b is detected as a bifurcation point and we keep tracking the branch vessels. The schematic diagram of bifurcation detection is shown in Figure 5.

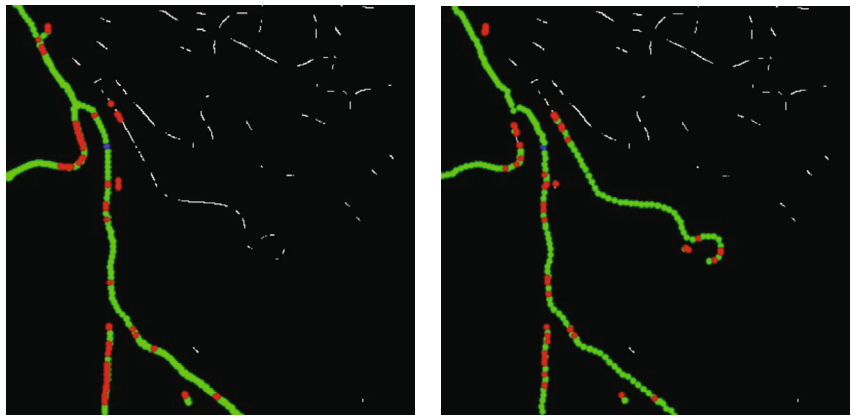


(a)

(b)

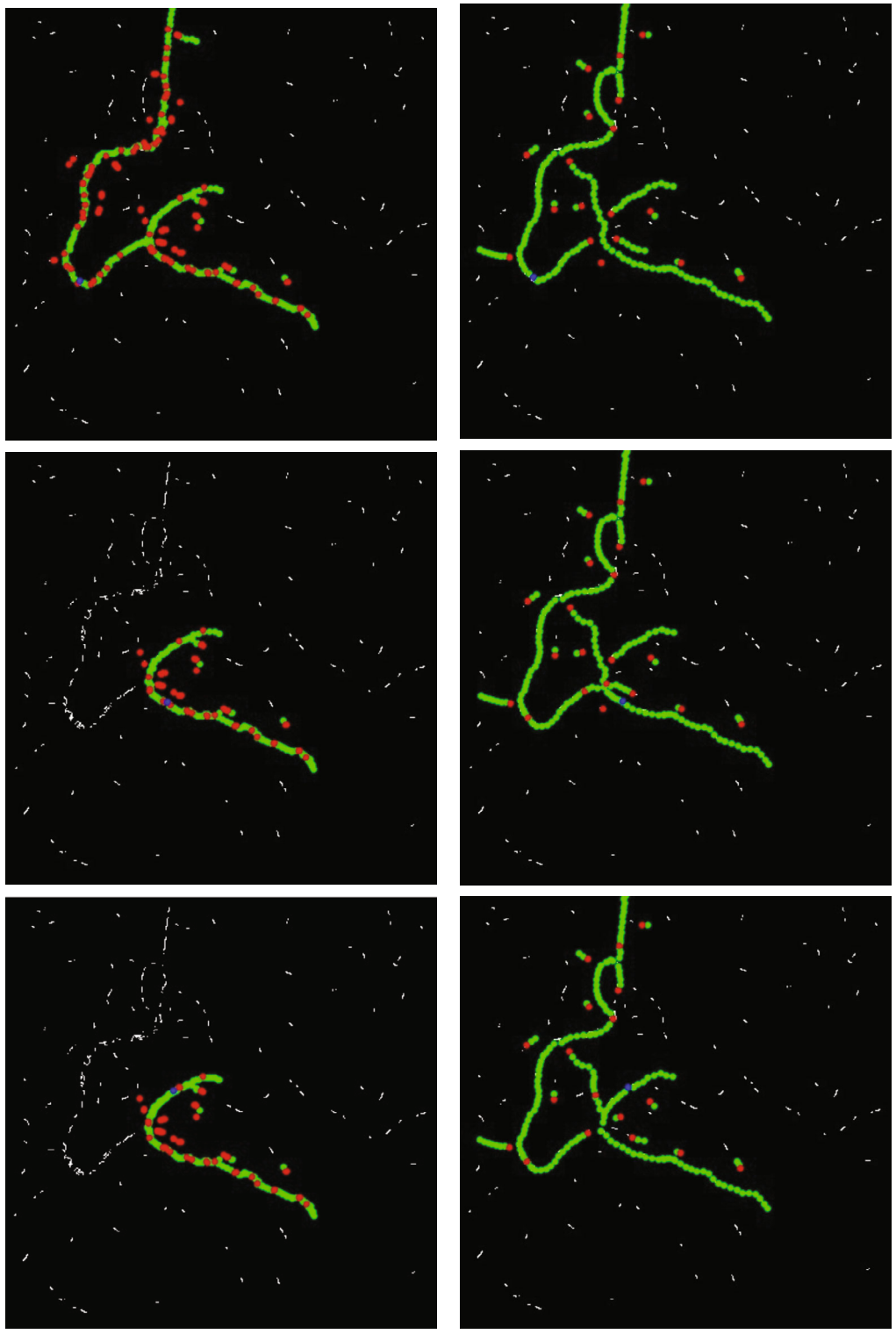
(c)

Figure 7: Experimental results of three original images obtained by applying the proposed method. (a) Images preprocessed. (b) Vascular contour segmentation. (c) Improved adaptive tracking (red dots are bifurcation points, green dots are normal tracking points).



(a) (b)

Figure 8: Comparison of the tracking effect between our proposed method and the method of [38]. (a) Results of [38]. (b) Results of the proposed method.



(a) (b)

Figure 9: Experimental results of different seed points (blue). (a) The method of [38]. (b) The proposed method.

In addition, before tracking, the ridge image can be preprocessed to remove the scattered and distributed ridge points, it can also reduce the misjudgments of bifurcations. The specific steps areas follows: set a threshold τ_R, count the number of surrounding ridge points for each ridge point $N_R(P)$, and then remove this ridge point if $N_R(P) < \tau_R$ and keep it otherwise.

5. Results and Analysis

In this section, we will conduct several experiments to justify the effectiveness of the proposed method. All the images are captured from the video data of coronary angiograms provided by Qilu Hospital (Qingdao). The experiments are implemented on an Intel Core i5-8300H and 8 GB of RAM processor using MATLAB software of version 2019b.

We carefully selected the parameters used. In multiscale enhancement, we set β, c, and σ to 0.5, 20, and $[1:10]$. In the proposed tracking method, we set the radius to 5 pixels and $\Delta\theta$ to 45° in forward tracking. For bifurcation detection, it needs a larger area for searching; thus, we set the radius (r_1, r_2) to $(7, 12)$ pixels and $\Delta\theta'$ to 135° which can avoid backward tracking. Note that we set other thresholds I_0, $\tau_1, \tau_2, d, \tau_P, \tau_B$, and τ_R to $10, 45°, 30°, 5, 4, 2$, and 3.

Three images with the different vascular structures were selected for independent experiments, which are shown in Figure 6. We applied our methods for these images, and the results are shown in Figure 7.

As can be seen from Figure 7, even though the vascular structures in the images are very different, the proposed method still has a nice experimental effect. From the images preprocessed (a), we can find that after the image preprocessing, the vascular structures were successfully highlighted and the background was restrained. The extraction of vessel contours (b) obtained vascular contours accurately and completely. The improved adaptive tracking method (c) is a core part of our work: compared with the original adaptive tracking method of [38], one of the main improvements in this approach is the bifurcation point detection part. We changed the originally fixed search radius to a proper search scope, which enhanced the capacity of the retrieval of bifurcation, and we used four conditions in Equation (17) to judge bifurcation point instead of only using the first condition, which greatly improved the detection accuracy and reduced the misjudgment. The results can be seen in Figure 8.

To achieve the completely automatic identification of vessels, we need to test the robustness of the proposed tracking method for randomly selected seed points from the detected ridge points. Taking Figure 6(c) as an example, three seed points were selected from different positions. The experimental results are shown in Figure 9. It can be seen that the results of the proposed method have strong robustness; that is, our method is generally applicable for automatically selected seed points. Meanwhile, our method is more accurate than the method of [38], which is clear in Figure 9 that the points of different types we identified are more approaching to the real vessel.

Even though our method has an improvement in accuracy and robustness compared to the former one, it still has some shortcomings. For example, the image preprocessing method is not effective enough for some images with complex vascular structures. Although the detection of bifurcation points has been improved compared with the method in [38], there are still a few misjudgments. This phenomenon can be seen in Figure 8(b). In the case of a more complex vascular structure, the tracking effect varies with the selection of seed points, and some vessel segments may be lost, as shown in Figure 7(c2).

6. Conclusion

In this paper, we designed a scheme of image preprocessing, used the C-V model, and proposed an improved adaptive tracking method, with which we can realize segmentation and automatic identification of vessels in coronary angiograms. Among these methods, the improved adaptive tracking method contains our major innovations that can enhance the capability of identifying vessels. Besides, we did many experiments to test our proposed method and the results turned out that our method is more robust and accurate than the former method.

Due to the complexity of coronary angiograms described above, traditional image processing methods are not effective enough. Hence, in the following work, we will continue to optimize the tracking algorithm and carry out image process research on deep learning to achieve a better effect.

Data Availability

The data supporting this study is from Qilu Hospital (Qingdao) of Shandong University.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yaofang Liu, Wenlong Wan, and Xinyue Zhang contributed equally.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 11771408 and 11871444), the Key Research and Development Project of Shandong Province (No. 2015GSF118026), the Qingdao Key Health Discipline Development Fund, and the People's Livelihood Science and Technology Project of Qingdao (No. 18-6-1-62-nsh).

References

[1] S. Z. Abildstrøm and M. Madsen, "The Danish heart register," *Scandinavian Journal of Public Health*, vol. 39, 7 Supplement, pp. 46–49, 2011.

[2] D. Rueckert and P. Burger, "Contour fitting using stochastic and probabilistic relaxation for cine MR images," *Computer Assisted Radiology*, vol. 137, pp. 137–142, 1995.

[3] C. Pellot, A. Herment, M. Sigelle, P. Horain, H. Maitre, and P. Peronneau, "A 3d reconstruction of vascular structures from two x-ray angiograms using an adapted simulated annealing algorithm," *IEEE Transactions on Medical Imaging*, vol. 13, no. 1, pp. 48–60, 1994.

[4] A. J. Bulpitt and E. Berry, "Spiral ct of abdominal aortic aneurysms: comparison of segmentation with an automatic 3d deformable model and interactive segmentation," *Medical Imaging 1998: Image Processing*, 1998.

[5] D. Rueckert, P. Burger, S. M. Forbat, R. D. Mohiaddin, and G. Z. Yang, "Automatic tracking of the aorta in cardiovascular

MR images using deformable models," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 581–590, 1997.

[6] T. Song, S. Pang, S. Hao, A. Rodríguez-Patón, and P. Zheng, "A parallel image skeletonizing method using spiking neural P systems with weights," *Neural Processing Letters*, vol. 50, no. 2, pp. 1485–1502, 2019.

[7] R. Toledo, X. Orriols, P. Radeva et al., "Eigensnakes for vessel segmentation in angiography," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Barcelona, Spain, September 2000.

[8] N. Armande, P. Montesinos, O. Monga, and G. Vaysseix, "Thin nets extraction using a multi-scale approach," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 248–257, 1999.

[9] D. L. Wilson and J. A. Noble, "Segmentation of cerebral vessels and aneurysms from MR angiography data," *International Conference on Information Processing in Medical Imaging*, 1997.

[10] J. F. O'Brien and N. F. Ezquerra, "Automated segmentation of coronary vessels in angiographic image sequences utilizing temporal, spatial and structural constraints," *Proceedings of SPIE-the International Society for Optical Engineering*, vol. 2359, 1994.

[11] G. Ayala, T. Leon, and V. Zapater, "Different averages of a fuzzy set with an application to vessel segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 3, pp. 384–393, 2005.

[12] S. A. Stansfield, "ANGY: a rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 2, pp. 188–199, 1986.

[13] J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation," *Graphical Models and Image Processing*, vol. 58, no. 3, pp. 246–261, 1996.

[14] S. Shiffman, G. D. Rubin, and Y. Napel, "Semiautomated editing of computed tomography sections for visualization of vasculature," *Proceedings of SPIE-The International Society for Optical Engineering*, vol. 2707, pp. 140–151, 1996.

[15] T. Song, X. Zeng, P. Zheng, M. Jiang, and A. Rodriguez-Paton, "A parallel workflow pattern modeling using spiking neural P systems with colored spikes," *IEEE Transactions on Nano Bioscience*, vol. 17, no. 4, pp. 474–484, 2018.

[16] P. Sahoo, C. Wilkins, and J. Yeager, "Threshold selection using Renyi's entropy," *Pattern Recognition*, vol. 30, no. 1, pp. 71–84, 1997.

[17] Y. Sato, S. Nakajima, N. Shiraga et al., "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis*, vol. 2, no. 2, pp. 143–168, 1998.

[18] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.

[19] X. Chen, Y. Lu, J. Bai et al., "Train a 3D U-Net to segment cranial vasculature in CTA volume without manual annotation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 559–563, Washington, DC, USA, April 2018.

[20] O. Wink, W. J. Niessen, and M. A. Viergever, "Multiscale vessel tracking," *IEEE Transactions on Medical Imaging*, vol. 23, no. 1, pp. 130–133, 2004.

[21] M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1531–1546, 2006.

[22] D. Zhang, C. Wang, and S. Zhou, "A new method of vessel centerline extraction from 3D CT coronary angiography based on open-snake," in *2015 IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015)*, Beijing, China, 2015.

[23] T. Song, L. Pan, T. Wu, P. Zheng, M. L. D. Wong, and A. Rodriguez-Paton, "Spiking neural P systems with learning functions," *IEEE Transactions on Nano Bioscience*, vol. 18, no. 2, pp. 176–190, 2019.

[24] D. Barbuzzi, A. Massaro, A. Galiano, L. Pellicani, G. Pirlo, and M. Saggese, "Multi-domain intelligent system for document image retrieval," *International Journal of Adaptive and Innovative Systems*, vol. 2, no. 4, pp. 282–297, 2019.

[25] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys*, vol. 36, no. 2, pp. 81–121, 2004.

[26] S. R. Aylward and E. Bullitt, "Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 61–75, 2002.

[27] F. Zamani Boroujeni, R. W. O. K. Rahmat, N. Mustapha, L. S. Affendey, and O. Maskon, "Coronary artery center-line extraction using second order local features," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 940981, 20 pages, 2012.

[28] K. Jain and A. Kumar, "An optimal RSSI-based cluster-head selection for sensor networks," *International Journal of Adaptive and Innovative Systems*, vol. 2, no. 4, pp. 349–361, 2019.

[29] Z. Shoujun, Y. Jian, W. Yongtian, and C. Wufan, "Automatic segmentation of coronary angiograms based on fuzzy inferring and probabilistic tracking," *Biomedical Engineering Online*, vol. 9, no. 1, pp. 40–40, 2010.

[30] R. Xiao, J. Yang, D. Ai et al., "Adaptive ridge point refinement for seeds detection in x-ray coronary angiogram," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 502573, 10 pages, 2015.

[31] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[32] D. F. Malin, "Unsharp masking," *AAS Photo Bulletin*, vol. 16, pp. 10–13, 1977.

[33] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Delhi, India, September 2014.

[34] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1998.

[35] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[36] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[37] T. F. Chan, B. Y. Sandberg, and L. A. Vese, "Active contours without edges for vector-valued images," *Journal of Visual Communication and Image Representation*, vol. 11, no. 2, pp. 130–141, 2000.

[38] R. Xiao, J. Yang, M. Goyal, Y. Liu, and Y. Wang, "Automatic vasculature identification in coronary angiograms by adaptive geometrical tracking," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 796342, 11 pages, 2013.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 2602688, 10 pages
https://doi.org/10.1155/2021/2602688

Research Article

Automatic Assessment of Mitral Regurgitation Severity Using the Mask R-CNN Algorithm with Color Doppler Echocardiography Images

Qinglu Zhang,[1,2,3] Yuanqin Liu,[4] Jia Mi,[2] Xing Wang,[2] Xia Liu,[2] Fenfen Zhao,[2] Cuihuan Xie,[2] Peipei Cui,[2] Qingling Zhang,[1,3,5] and Xiangming Zhu[1,3,5]

[1]*Anhui Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong 250012, China*
[2]*Department of Special Examination, Shandong Provincial Third Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong 250031, China*
[3]*Department of Ultrasound, The First Affiliated Hospital of Wannan Medical College, Wuhu, Anhui 241001, China*
[4]*Department of Neurosurgery, The First Affiliated Hospital of Shandong First Medical University, Jinan, Shandong 250014, China*
[5]*Anhui Provincial Hospital, Hefei, Anhui 230001, China*

Correspondence should be addressed to Qingling Zhang; qingling.go@gmail.com and Xiangming Zhu; zhuxmyjsyy@163.com

Received 2 August 2021; Accepted 18 August 2021; Published 13 September 2021

Academic Editor: Tao Song

Accurate assessment of mitral regurgitation (MR) severity is critical in clinical diagnosis and treatment. No single echocardiographic method has been recommended for MR quantification thus far. We sought to define the feasibility and accuracy of the mask regions with a convolutional neural network (Mask R-CNN) algorithm in the automatic qualitative evaluation of MR using color Doppler echocardiography images. The authors collected 1132 cases of MR from hospital A and 295 cases of MR from hospital B and divided them into the following four types according to the 2017 American Society of Echocardiography (ASE) guidelines: grade I (mild), grade II (moderate), grade III (moderate), and grade IV (severe). Both grade II and grade III are moderate. After image marking with the LabelMe software, a method using the Mask R-CNN algorithm based on deep learning (DL) was used to evaluate MR severity. We used the data from hospital A to build the artificial intelligence (AI) model and conduct internal verification, and we used the data from hospital B for external verification. According to severity, the accuracy of classification was 0.90, 0.89, and 0.91 for mild, moderate, and severe MR, respectively. The Macro F1 and Micro F1 coefficients were 0.91 and 0.92, respectively. According to grading, the accuracy of classification was 0.90, 0.87, 0.81, and 0.91 for grade I, grade II, grade III, and grade IV, respectively. The Macro F1 and Micro F1 coefficients were 0.89 and 0.89, respectively. Automatic assessment of MR severity is feasible with the Mask R-CNN algorithm and color Doppler electrocardiography images collected in accordance with the 2017 ASE guidelines, and the model demonstrates reasonable performance and provides reliable qualitative results for MR severity.

1. Introduction

Mitral regurgitation (MR) is a common valvular heart condition. A study by the 2016 American Heart Association (AHA) in the USA estimated that the incidence rate of moderate or worse MR is 1.7%, which is approximately 4-fold higher than that of aortic stenosis [1]. Furthermore, the incidence increases with age, and the proportion can reach 10% in the population over 75 years old [2]. The therapeutic method varies based on the degree of MR. According to the Society of Thoracic Surgeons national database, the number of mitral valve surgeries increased by an average of 4% every year between 2010 and 2015. When deciding which patients are suitable for mitral valve (MV) surgery, the guidelines of the American College of Cardiology (ACC) and AHA for the management of valvular heart

disease emphasize the severity of MR [3]. Thus, accurate assessment of MR severity is crucial for clinical decision-making, prognostication, and decisions regarding the timing of surgical intervention [4]. Transthoracic echocardiography (TTE) is the most important imaging method for MR diagnosis and evaluation due to its widespread availability, low cost, acceptability, and safety profile [5]. However, the MR evaluation parameters listed in the 2017 American Society of Echocardiography (ASE) guidelines are numerous and complex and are very challenging to use in practice [6]. There is currently no single recommended MR evaluation method in this setting. Herein, we attempt to validate a convenient and automatic method for evaluating MR severity.

Since John MacCarthy first proposed "artificial intelligence (AI)" in 1956, researchers have made great efforts to apply AI to almost all stages of clinical practice. At present, the development of AI in the field of ultrasound medicine to improve the accuracy of ultrasound diagnosis, reduce the misdiagnosis rate, and meet growing clinical needs is a hot research topic. Deep learning (DL) is a subset of AI inspired by the workings of the human brain, commonly referred to as an artificial neural network (ANN) [7]. Convolutional neural networks (CNNs) are a subtype of ANNs that mimic the visual cortex. Regions with CNN features (R-CNN) apply CNNs in object detection. To improve efficiency, Fast R-CNN combines the feature extraction, classification, and bounding box prediction of R-CNN and incorporates a method called region of interest pooling (RoIPool) [8]. Then, researchers developed Faster R-CNN, which has similar accuracy to Fast R-CNN, but the training time and testing time are 10 times shorter. He et al. proposed a new method called Mask R-CNN in 2017, which expands Faster R-CNN by adding branches used to predict the segmentation mask on each of the RoIs classified with existing branches and border frame returns [9]. Compared to Faster R-CNN, the mask branch only adds a small computational overhead, enabling a fast system and rapid experimentation. Thus, our study chose Mask R-CNN algorithm. Such an AI system has great potential for effective improvement of diagnosis.

We aimed to evaluate the feasibility and accuracy of MR severity detection with AI data models using MR color Doppler echocardiography images collected based on the 2017 ASE guidelines.

2. Methods

2.1. Establishment of the MR Color Doppler Echocardiography Case Database

2.1.1. Ultrasound Instrument. Ultrasound was performed by echocardiographers using a Philips ultrasound machine (EPIQ 7C, Philips Medical Systems, Bothell, WA), GE ultrasound machine (VIVID E95, GE Medical Systems, Horten, Norway), and Siemens ultrasound machine (SC2000, Siemens Medical Solutions USA, Inc.).

2.1.2. Patients and MR Image Classification Standard. This study involved data from two large general hospitals in different regions. Hospital A is Shandong Provincial Third Hospital, and hospital B is The First Affiliated Hospital of Wannan Medical College. The Institutional Review Boards of the two hospitals approved this study protocol and waived the need for informed consent due to a minimal potential for harm.

All echocardiographers were well experienced, had worked more than five years, and had undergone thorough professional training before the study. According to the quantitative methods of MR evaluation from the 2017 ASE guidelines (see Figure 1), the severity of MR can be classified into three types: mild, moderate, and severe. This classification is relatively broad and cannot well reflect the severity of MR. Then, MR was further subclassified into four grades: grade I (mild), grade II (moderate), grade III (moderate), and grade IV (severe). A total of 1132 and 295 MR cases were collected from hospital A and hospital B, respectively, from January 2019 to December 2020. There were a similar number of cases for each grade. The 2017 ASE guidelines provide distinct criteria for the classification of chronic MR using color Doppler echocardiography: vena contracta (VC), effective regurgitant orifice (ERO), regurgitant volume (RVol), and regurgitation fraction (RF) [6]. VC is a parameter used for determination of the regurgitant orifice. To obtain the VC, we measure the narrowest width of the jet as it emerges from the orifice in zoom mode on the long axis view of the sternum. When determining the ERO, it is important to carefully measure the proximal isovelocity surface area (PISA) and obtain the greatest PISA radius at the time of peak MR velocity. To obtain the most hemispheric flow convergence, we adjust the lower Nyquist limit to 30-40 cm/sec. The Nyquist limit should be set at 50-70 cm/sec when measuring RF. RVol is measured in the case of multiple jets or eccentric jets, as it is more accurate. Color Doppler echocardiography images are acquired from the standard two-dimensional (2D) apical 4-chamber view of TTE or the standard view with the most regurgitation.

2.1.3. Exclusion Criteria. Cases were excluded if the image quality was very poor or TTE images could not be clearly displayed.

2.2. Image Marking. The LabelMe software (3.167) was used to demarcate the region of interest (RoI) in MR ultrasound images for automatic analysis by machine DL technology. The workflow of LabelMe is shown in Figure 2. At the "Annotation" step, tracing the contour of MR, the more accurate the better (see Figure 3).

2.3. Establishment and Validation of the Data Model

2.3.1. Network Architecture. Mask R-CNN is a method of object detection and segmentation that can distinguish different objects in images and draw bounding boxes (bbox) around specific objects. It can also mark and classify targets and identify other detection key points. The network architecture was constructed in the Google TensorFlow framework, and the network architecture of the Mask R-CNN algorithm is illustrated in Figure 4. We defined a multitask loss on each sampled RoI as $L = L_{\text{cls}} + L_{\text{bbox}} + L_{\text{mask}}$. The

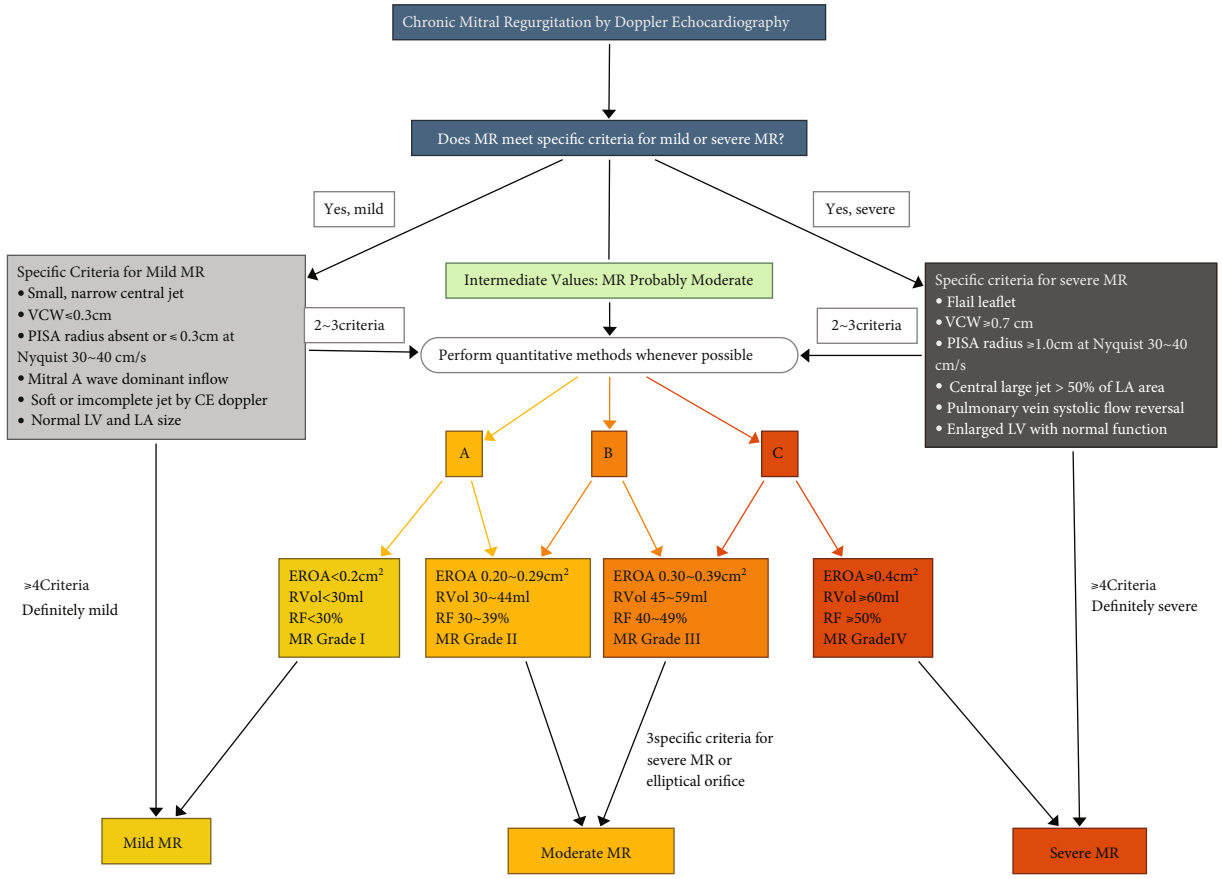


Figure 1: Algorithm for the integration of multiple parameters of MR severity by Doppler echocardiography adapted from ASE 2017.

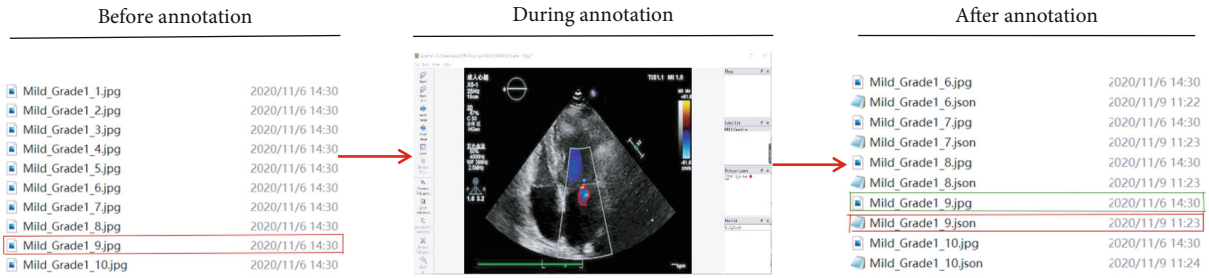


Figure 2: The workflow of LabelMe. (a) Create dataset on our local computer. (b) Perform annotation and save annotation results on each image by pressing "Save" button. (c) For each image, get a.json file, which contains for the labels created.

classification loss (L_{cls}) and bounding box loss (L_{bbox}) were identical to those defined in Faster R-CNN [8].

$$L_{cls} = \frac{1}{N_{cls}} \sum_{i} L_{cls}(pi, p_i^*), \tag{1}$$

$$L_{bbox} = \frac{1}{N_{box}} \sum_{i} p_i^* L_1^{smooth}(ti - t_i^*), \tag{2}$$

$$L_{cls}(\{pi, p_i^*\}) = -p_i^* \log p_i^* - (1 - p_i^*) \log (1 - p_i^*). \tag{3}$$

L_{mask} is the average binary cross-entropy loss.

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \le i,j \le m} \left[y_{ij} \log oy_{ij}^k + \left(1 - y_{ij}^k\right) \log \left(1 - oy_{ij}^k\right) \right]. \tag{4}$$

The loss function value (L), $L_{cls} + L_{bbox} + L_{mask}$, in the Mask R-CNN was minimized.

2.3.2. Network Training and Testing. The Mask R-CNN was trained using the MR ultrasound images. The MR images acquired from hospital A made up dataset A, and the MR

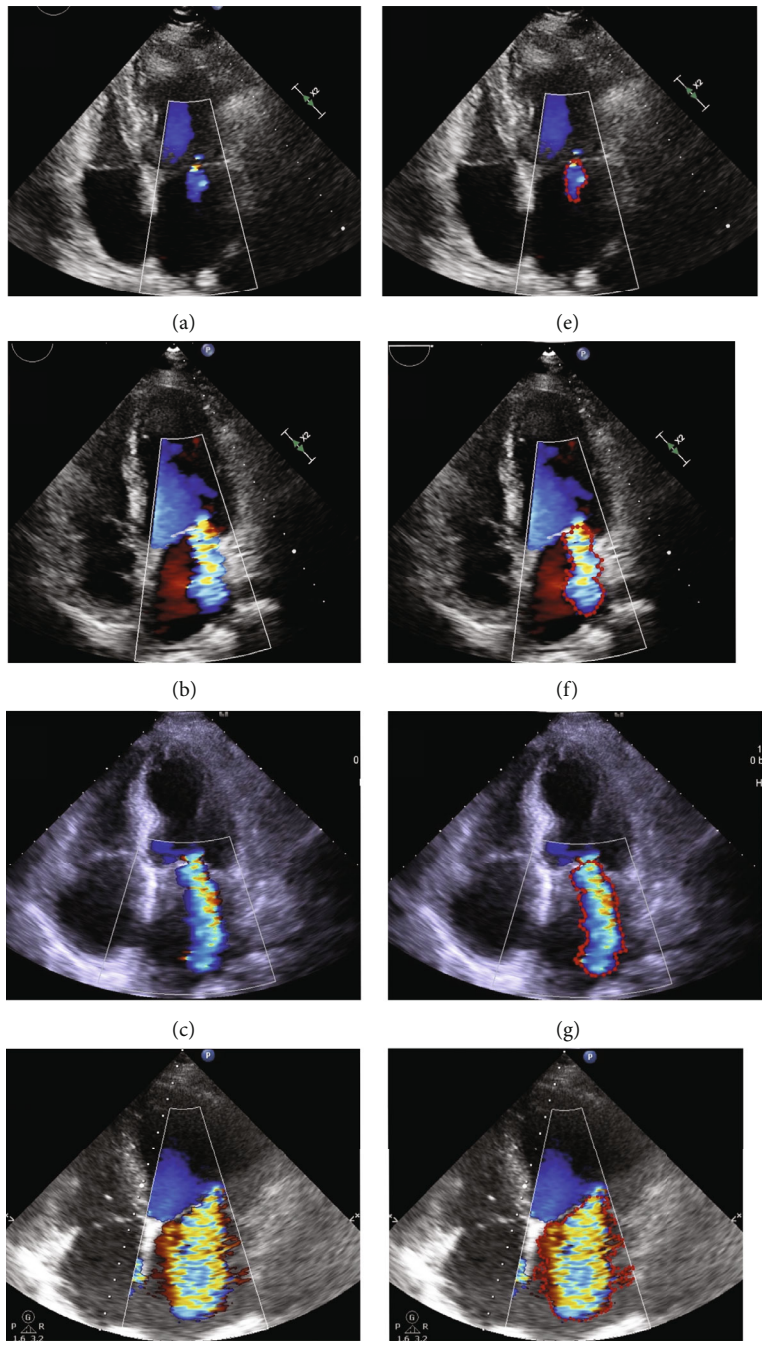


Figure 3: Example of MR contour. (a–d) The original image of MR of grade I, grade II, grade III, and grade IV. (e–h) The contour map of the MR traced by the LabelMe software corresponding to (a–d).

images acquired from hospital B made up dataset B. Dataset A was used for training of the AI model. To ensure the accuracy and stability of the model, we used dataset B to verify the model. The ratio of dataset A to dataset B is approximately 8 : 2. The ratio of each grade in the two datasets is also approximately 8 : 2. The trained model was applied for prediction in the test set. The training parameters were set as follows: For the backbone and region proposal network (RPN), learning rate was 0.001; for the R-CNN and Mask heads, learning rate was 0.0001. Throughout the training process, the momentum was set to 0.9 and the stochastic gradient descent optimizer was used. The learning rate and momentum were set by monitoring the loss during training. With a low learning rate, the improvements will be linear.

2.3.3. Evaluation Metrics. The overall performance of the AI model for the assessment of MR severity was validated with accuracy, precision, recall, F1-score, Macro F1, and Micro F1.

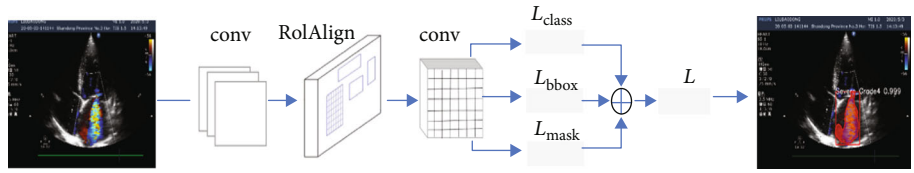


Figure 4: The network architecture of Mask R-CNN.

Table 1: Medical history, clinical presentation, and baseline TTE characteristics of the study patients.

	Age (Yrs)	Male	Medical history and clinical presentation						Baseline TTE characteristics					Time (minutes)
			CHD	MI	HTN	DM	TIA	NYHA≥III	LVDd, cm	LVDs, cm	LA, cm	EF, %	Reduced EF, <50%	
Mild (grade I)	61 (29-82)	126	59	14	56	41	29	19	4.3 ± 0.57	3.1 ± 0.59	3.4 ± 0.52	58 (46-67)	7	5.8 ± 4.5
Moderate (grade II)	68 (31-88)	117	102	51	82	37	27	47	4.4 ± 0.61	3.3 ± 0.58	3.5 ± 0.69	57 (44-66)	19	10.6 ± 2.7
Moderate (grade III)	71 (41-90)	121	125	62	103	38	44	91	5.1 ± 0.53	37 ± 0.65	4.2 ± 0.71	50 (36-63)	55	11.1 ± 2.4
Severe (grade IV)	73 (43-92)	135	140	71	134	47	48	135	5.5 ± 0.70	42 ± 0.87	4.5 ± 0.73	44 (30-61)	98	5.5 ± 4.4

CHD: coronary artery heart disease; MI: myocardial infarction; HTN: hypertension; DM: diabetes mellitus; TIA: transient ischemic attack; NYHA: New York Heart Association; time: the time taken to quantitatively evaluate the severity of MR according to the 2017 ASE guidelines. Values are median (interquartile range), mean ± SD, or *n* (%), unless otherwise indicated.

Accuracy is the ratio of the number of examples consistent with the results of the 2017 ASE guidelines and the total number of examples.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{6}$$

$$\text{F1} - \text{score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \tag{7}$$

Macro F1. Split the evaluations of *n* categories into *n* two-category evaluations, calculate the F1-score of each two-category, and the average value of the n F1-scores is Macro F1.

Micro F1. Divide the evaluations of n categories into *n* two-category evaluations, and add the corresponding TP, FP, and RN of the *n* two-category evaluations to calculate the precision and recall. The F1-score calculated from these precision and recall is Micro F1.

TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

3. Results

In this study, 1132 MR ultrasound images (288 grade I, 278 grade II, 270 grade III, and 296 grade IV) in dataset A and 295 MR ultrasound images (82 grade I, 75 grade II, 74 grade III, and 64 grade IV) in dataset B were finally applied. The baseline demographic and TTE characteristics of the study patients are summarized in Table 1.

Figure 5 shows the model performance evaluation metrics and results. The total loss was 0.0493, the bbox loss was 0.0055, the class loss was 0.0012, and the mask loss was 0.0427.

Figure 6 shows four test examples for the assessment of MR severity. Figures 6(a)–6(d) are graded MR images obtained by the evaluation methods described in the 2017 ASE guidelines. Figures 6(e)–6(h) are the results of the test using this AI model, which are consistent with the results obtained by the evaluation methods described in the 2017 ASE guidelines.

Figure 7 shows the confusion matrix of the MR classification and grading results for the validation. The accuracy of classification according to severity was 0.90, 0.89, and 0.91 for mild, moderate, and severe MR, respectively. The accuracy of classification according to grade was 0.90, 0.87, 0.81, and 0.91 for grade I, grade II, grade III, and grade IV, respectively.

Figure 8 shows the comparative histograms of precision, recall, and F1-score between classification indexes (Figure 8(a) is the classification according to severity, and Figure 8(b) is the classification according to grading). The precision of classification according to severity was 0.94, 0.93, and 0.87 for mild, moderate, and severe MR, respectively. The precision of classification according to grade was 0.94, 0.88, 0.88, and 0.87 for grade I, grade II, grade III, and grade IV, respectively. The recall of classification according to severity was 0.94, 0.90, and 0.92 for mild, moderate, and severe MR, respectively. The recall of classification according to grade was 0.94, 0.89, 0.82, and 0.92 for grade I, grade II, grade III, and grade IV, respectively. The F1-score of classification according to severity was 0.94, 0.91, and 0.89 for mild, moderate, and severe MR, respectively. The F1-score of

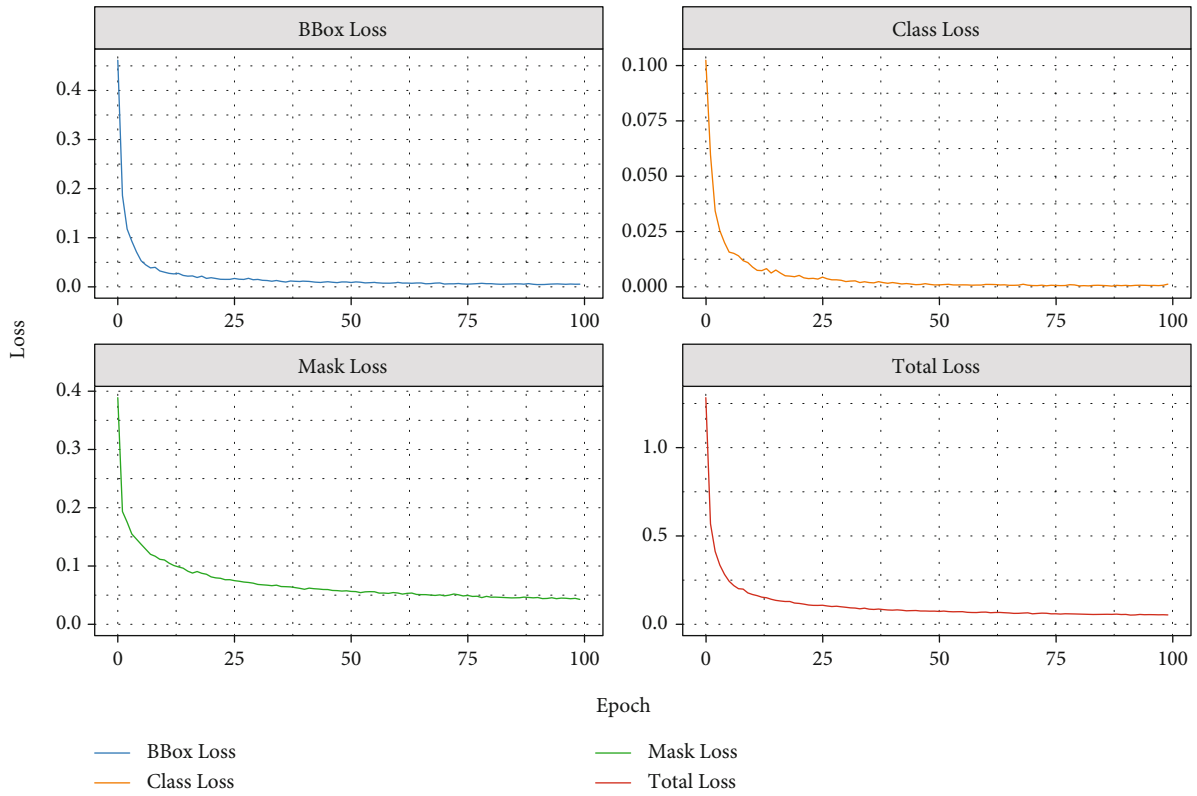


Figure 5: Model performance evaluation metric and results. The abscissa axis is epoch, and the ordinate axis is loss. These four ordinates represent bbox loss, class loss, mask loss, and total loss, respectively.

classification according to grade was 0.94, 0.88, 0.85, and 0.89 for grade I, grade II, grade III, and grade IV, respectively. It can be observed that this model produces satisfactory precision, recall, and F1-score results in the evaluation of MR severity.

Table 2 shows the comparison results of Macro F1 and Micro F1 in each classification. This shows a satisfactory classification result.

4. Discussion

We validated the Mask R-CNN algorithm for the evaluation of MR severity. The present study demonstrated the feasibility and accuracy of the Mask R-CNN algorithm for qualitative assessment of MR and demonstrated the reasonable performance of the model.

TTE is the most common imaging technique by which MR severity and etiology are determined. Although many recent studies have shown that 2D technology is not the most accurate method for quantitatively evaluating MR, the 2D TTE technique is currently the most commonly used method for quantitatively evaluating MR compared with cardiac magnetic resonance (CMR), transesophageal echocardiography (TEE), and the 3D TTE technique [10]. However, there is currently no single echocardiographic parameter that is precise enough to quantify MR. Integration of multiple parameters is required for a more accurate assessment of MR severity [11]. When multiple parameters are concordant, MR severity, especially mild and severe MR, can be determined with high confidence. In our study, all MR grades were determined independently by two well-experienced echocardiographers according to the 2017 ASE guidelines. It is necessary to emphasize that when there is consistent evidence from different parameters, it is easy to grade MR severity with confidence. When different parameters are contradictory, one must look carefully for technical and physiologic factors to explain the discrepancies and repeat the measurements according to the 2017 ASE guidelines. If the discrepancy remained, a third investigator's recommendation was used as a reference. Errors in measurement can be prevented.

AI is a powerful technological driving force at present. Increasing efforts have been made by medical ultrasound experts, mathematicians, and computer scientists to promote the integration of ultrasound, medicine, and AI, thereby improving the accuracy of ultrasonic diagnosis, reducing the misdiagnosis rate, shortening the reporting time, and meeting growing clinical needs [12].

AI has made some progress in the assessment of MR; here, we review some recent studies. Many studies of MR diagnosis have been carried out to investigate heart sounds (HSs). Maglogiannis et al. used Doppler heart sound (DHS) data with wavelet decomposition followed by a three-step diagnosis phase based on support vector machine (SVM) classifier to classify heart valve disease. The reported accuracy for aortic stenosis (AS) and MR classification is 91.67% [13]. Safara et al. developed a multilevel basis selection (MLBS) method with an SVM classifier to classify

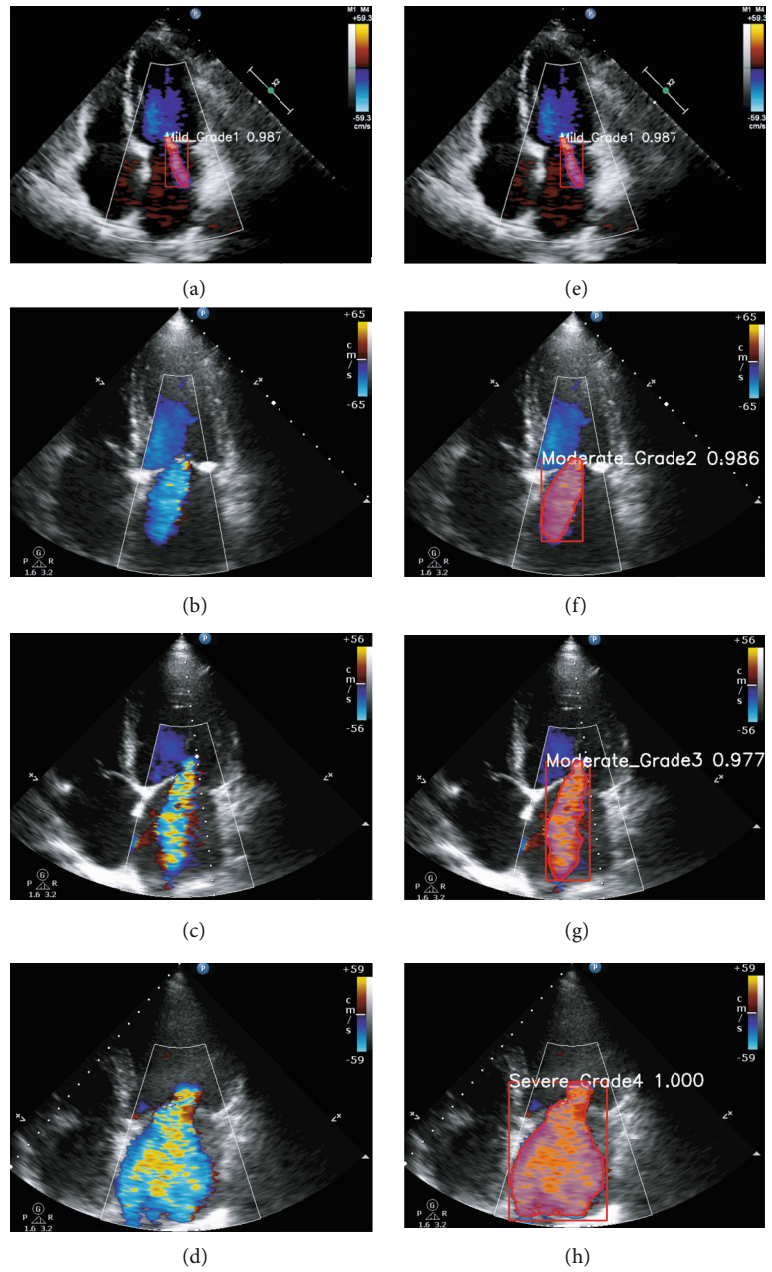


Figure 6: Test example of assessment of MR severity. (a–d) The MR images obtained by the quantitative evaluation methods of 2017 ASE guidelines. (e–h) The grading result obtained by the Mask R-CNN algorithm. The confident scores for the four cases were 0.987, 0.986, 0.997, and 1.000, respectively.

normal AS MR and AR samples, and the accuracy of classification was 97.56% [14]. There are some other AI studies on the detection of MR. An intelligent diagnostic system based on automatic diagnostic feature extraction for diagnosing heart diseases developed by Sun could discriminate MR with an accuracy of 98.4% [15]. Kwon, MD, and colleagues developed and validated an AI algorithm for detecting MR using electrocardiography (ECG); they demonstrated a promising performance of the AI algorithm for accurate MR detection. During the internal and external validation, the accuracy of MR detection was 0.816 and 0.877, respectively [16].

However, the abovementioned studies only used AI algorithms to detect MR, and there were no further qualitative studies. Recently, some studies have focused on detecting the severity of MR using automatic detection methods. Moghaddasi and Nourian developed a novel method for grading MR according to novel textural features with machine learning methods [17]. The proposed method achieved satisfactory accuracy for the detection of MR severity in normal subjects. This method is based on echocardiography videos. In their study, MR was graded into three types: mild, moderate, and severe. They did not further subdivide moderate MR into grade II and grade III. This does not reflect the severity of MR well. Studies by Uretsky et al. highlighted the accuracy and reproducibility of CMR in quantifying MR and have begun to link CMR to clinical

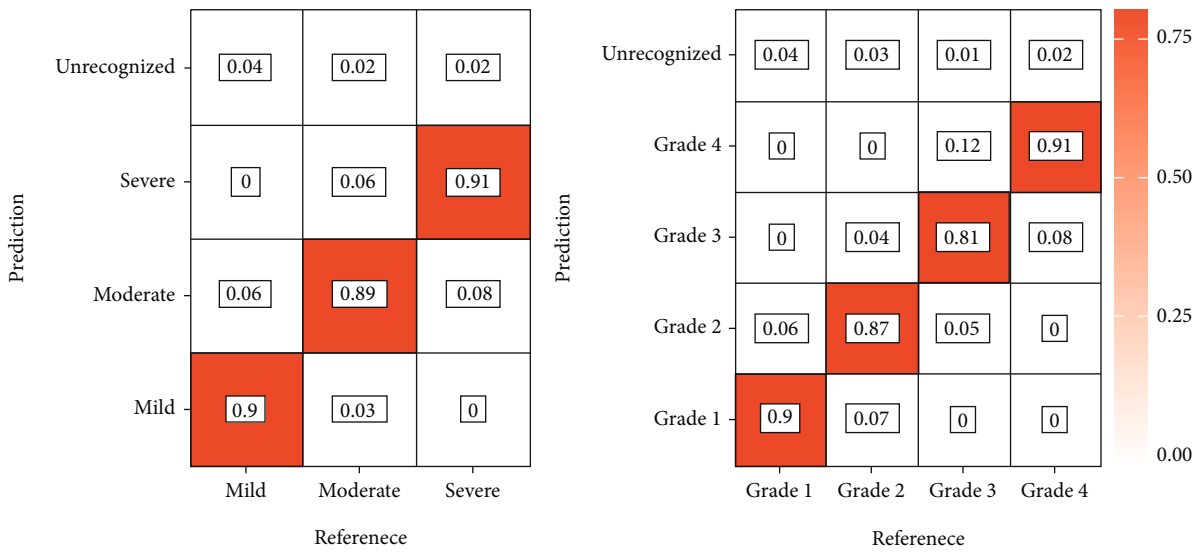


Figure 7: The confusion matrix of classification results and grading results of MR. The abscissa axis is the reference results obtained through the 2017 ASE guidelines, and the ordinate axis represents the results of this research model. Unrecognized means that the model cannot recognize this MR.

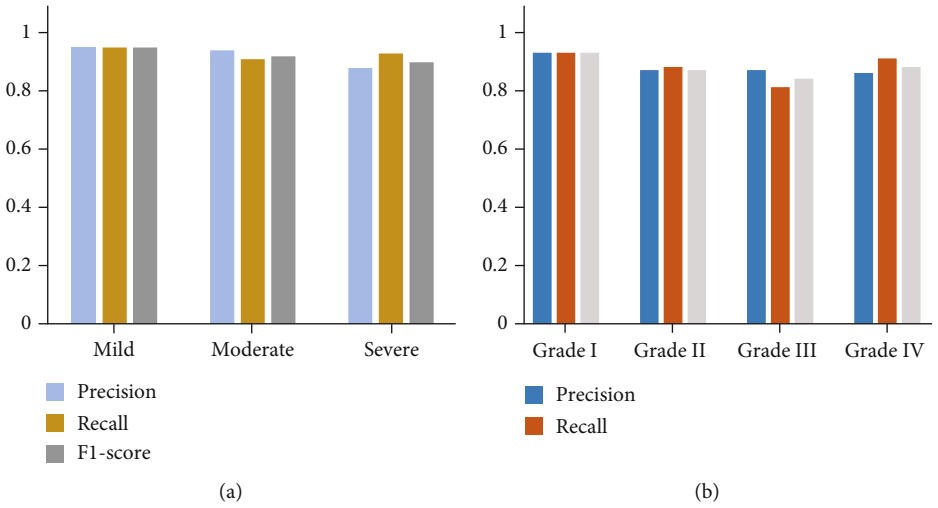


Figure 8: Comparative histograms of precision, recall, and F1-score between classification indexes ((a) classification according to severity; (b) classification according to grading).

Table 2: The comparison results of Macro F1 and Micro F1 in each classification.

Classification	Macro F1	Micro F1
According to severity	0.91	0.92
According to grading	0.89	0.89

outcomes [18]. However, in our daily practice, CMR is not widely available and is time-consuming. Moreover, in some emergency situations, CMR cannot be the first choice, and there are contraindications for it in some patients. Some studies have also pointed out that the degree of MR measured by TEE is more accurate than that measured by TTE. Militaru et al. evaluated the accuracy of MR volume quantified with 3D color Doppler TEE using new semiautomated software. The new software enabled semiautomated 3D MR flow quantification in complex MR with multiple eccentric jets and showed a satisfactory result [19]. However, TEE is operator dependent and semi-invasive, typically requiring patient sedation [20]. It is not suitable for routine examinations.

In our study, when classifying according to severity, we achieved accuracies of 0.90, 0.89, and 0.91, and when classifying according to grading, we achieved accuracies of 0.90, 0.87, 0.81, and, 0.91. Among the grading classifications, grade III has the lowest accuracy, which is mostly because the characteristics of grade III have some overlap with the characteristics of severe MR. In model verification, the unrecognized rate of grade I reached 0.04, which is probably

because the VC in some images of grade I is too small to be identified. Our model also obtained better precision, recall, F1-score, Macro F1, and Micro F1. All these suggest that our model has good performance. In the process of collecting cases, the quantitative methods for MR identification in the 2017 ASE guidelines were time-consuming, and for each case, it took a few minutes to take the pictures required to obtain the results. Grade I and grade IV take less than 10 minutes to classify; however, grade II and grade III take more than 10 minutes (see Table 1). This is because when VCW ≤ 0.3 cm, VCW ≥ 0.7 cm, or some other obvious condition is present (Figure 1), it is easy to determine whether MR is mild or severe, and no further evaluation is needed. In contrast, assessing MR severity with our AI model requires a shorter amount of time, which could greatly reduce working time. This can significantly improve the work efficiency of clinicians.

In this study, we designed an experimental dataset and a validation dataset. Hospital A and hospital B are in different regions, and both hospitals are large tertiary general hospitals. This can effectively address the influence of regional differences. Three commonly used and well-known brands of ultrasound machines were used, so the accuracy and quality of performance were good. The results prove that our AI model is universally applicable and has good performance and high accuracy. More importantly, it greatly shortens the diagnosis time. Due to these advantages, this AI model has the potential to be used for diagnosis in daily clinical practice.

5. Conclusions

Accurate assessment of the severity of MR is crucial in clinical treatment. In this study, we chose the Mask R-CNN algorithm to qualitatively evaluate MR using color Doppler echocardiography images collected based on the 2017 ASE guidelines. This demonstrated that the model has good performance and could evaluate the severity of MR with good accuracy. Thus, with the combination of MR echocardiography images and DL, the time required to analyze cardiac-related parameters is decreased, and clinical decision-making can be expedited. This model can serve as a new tool for the evaluation of MR severity.

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.

Authors' Contributions

Qinglu Zhang, Yuanqin Liu, and Jia Mi contributed equally to this study and share first authorship.

Acknowledgments

This work is supported by the Climb Science Foundation of the Hospital-Level Project of The First Affiliated Hospital of Wannan Medical College (PF2019008, applicant: Xiangming Zhu).

References

[1] Writing Group Members, D. Mozaffarian, E. J. Benjamin et al., "Heart disease and stroke Statistics-2016 update: a report from the American Heart Association," *Circulation*, vol. 133, no. 4, pp. e38–e360, 2016.

[2] E. J. Benjamin, P. Muntner, A. Alonso et al., "Heart disease and stroke statistics-2019 update: a report from the American Heart Association," *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019.

[3] V. Badhwar, J. S. Rankin, J. P. Jacobs et al., "The Society of Thoracic Surgeons adult cardiac surgery database: 2016 update on research," *The Annals of Thoracic Surgery*, vol. 102, no. 1, pp. 7–13, 2016.

[4] P. Thavendiranathan, D. Phelan, J. D. Thomas, S. D. Flamm, and T. H. Marwick, "Quantitative assessment of mitral regurgitation: validation of new methods," *Journal of the American College of Cardiology*, vol. 60, no. 16, pp. 1470–1483, 2012.

[5] R. A. Nishimura, C. M. Otto, R. O. Bonow et al., "2014 AHA/ACC guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines," *Journal of the American College of Cardiology*, vol. 63, no. 22, pp. e57–e185, 2014.

[6] W. A. Zoghbi, D. Adams, R. O. Bonow et al., "Recommendations for noninvasive evaluation of native valvular regurgitation a report from the American Society of Echocardiography developed in collaboration with the Society for Cardiovascular Magnetic Resonance," *Journal of the American Society of Echocardiography*, vol. 30, no. 4, pp. 303–371, 2017.

[7] N. Gahungu, R. Trueick, S. Bhat, P. P. Sengupta, and G. Dwivedi, "Current challenges and recent updates in artificial intelligence and echocardiography," *Current Cardiovascular Imaging Reports*, vol. 13, no. 2, p. 5, 2020.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[9] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2020.

[10] H. Baumgartner, V. Falk, J. J. Bax et al., "2017 ESC/EACTS guidelines for the management of valvular heart disease," *European Heart Journal*, vol. 38, no. 36, pp. 2739–2791, 2017.

[11] E. Avenatti, G. B. Mackensen, K. C. El-Tallawi et al., "Diagnostic value of 3-dimensional vena contracta area for the quantification of residual mitral regurgitation after MitraClip procedure," *JACC. Cardiovascular Interventions*, vol. 12, no. 6, pp. 582–591, 2019.

[12] J. Jeganathan, Z. Knio, Y. Amador et al., "Artificial intelligence in mitral valve analysis," *Annals of Cardiac Anaesthesia*, vol. 20, no. 2, pp. 129–134, 2017.

[13] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, "Support vectors machine-based identification of heart valve

diseases using heart sounds," *Computer Methods and Programs in Biomedicine*, vol. 95, no. 1, pp. 47–61, 2009.

[14] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. A. Ramaiah, "Multi-level basis selection of wavelet packet decomposition tree for heart sound classification," *Computers in Biology and Medicine*, vol. 43, no. 10, pp. 1407–1414, 2013.

[15] S. Sun, "An innovative intelligent system based on automatic diagnostic feature extraction for diagnosing heart diseases," *Knowledge-Based System*, vol. 75, pp. 224–238, 2015.

[16] J.-m. Kwon, K.-H. Kim, Z. Akkus, K.-H. Jeon, J. Park, and B.-H. Oh, "Artificial intelligence for detecting mitral regurgitation using electrocardiography," *Journal of Electrocardiology*, vol. 59, pp. 151–157, 2020.

[17] H. Moghaddasi and S. Nourian, "Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos," *Computers in biology and medicine*, vol. 73, pp. 47–55, 2016.

[18] S. Uretsky, E. Argulian, J. Narula, and S. D. Wolff, "Use of cardiac magnetic resonance imaging in assessing mitral regurgitation: current evidence," *Journal of the American College of Cardiology*, vol. 71, no. 5, pp. 547–563, 2018.

[19] S. Militaru, O. Bonnefous, K. Hami et al., "Validation of semi-automated quantification of mitral valve regurgitation by three-dimensional color Doppler transesophageal echocardiography," *Journal of the American Society of Echocardiography*, vol. 33, no. 3, pp. 342–354, 2020.

[20] P. G. Chew, K. Bounford, S. Plein, D. Schlosshan, and J. P. Greenwood, "Multimodality imaging for the quantitative assessment of mitral regurgitation," *Quantitative Imaging in Medicine and Surgery*, vol. 8, no. 3, pp. 342–359, 2018.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 1562502, 8 pages
https://doi.org/10.1155/2021/1562502

Research Article

Clinical Significance of Hyperdense Lesions on Non-enhanced Brain CT Obtained Immediately after Arterial Revascularization in Acute Ischemic Stroke Patients

Changbin Wang,[1] Zudong Yin,[1] Xinyi Zhang,[1] and Xiumin Zhao[2]

[1]*Department of Radiology, Shandong Provincial Third Hospital, Cheeloo College of Medicine, Shandong University, Jinan 250031, China*
[2]*Department of Neurology, Shandong Provincial Third Hospital, Cheeloo College of Medicine, Shandong University, No. 12, Wuyingshan Middle Road, Tianqiao District, Jinan 250031, China*

Correspondence should be addressed to Xiumin Zhao; wusuowei6960@sina.com

Received 25 May 2021; Accepted 17 August 2021; Published 6 September 2021

Academic Editor: Tao Song

Purpose. To analyze the characteristics of hyperdense lesions on brain CT conducted immediately after arterial revascularization (AR) in patients with acute ischemic stroke (AIS), track the outcome of those lesions and investigate their clinical significance. *Materials and Methods*. 97 AIS patients were enrolled in our study. Among them, 52 patients showed hyperdense lesions and were divided into three categories: type I, type II and type III according to the morphologic characteristics of hyperdense lesions. All patients underwent several follow-up CT/MR examinations to visualize the outcomes of the lesions. *Results*. Among the 52 patients, 22 showed contrast extravasation, 23 displayed contrast extravasation combined with hemorrhagic transformation (HT) and 7 confirmed symptomatic intracranial hemorrhage (SICH) in follow-up CT/MR. Among the without hyperdense lesions group, only 7 converted to hemorrhage, and no SICH occurred. All type I lesions showed contrast extravasation; 23 type II lesions turned to hemorrhage, 2 revealed SICH and 6 were pure contrast extravasation; all of the type III developed into SICH. *Conclusion*. Hyperdense lesions on non-enhanced brain CT obtained immediately after arterial revascularization (AR) exhibited varying features. Type I indicated a pure contrast extravasation. Type II and type III hyperdense lesions suggested higher incidence of HT, the presence of type III lesions indicated an ominous outcome.

1. Introduction

Arterial revascularization (AR) has become the preferred treatment for patients with acute ischemic stroke (AIS) [1–3]. Non-enhanced brain CT conducted immediately after AR often presents hyperdense lesions in the parenchyma [4, 5]. Such hyperdense lesions have been reported and analyzed in various literature since 1993 [6]. However, it is still a tricky problem to determine whether the hyperdense lesions are contrast extravasation or cerebral hemorrhage [7]. The purpose of this study was to analyze the imaging characteristics of intracranial hyperdense lesions and to classify them according to their morphological characteristics. And the final outcome and clinical significance of different types of hyperdense lesions was analyzed based on its spontaneous regression, HT, or even SICH.

2. Materials and Methods

2.1. Patients. From August 2016 to Match 2020, 97 AIS patients (51 males and 46 females) were included in our study with an average age of 61.73 ± 8.44 years (range, 41-83 years). This study was approved by the Medical Ethics Committee of our hospital. Written informed consent was obtained from all patients.

2.2. Arterial Thrombolysis or/and Mechanical Thrombectomy. All patients received arterial thrombolysis or/and mechanical thrombectomy within 6 hours after onset. The findings of brain CT immediately after AR were analyzed in all cases. The contrast agent used for cerebral angiography during AR was iohexol (300mgI/ml), with the dosage of 150-250 ml. The time window for arterial thrombolysis (from the initial onset to the onset of arterial thrombolysis) is 3-6 hours. Thrombolysis lasts for 30 minutes to 2.5 hours. The thrombolytic agent used was urokinase. Doses of urokinase ranged from 60000 to 600000 U, with 10 mL saline/60000 U, in boluses. 40 of the patients were treated with thrombolysis alone and 57 were treated with mechanical thrombectomy combined with arterial thrombolysis. Among the 97 patients, the site of occlusion was the internal carotid artery in 32 patients, the M1 segment of middle cerebral artery in 54 patients, the M2 segment in 8 patients, and the basilar artery in 3 patients.

2.3. CT Acquisition. After cerebral angiography and revascularization of obstructed vessels, all patients underwent non-enhanced CT scan of brain immediately after AR. Follow-up CT was performed within 24 hours or 48 hours after the procedure. Gd-free MR scanning was performed in 21 patients within 48 hours after procedure. Using Philips Brilliance 256 rows of iCT scanners, all patients were scanned from the base of the skull to the top of the skull. Scanning conditions: 120KV, 300MA, layer thickness of 3 mm. Philips Achieva 1.5 T MR scanner was used to collect the images with the skull phased-front circle. The scanning sequence includes T1-weighted sequence, T2- weighted sequence, fluid attenuated inversion recovery (Flair), diffusion weighted imaging (DWI), apparent diffusion coefficient (ADC), susceptibility weighted imaging (SWI) etc.

2.4. Imaging Evaluation. The characteristics of non-enhanced brain CT of all patients were analyzed and evaluated by two experienced attending radiologists or above. According to the shape, location and with/without mass effect of hyperdense lesions, the patients were divided into three categories: type I (sporadic patchy lesions with unclear boundaries which mostly distributed in the cerebral cortex and sulti, without mass effect (Figure 1(a))), type II (solid "mass" shaped lesions with well-defined boundaries which mostly distributed in the basal ganglion and without mass effect or mild mass effect (Figures 2(a) and 3(a))) and type III (diffuse patchy lesions occupying a larger area with obvious mass effect (Figure 4(a))). Once divergence occurred during the diagnosis and classification evaluation, the two parties shall solve the difference through consultation.

2.5. Statistical Analysis. All statistical analyses were conducted using SPSS version 18 (SPSS Inc, Chicago, USA). Quantitative data were expressed as means ± standard (SD) and compared by the two-sample independent Student's t-test; Qualitative data was expressed as rate and compared by Chi-square test. $P < 0.05$ indicated that the difference was statistically significant.

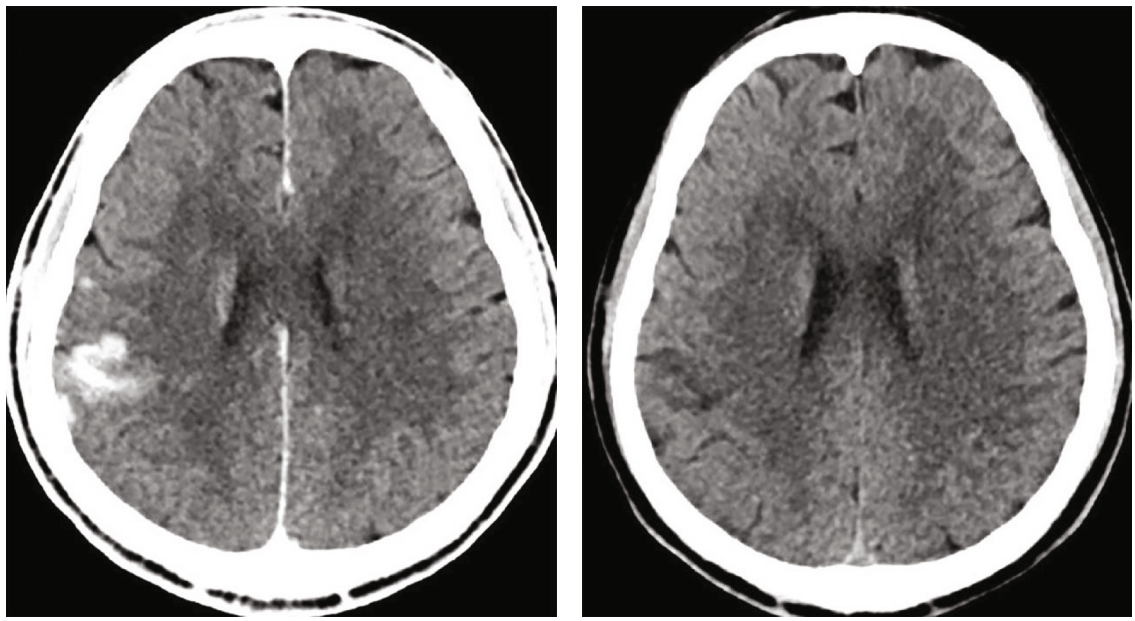
3. Results

3.1. Overall Results. Among 97 patients, 52 (53.6%) patients showed hyperdense lesions around the intracranial infarction area on non-enhanced CT conducted immediately after AR, whereas the other 45 (46.4%) patients (the control group) had no hyperdense lesions. The basic clinical data of patients in the hyperdense group and without hyperdense group were summarized in Table 1. There were no significant differences in age, sex, hypertension, diabetes, NIHSS score, history of antiplatelet therapy before onset, degree of recanalization, methods of arterial revascularization. 30 (57.7%) patients in the intracranial hyperdense lesions group developed HT after AR, Among the control group, 7 patients converted to AICH (Asymptomatic Intracranial Hemorrhage) in follow-up CT/MR within 24 or 48 hours after AR (Table 2), and no SICH occurred. Significantly higher HT incidence was observed in the hyperdense lesions group when compared with the without hyperdense lesions group ($P < 0.001$).

3.2. Outcome of the Hyperdense Lesions. In the hyperdense group, hyperdense lesions were completely disappeared in 22 (42.3%) patients (Figures 1 and 3), 23 (44.2%) patients displayed contrast extravasation combined with HT (Figure 2) and 7 (13.5%) patients were confirmed SICH (Figure 4) in follow-up CT/MR within 24 or 48 hours after AR. In follow-up CT 24 or 48 hours after AR, the outcome of different types of hyperdense lesions were as follows (Table 3): sporadic patchy type I lesions (16 lesions) were completely disappeared, showing contrast extravasation; 23 type II lesions (74.2%) turned to hemorrhage (AICH), 2 lesions (6.5%) revealed SICH and 6 lesions (19.4%) were contrast extravasation; all of the type III (5 cases) lesions developed into SICH. Type II and type III had a significantly higher incidence of HT than type I ($P < 0.01$). The sensitivity, specificity, positive predictive value and negative predictive value of hyperdense lesions on non-enhanced CT conducted immediately after AR for the prediction of HT were 81.1%, 63.3%, 57.7% and 84.4%, respectively.

4. Discussion

With the development of medical imaging equipment and interventional medical technology, patients with AIS have more opportunities for transductal AR (including intra-arterial drug thrombolysis, mechanical thrombectomy), which improves the rate of vascular recanalization in stroke patients and greatly improves the patients' clinical prognosis and survival rate. However, intracranial hemorrhage, a complication associated with intravascular interventional procedures, poses a challenge to clinical diagnosis and treatment of AIS. Immediately after completing intra-arterial procedures, a non-enhanced CT is often performed to assess the progression of the AIS and whether hemorrhage has occurred [6]. Non-enhanced CT often detects hyperdense lesions in the parenchyma. Not all hyperdense lesions on non-enhanced CT conducted immediately after AR represent hemorrhage, it may be attributed to either contrast



(a) (b)

Figure 1: Non-enhanced brain CT scan of a 58-year-old woman who had a sudden onset of impaired left limb movement. (a). The non-enhanced brain CT scan obtained immediately after intra-arterial thrombolysis showed that a type I hyperdense lesion was located in the right temporal lobe cortex with a CT value of 90-135 Hu. There was no mass effect. (b). Follow-up CT scan obtained 24 hours after the end of AR, the hyperdense lesion on the right temporal lobe cortex was completely disappeared.

extravasation or cerebral hemorrhage [6, 8]. So neurologists and neuroradiologists need to identify the true nature of the hyperdense lesions so that appropriate interventions can be taken in time, otherwise it will seriously affect the prognosis of patients.

A large amount of contrast agent is injected during the angiography and AR in patients with AIS, so extravascular exudation of contrast agent may be the pathological mechanism of CT intracranial hyperdense lesions. Contrast extravascular exudation is based on damage to the blood-brain barrier (BBB) [9]. For patients with AIS, the permeability of occluded vessels in the infarcted area was changed. After recanalization, the blood flow in the infracted area increased significantly and hyperperfusion occurred. When the perfusion pressure of the distal capillary bed exceeds the bearing capacity of the vessel wall, the BBB is destroyed and contrast agent exudates. The removal time of the contrast agent injected during the recanalization is prolonged due to ischemia and hypoxia in the cerebral infarction area, which results in the local retention of contrast agent [10]. During interventional procedure, microcatheter and microguide wire may lead to vascular intimal damage, which is also one of the common causes of contrast extravasation [11, 12]. The manifestations of post-procedure CT intracranial hyperdense lesions vary with the degree of cerebral microvascular injury. When ischemic injury only destroys the permeability barrier of endothelial cells, the intracranial hyperdense lesions may be a single contrast agent without hemorrhage. However, when ischemic infarcts break down the structural barrier-basement membrane, the hyperdense lesions may be associated with some form of hemorrhage, or a mixture of contrast and blood. Disruption of the blood–brain barrier may be an essential condition for hemorrhagic transformation [11–13].

There were no significant differences in terms of age, sex, hypertension, diabetes mellitus, National Institute of Health Stroke Scale score (NIHSSs), history of antiplatelet therapy before onset, recanalization rate and method of thrombolysis between patients with intracranial hyperdense lesions group and those without. The incidence of HT in patients with hyperdense lesions group was significantly higher than that without hyperdense lesions group, and the difference was statistically significant (Table 1).

In this study, 22 (42.3%) patients of the intracranial hyperdense lesions were completely disappeared on a follow-up CT obtained within 24 hours or 48 hours after procedure. Among them, 16 patients had type I hyperdense lesions and 6 patients had type II lesions. Follow-up CT scan showed no recurrence of HT in these patients. This rapidly fading intracranial hyperdense lesions was contrast extravasation without HT. That is, ischemic injury in these patients may be limited only to the endothelial permeability barrier. Wildenhain et al. reported that rapid dissolution of intracranial hyperdense lesions is a good prognostic indicator [2].

There are several definitions of contrast extravasation after AR in patients with AIS. Nakano et al. [14] believed that the rapidly fading of high-density shadow was the main basis for the diagnosis of the extravasation of contrast agent. Mericle et al. [7] showed that extravasation of contrast medium should be defined as a hyperdense lesions with a maximal Hu measurement>90 and/or the hyperdense lesions rapidly subsided within 24 hours after AR. According to our study, we are more inclined to suggest that rapid resolution of intracranial hyperdense lesions is the key to the

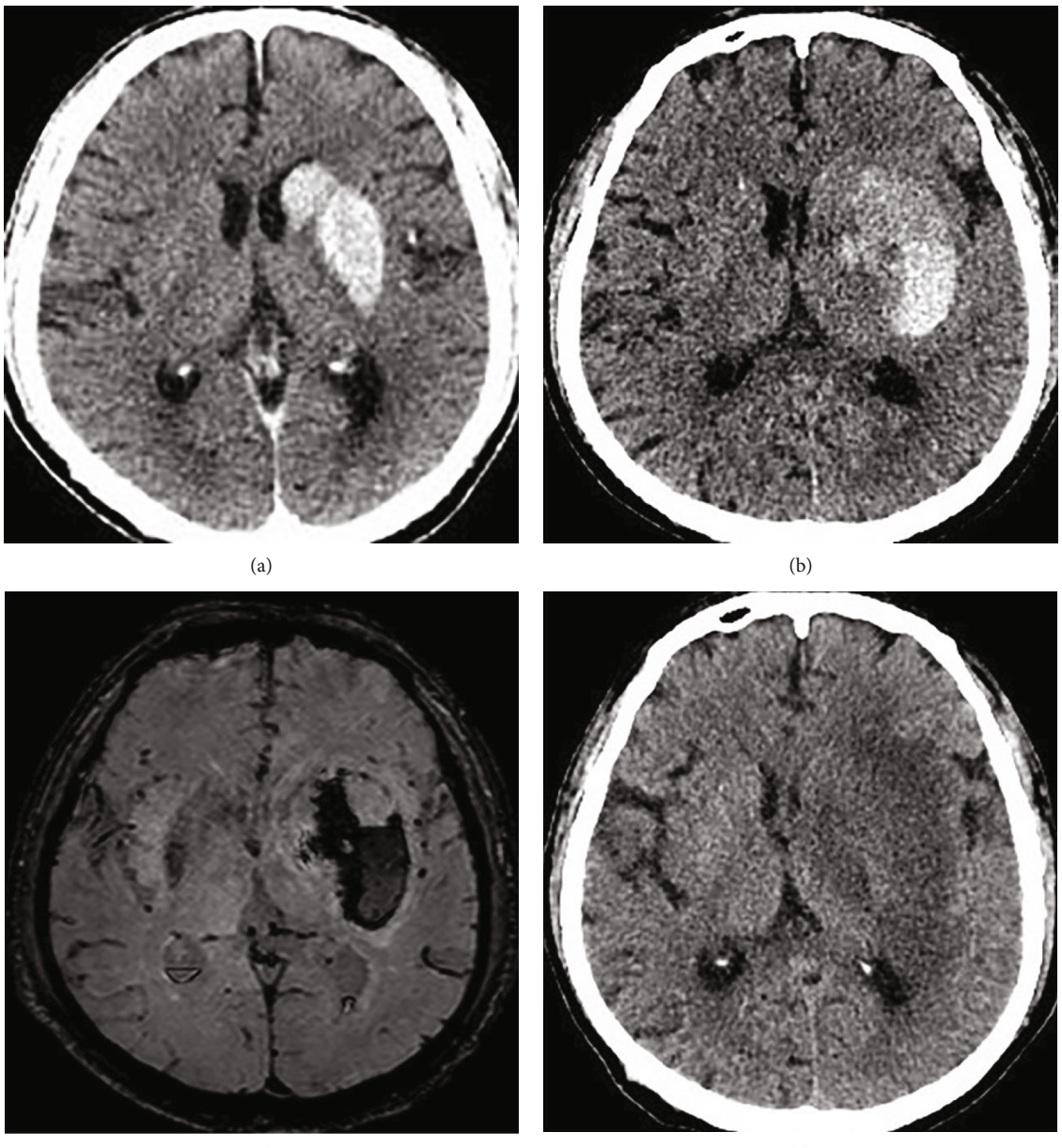
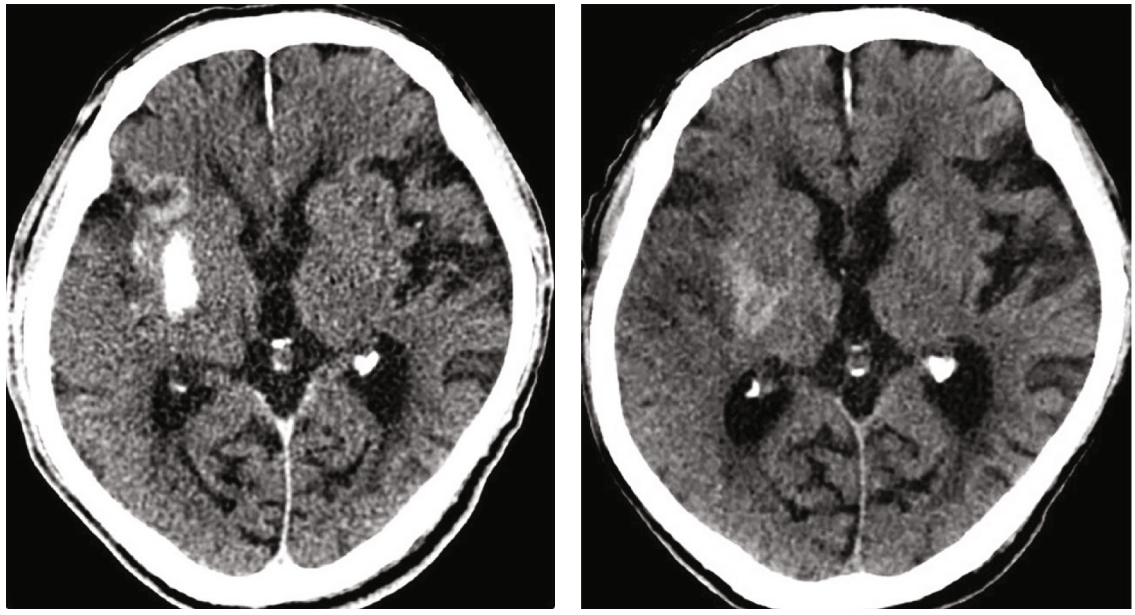


Figure 2: CT scan of a 62-year-old man who had a sudden onset of impaired right limb movement. (a). The non-enhanced brain CT scan obtained immediately after intra-arterial thrombolysis showing a hyperdense area in the head portion of the left caudate nucleus and the entire left lentiform nucleus. Representative type II, namely solid "mass" shaped lesions with well-defined boundaries and without mass effect. (b). Follow-up CT scan obtained 48 hours after the end of intra-arterial thrombolysis. The range of the hyperdense area increased slightly and the attenuation decreased. The patient was not accompanied by deterioration of neurological symptoms. (c). On the 2nd day after the procedure, susceptibility weighted imaging showed patchy low signal in the left basal ganglia region, indicating hematoma formation. (d). Two weeks after the procedure, the left basal ganglia hematoma had been completely absorbed.

diagnosis of contrast extravasation, and there was no mass effect of the intracranial hyperdense lesions. The maximal Hu measurement>90 of intracranial hyperdense lesions was not an absolute indication for the diagnosis of the extravasation of contrast medium.

If follow-up CT within 24 hours or 48 hours after procedure showed that the attenuation of the intracranial hyperdense lesions was mildly reduced compared with that of non-enhanced CT conducted immediately after completing inter-arterial procedure, and the size and contour of intracranial hyperdense lesions changed slightly, even the hyperdense lesions appeared space occupying effect. The above CT findings are known as contrast extravasation combined with HT. HT is the most serious complication after AR in AIS patients [12]. HT after arterial thrombolysis can be divided into SICH and AICH. SICH was defined as a decrease in the NIHSSs ≥4 within 36 hours after thrombolysis and intracranial hemorrhage confirmed by imaging was correlated



(a) (b)

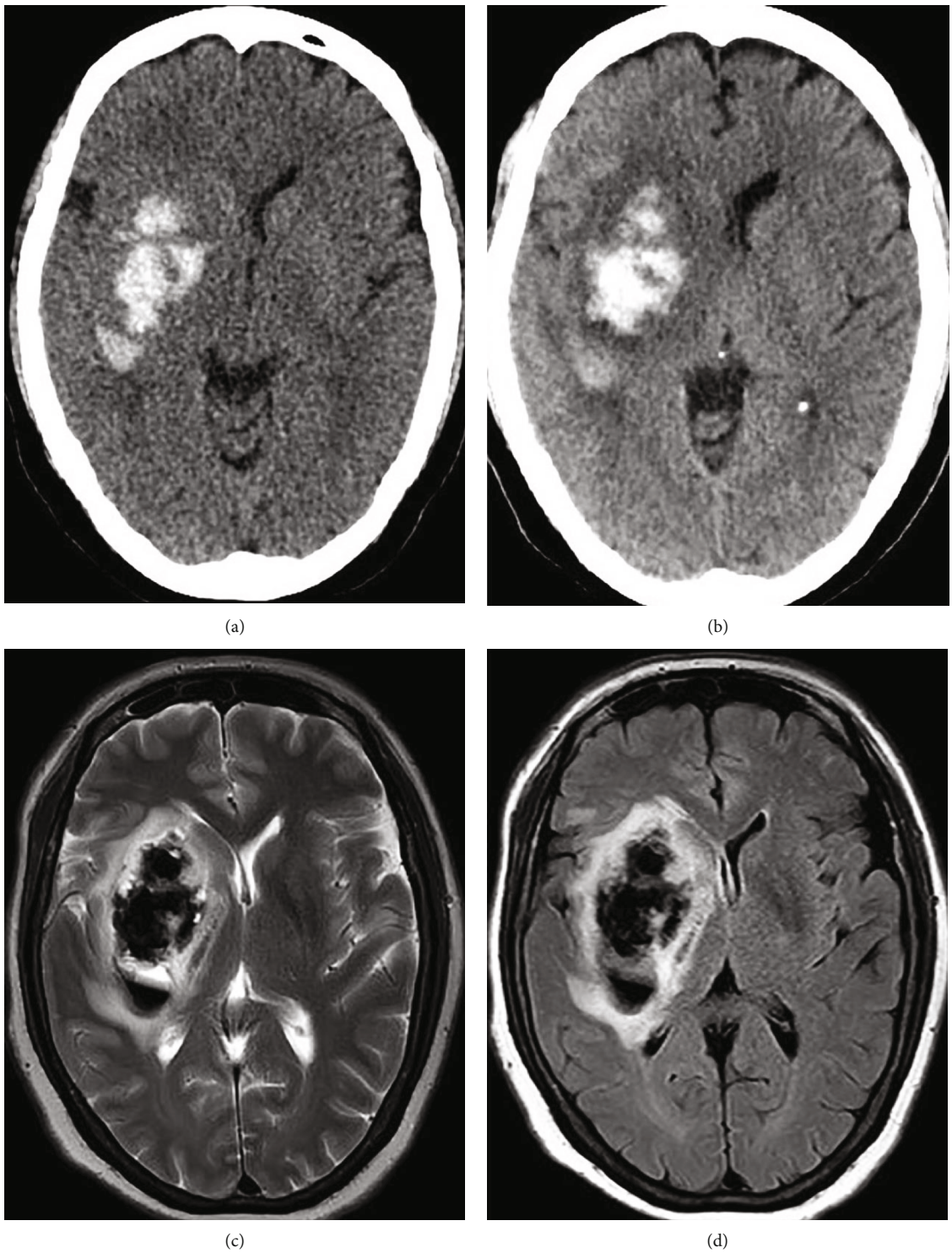
Figure 3: CT scan of a 58-year-old woman who had a sudden onset of impaired left limb movement. (a). The non-enhanced brain CT scan obtained immediately after intra-arterial thrombolysis showing a type II hyperdense lesion in the right lentiform nucleus with a CT value of 90-135 Hu. A type I hyperdense lesion was located in the right insular cortex with a CT value of 52-57 Hu (the latter was not included in this study). There was no mass effect in the above two lesions. (b). Follow-up CT scan obtained 24 hours after the end of AR, the hyperdense lesion of the right lentiform nucleus was obviously absorbed. The patient was not accompanied by deterioration of neurological symptoms.

with the deterioration of clinical symptoms in time. AICH was defined as bleeding without deterioration of neurological symptoms [15].

HT is a complex and multifactorial phenomenon. Known risk factors include age, blood glucose, low platelet count, high NIHSSs, size and location of ischemic area, poor collateral vessels, arterial stiffness, thrombolytic agent used and time window allowed for the initiation of the therapy [16, 17]. And its occurrence is closely related to the damage of BBB, the reperfusion injury of ischemic area, the use of microcatheter, the establishment of collateral circulation and prolonged procedure time [18, 19], which is the result of the combined action of the above factors [4]. In this study, 30 (57.7%) patients in the intracranial hyperdense lesions group developed HT after AR, including 23 (44.2%) patients with AICH and 7(13.5%) patients with SICH. All of the 23 (44.2%) patients with AICH were type II hyperdense lesions. Among the 7 patients with SICH, 2 were type II hyperdense lesions and 5 were type III.

SICH, a massive intracerebral hematoma (accompanied by obvious space occupying effect) with worsening neurological symptoms, is a clinical and CT imaging hybrid [20]. It is also the most intractable problem in the reperfusion treatment of AIS patients. The rapid disappearance of intracranial hyperdense lesions after AR had nothing to do with SICH. Since SICH is a serious microvascular injury involving the basement membrane, no hyperdense lesions on non-enhanced CT immediately after AR suggests that the permeability and structural barrier of microvascular endothelial cells are not destroyed, which may be a reliable negative predictor of SICH, in other words, the incidence of SICH will be extremely low or will not occur at all. Kass-Hout et al. [21] indicated that a longer procedure time is an independent predictor for SICH in patients receiving mechanical thrombectomy. The extended procedure time may be due to difficulty getting catheters into blocked vessels, which may be related to arterial stiffness or more comorbidities [22]. Prolonged procedure time means more attempts to recanalize the occluded vessels. The above factors could greatly increase the incidence of HT [23, 24]. A prolonged procedure time was significantly associated with a higher rate of SICH, even in patients with successful recanalization [19].

Nakano et al. [14] reported that 48% of patients showed intracranial hyperdense lesions on non-enhanced CT immediately after arterial thrombolysis, and 29.7% of these patients developed SICH, while patients without intracranial hyperdense lesions on non-enhanced CT immediately after AR had no SICH. Jang et al. [6] reported that 33% (31/94) of patients displayed intracranial hyperdense lesions on non-enhanced CT immediately after intraarterial procedures. Among them, 58% (18/31) developed cerebral hemorrhage, and 19.4% (6/31) developed SICH. In this study, 53.6% (52/97) of the patients showed intracranial hyperdense lesions around the intracranial infarction area on non-enhanced CT immediately after AR, and 13.5% (7/52) of them developed SICH. None of the patients without intracranial hyperdense lesions developed SICH. In our study, the incidence of intracranial hyperdense lesions on non-enhanced CT immediately after AR was higher than that in the results of Nakano et al. [14] and Jang et al. [6], and the incidence of SICH was lower than that in the results of



(a) (b)

(c) (d)

Figure 4: CT scan of a 66-year-old woman who experienced a sudden left limb inactivity accompanied by loss of speech function for 4 hours. There was no intracranial hemorrhage on CT scan before thrombolytic therapy. (a). The non-enhanced brain CT scan obtained immediately after intra-arterial thrombolysis showing a type III hyperdense lesion in the right basal ganglia with a CT value of 55-73 Hu. The right ventricle was significantly compressed. The patient had a marked symptom of neurological deterioration. (b). Follow-up CT scan obtained 24 hours after the end of AR revealed that the contour of the hyperdense lesion located in the right basal ganglia region was slightly expanded, the attenuation was increased, and the surrounding low-density edema occurred, with obvious space-occupying effect. (c). T2WI and D FLAIR obtained 24 hours after procedure showed large patchy low signals in the right basal ganglia region surrounded by high signal. Low signal denotes hematoma and high signal denotes edema.

Table 1: Baseline Characteristics of Patients in the Hyperdense Group and without hyperdense Group.

	Hyperdense group (n = 52)	Without hyperdense group (n = 45)	P
Age (mean ± SD)	61.5 ± 8.66	62.0 ± 8.26	0.773
Gender (male/female)	28/24	23/22	0.788
Hypertension	24	20	0.866
Diabetes	18	14	0.714
NIHSS score (mean ± SD)	13.94 ± 3.9	13.93 ± 4.2	0.991
History of antiplatelet therapy before onset	29	25	0.983
Successful recanalization (%)	34 (65.4)	26 (57.8)	0.442
Thrombolysis	22	18	0.818
Thrombolysis and mechanical thrombectomy	30	27	0.818
HT (%)	30 (57.7)	7 (15.6)	<0.001

NIHSS=National Institute of Health Stroke Scale. HT = hemorrhagic transformation.

Table 2: Distribution of HT between the hyperdense group and the without hyperdense group.

	No hemorrhage	AICH	SICH
Hyperdense group (n = 52)	22	23	7
Without hyperdense group (n = 45)	38	7	0
P		<0.001	

HT = hemorrhagic transformation, AICH = Asymptomatic Intracranial Hemorrhage, SICH = Symptomatic Intracranial Hemorrhage.

Table 3: Distribution of HT among the different types of hyperdense lesions.

	No hemorrhage	AICH	SICH
Type I (n = 16)	16	0	0
Type II (n = 31)	6	23	2
Type III (n = 5)	0	0	5
P		<0.01	

HT = hemorrhagic transformation, AICH = Asymptomatic Intracranial Hemorrhage, SICH = Symptomatic Intracranial Hemorrhage.

Nakano et al. [14] and Jang et al. [6]. There was no significant difference between cases selection and vascular reconstruction methods, so this difference may be related to the insufficient sample size, different generations of thrombectomy devices and procedure time. Whether it is related to other factors remains further analysis with a larger group of patients. Our results showed that none of the patients without intracranial hyperdense lesions developed SICH, which was consistent with the study of Nakano et al. [14].

This study showed that type I intracranial hyperdense lesions (sporadic patchy lesions with unclear boundaries), which completely disappeared within 24 hours post-procedure, is a single, pure contrast extravasation. A large proportion (80.6%,25/31) of type II intracranial hyperdense lesions (solid "mass" shaped lesions with well-defined boundaries) converted to hemorrhage, while a small proportion (19.4%,6/31) were only contrast extravasation. Type III intracranial hyperdense lesions (diffuse patchy lesions with mass effect) invariably developed into SICH. The occurrence of intracranial hyperdense lesions after AR has a high sensitivity and negative predictive value for the prediction of cerebral HT, but the specificity and positive predictive value are not high, which is consistent with the research results of Kim et al. [25].

This study has some limitations. Firstly, the retrospective design has its inherent limitations of such study. Secondly, the sample was small, which may have influenced the results in this study. Another limitation is that our study was a single-center study and we did not analyze the effect of the duration of procedure on the occurrence of hyperdense lesions on non-enhanced CT conducted immediately after AR. Therefore, we will conduct multicenter studies in future to collect more cases of ischemic stroke with transarterial revascularization and follow up for a longer time, so as to confirm the availability and practicability of the morphological classification of intracranial hyperdense lesions in the current study.

5. Conclusions

In conclusion, the occurrence of intracranial hyperdense lesions on non-enhanced CT conducted immediately after AR has its limitations for the accurate prediction of cerebral HT (including SICH and AICH). Type I hyperdense lesions suggested pure contrast extravasation and was not associated with hemorrhagic complications. The presence of Type II and type III hyperdense lesions indicated higher incidence of HT, in particular, the presence of type III hyperdense lesions indicated an ominous outcome. When intracranial hyperdense lesions of type II and type III appear around the intracranial infarction area in stroke patients, high attention should be paid to the progress of the patient's neurological symptoms and timely and effective treatment should be given to minimize the complications of cerebral hemorrhage.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

References

[1] W. S. Smith, "Safety of mechanical Thrombectomy and intravenous tissue plasminogen activator in acute ischemic stroke. Results of the multi mechanical Embolus removal in cerebral ischemia (MERCI) trial, part I," *AJNR. American Journal of Neuroradiology*, vol. 27, no. 6, pp. 1177–1182, 2006.

[2] V. Yedavalli and S. Sammet, "Contrast extravasation versus hemorrhage after Thrombectomy in patients with acute stroke," *Journal of Neuroimaging*, vol. 27, no. 6, pp. 570–576, 2017.

[3] E. V. Kuklina, X. Tong, M. G. George, and P. Bansil, "Epidemiology and prevention of stroke: a worldwide perspective," *Expert Review of Neurotherapeutics*, vol. 12, no. 2, pp. 199–208, 2012.

[4] O. Nikoubashman, A. Reich, M. Gindullis et al., "Clinical significance of post-interventional cerebral hyperdensities after endovascular mechanical thrombectomy in acute ischaemic stroke," *Neuroradiology*, vol. 56, no. 1, pp. 41–50, 2014.

[5] G. Parrilla and B. García-Villalba, "hemorrhage/contrast staining areas after mechanical intra-arterial thrombectomy in acute ischemic stroke: imaging findings and clinical significance," *AJNR. American Journal of Neuroradiology*, vol. 33, no. 9, pp. 1791–1796, 2012.

[6] Y. M. Jang, D. H. Lee, H. S. Kim et al., "The fate of high-density lesions on the non-contrast CT obtained immediately after intra-arterial thrombolysis in ischemic stroke patients," *Korean Journal of Radiology*, vol. 7, no. 4, pp. 221–228, 2006.

[7] R. A. Mericle, D. K. Lopes, M. D. Fronckowiak, A. K. Wakhloo, L. R. Guterman, and L. N. Hopkins, "A grading scale to predict outcomes after intra-arterial thrombolysis for stroke complicated by contrast extravasation," *Neurosurgery*, vol. 46, no. 6, pp. 1307–1315, 2000, discussion 1314-5.

[8] R. G. Nogueira, D. S. Liebeskind, G. Sung, and G. Duckwiler, "Predictors of good clinical outcomes, mortality, and successful revascularization in patients with acute ischemic stroke undergoing thrombectomy: pooled analysis of the mechanical Embolus removal in cerebral ischemia (MERCI) and multi MERCI trials," *Stroke*, vol. 40, no. 12, pp. 3777–3783, 2009.

[9] S. Payabvash, M. H. Qureshi, S. M. Khan et al., "Differentiating intraparenchymal hemorrhage from contrast extravasation on post-procedural noncontrast CT scan in acute ischemic stroke patients undergoing endovascular treatment," *Neuroradiology*, vol. 56, no. 9, pp. 737–744, 2014.

[10] W. Yoon, J. J. Seo, J. K. Kim, K. H. Cho, J. G. Park, and H. K. Kang, "Contrast enhancement and contrast extravasation on computed tomography after intra-arterial thrombolysis in patients with acute ischemic stroke," *Stroke*, vol. 35, no. 4, pp. 876–881, 2004.

[11] R. C. Lisboa, B. D. Jovanovic, and M. J. Alberts, "Analysis of the safety and efficacy of intra-arterial thrombolytic therapy in ischemic stroke," *Stroke*, vol. 33, no. 12, pp. 2866–2871, 2002.

[12] J. Kaesmacher, M. Kaesmacher, C. Maegerlein et al., "Hemorrhagic transformations after thrombectomy: risk factors and clinical relevance," *Cerebrovascular Diseases*, vol. 43, no. 5-6, pp. 294–304, 2017.

[13] H. S. Chong, S. C. Jung, S. J. Cho et al., "MRI for prediction of hemorrhagic transformation in acute ischemic stroke: a systematic review and meta-analysis," *Acta Radiologica*, vol. 61, no. 7, 2019.

[14] S. Nakano, T. Iseda, H. Kawano, T. Yoneyama, T. Ikeda, and S. Wakisaka, "Parenchymal hyperdensity on computed tomography after intra-arterial reperfusion therapy for acute middle cerebral artery occlusion: incidence and clinical significance," *Stroke*, vol. 32, no. 9, pp. 2042–2048, 2001.

[15] C. Berger, M. Fiorelli, T. Steiner et al., "Hemorrhagic transformation of ischemic brain tissue: asymptomatic or symptomatic?," *Stroke*, vol. 32, no. 6, pp. 1330–1335, 2001.

[16] M. Acampa, S. Camarri, P. E. Lazzerini et al., "Increased arterial stiffness is an independent risk factor for hemorrhagic transformation in ischemic stroke undergoing thrombolysis," *International Journal of Cardiology*, vol. 243, pp. 466–470, 2017.

[17] K. R. Van Kranendonk, K. M. Treurniet, A. M. Boers et al., "Hemorrhagic transformation is associated with poor functional outcome in patients with acute ischemic stroke due to a large vessel occlusion," *Journal of Neurointerventional Surgery*, vol. 11, 2018.

[18] A. M. Spiotta, J. Vargas, R. Turner, M. I. Chaudry, H. Battenhouse, and A. S. Turk, "The golden hour of stroke intervention: effect of thrombectomy procedural time in acute ischemic stroke on outcome," *Journal of NeuroInterventional Surgery*, vol. 6, no. 7, pp. 511–516, 2014.

[19] X. Huang, Q. Cai, L. Xiao et al., "Influence of procedure time on outcome and hemorrhagic transformation in stroke patients undergoing thrombectomy," *Journal of Neurology*, vol. 266, no. 10, pp. 2560–2570, 2019.

[20] N. Lummel, G. Schulte-Altedorneburg, C. Bernau et al., "Hyperattenuated intracerebral lesions after mechanical recanalization in acute stroke," *AJNR. American Journal of Neuroradiology*, vol. 35, no. 2, pp. 345–351, 2014.

[21] T. Kass-Hout, O. Kass-Hout, C. J. Sun, T. A. Kass-Hout, R. Nogueira, and R. Gupta, "Longer procedural times are independently associated with symptomatic intracranial hemorrhage in patients with large vessel occlusion stroke undergoing thrombectomy," *Journal of NeuroInterventional Surgery*, vol. 8, no. 12, pp. 1217–1220, 2016.

[22] M. Ribo, A. Flores, M. Rubiera et al., "Difficult catheter access to the occluded vessel during endovascular treatment of acute ischemic stroke is associated with worse clinical outcome," *Journal of NeuroInterventional Surgery*, vol. 5, Supplement 1, pp. i70–i73, 2013.

[23] Y. Hao, D. Yang, H. Wang et al., "Predictors for symptomatic intracranial hemorrhage after endovascular treatment of acute ischemic stroke," *Stroke*, vol. 48, no. 5, pp. 1203–1209, 2017.

[24] R. Bourcier, S. Saleme, J. Labreuche et al., "More than three passes of stent retriever is an independent predictor of parenchymal hematoma in acute ischemic stroke," *Journal of NeuroInterventional Surgery*, vol. 11, no. 7, pp. 625–629, 2019.

[25] J. T. Kim, S. H. Heo, B. H. Cho et al., "Hyperdensity on noncontrast CT immediately after intra-arterial revascularization," *Journal of Neurology*, vol. 259, no. 5, pp. 936–943, 2012.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 1102083, 10 pages
https://doi.org/10.1155/2021/1102083

Research Article

Classification Framework for Healthy Hairs and Alopecia Areata: A Machine Learning (ML) Approach

Choudhary Sobhan Shakeel,[1] **Saad Jawaid Khan**,[1] **Beenish Chaudhry**,[2] **Syeda Fatima Aijaz**,[1] **and Umer Hassan**[1]

[1]*Department of Biomedical Engineering, Ziauddin University, Faculty of Engineering, Science, Technology and Management, Karachi, Pakistan*
[2]*School of Computing and Informatics, University of Louisiana at Lafayette, USA*

Correspondence should be addressed to Saad Jawaid Khan; sj.khan@zu.edu.pk

Received 13 June 2021; Revised 18 July 2021; Accepted 29 July 2021; Published 16 August 2021

Academic Editor: Iman Yi Liao

Alopecia areata is defined as an autoimmune disorder that results in hair loss. The latest worldwide statistics have exhibited that alopecia areata has a prevalence of 1 in 1000 and has an incidence of 2%. Machine learning techniques have demonstrated potential in different areas of dermatology and may play a significant role in classifying alopecia areata for better prediction and diagnosis. We propose a framework pertaining to the classification of healthy hairs and alopecia areata. We used 200 images of healthy hairs from the Figaro1k dataset and 68 hair images of alopecia areata from the Dermnet dataset to undergo image preprocessing including enhancement and segmentation. This was followed by feature extraction including texture, shape, and color. Two classification techniques, i.e., support vector machine (SVM) and k-nearest neighbor (KNN), are then applied to train a machine learning model with 70% of the images. The remaining image set was used for the testing phase. With a 10-fold cross-validation, the reported accuracies of SVM and KNN are 91.4% and 88.9%, respectively. Paired sample T-test showed significant differences between the two accuracies with a $p < 0.001$. SVM generated higher accuracy (91.4%) as compared to KNN (88.9%). The findings of our study demonstrate potential for better prediction in the field of dermatology.

1. Introduction

The "falling of scalp hairs in sufficient quantity" is defined as hair loss [1]. Alopecia areata is an autoimmune disorder that involves nonscarring hair loss in well-defined patches that can affect the entire scalp region and, ultimately, lead to baldness [2, 3]. The disorder impacts millions of people worldwide [4], especially those with a family history of alopecia areata [5]. It begins when the body's autoimmune system starts to target the hair follicles, disturbing their normal functioning and preventing subsequent hair growth. The outcome is hair loss. Hair loss can be attributed to various causes, and trichoscopies and biopsies are generally necessary to ensure the cause is alopecia areata. However, the limitations of these diagnostic methods are the uncertainty surrounding the number of tests required for adequate diagnosis. Hence, there is a vast scope for researching new techniques pertaining to the classification and diagnosis of alopecia areata [6].

Machine learning (ML) techniques have shown effectiveness in the prediction and classification of various diseases and disorders. Machine learning encapsulates the study of different computer algorithms that exhibit the potential to learn and adapt [7]. Machine learning (ML) algorithms and their advanced versions have been incorporated in various medical disciplines for diagnostic purposes. For instance, machine learning (ML) techniques have exhibited accurate results using magnetic resonance imaging (MRI) and

computed tomography (CT) images for the diagnosis of brain tumors, breast cancer, ovarian cancer, pulmonary disease, and dermatological diseases [8–12]. Machine learning (ML) techniques have also shown their credibility during the COVID-19 pandemic and have aided medical professionals in identifying the coronavirus disease along with its levels [13].

In dermatology, effective diagnosis and prediction have been achieved by different machine learning methods. Scalp analysis systems have been developed utilizing SVM and KNN to classify scalp images. Scalp images have been used for classification of conditions such as dandruff with the employment of machine learning techniques of SVM, KNN, and decision trees [14–20]. All these techniques use scalp and/or skin images to develop prediction models. To the best of our knowledge, so far, none of the machine learning techniques has been applied on human hair images.

In this paper, we propose a framework that encapsulates practical application in effectively classifying alopecia areata and healthy hairs using hair images as previous work has been carried out with scalp and skin images only. Our proposal demonstrates the practical application of machine learning techniques for distinguishing alopecia areata. The results from our study exhibit the future potential of this framework to distinguish hair disorders that cannot be determined by the naked eye.

2. Related Works

Most researchers have used scalp images to extract skin features characteristic of alopecia areata. A trichoscopy method was proposed that involved extraction of hair loss feature by processing of scalp images using encapsulated techniques such as grid line selection and eigenvalue. The system was novel in terms of using a combination of computer vision and image processing techniques for alopecia areata diagnosis [14]. In another study, an automated classification method for the early diagnosis and treatment of alopecia was proposed using artificial neural networks (ANN). The system used a feedforward artificial neural network, and the results exhibited an accuracy of 91% [15]. In another work, scalp images were classified according to three scalp conditions namely, alopecia areata, dandruff, and normal hair. The classification yielded an accuracy of 85% [16]. In another study, texture analysis was executed on scalp images using Severity of Alopecia Tool (SALT) score. The proposed system permitted analysis of hair density changes exhibited in alopecia areata [17].

Other systems have used scalp images to analyze hair density and loss that manifest due to various reasons, including alopecia areata. A system referred to as TrichoScan was developed using epiluminescence microscopy to analyze hair cover in scalp images of people with androgenic alopecia (AGA). Four parameters, namely, hair density, hair diameter, hair growth rate, and anagen/telogen ratio, were extracted, and the results reported a correlation of approximately 91% [21]. In another system, hair loss was diagnosed via the application of artificial neural networks (ANNs). Scalp images were acquired from three hundred and forty-eight participants, and the results of the study exhibited that artificial neural networks can be utilized for detecting hair loss [22]. Shih [23] captured forty microscopic scalp images with a magnification factor of eighty-five to propose a hair counting algorithm involving features such as density, diameter, length, and hair oiliness level. The algorithm was observed to be more accurate than the traditional Hough-based one and was more reliable in counting hairs on an individual's scalp as compared to manual counting.

Researchers have also used scalp images to develop machine learning models for diagnosing different diseases. An intelligent scalp analysis system was proposed employing different machine learning methods such as SVM, linear discriminant analysis (LDA), KNN, and decision trees. Classification was carried out between two groups, namely, bacteria 1 that pertains to blisters or boils in the scalp and bacteria 2 that comprises of scalp skin exhibiting red spots. The highest accuracy of 80% was achieved with the application of SVM [18]. Another scalp analysis system used optical coherence tomography (OCT) and machine learning to identify fungal infection. A-line features comprising of attenuation coefficient values and B-scan features involving texture parameters such as energy, kurtosis, and skewness were extracted from the captured scalp images. Classification was carried out between nondandruff and dandruff scalps with the application of machine learning algorithms including decision tree, SVM, neural network, and extreme learning machine (ELM). The highest accuracy of 87.5% was acquired via SVM followed by neural network, decision tree, and ELM with 83.3%, 79.16%, and 75.23% accuracies, respectively [19]. A webcam and microscope camera sensor system was proposed for executing a hair and scalp analysis with reference to the Norwood-Hamilton scale. K-means clustering was applied, and the level of baldness was determined. The results exhibited accuracy in the range of 71% to 84% for different circumstances such as oily scalp, swollen/red scalp, and dry scalp [20].

The literature review demonstrates that no work has been done with hair images for identification of alopecia areata (hair disorder). Previous work has been carried out with dermoscopic and scalp images. Similar image preprocessing steps were used in [16]; however, the study made use of scalp images and applied only SVM with 85% accuracy. Furthermore, feature extraction techniques were also different in [16] as compared to our proposed framework. Hence, our work demonstrates a novel and innovative framework for classifying alopecia areata using color, texture, and shape as features and SVM and KNN as classification algorithms.

3. Materials and Methods

3.1. Datasets

3.1.1. Healthy Hair Image Dataset. A total of 200 healthy hair images have been retrieved from the Figaro1k dataset. Figaro1k is a publicly available dataset containing different classes of hair images such as straight, wavy, and curly [24]. A normalization procedure has been applied on the dataset to ensure that the size and the aspect ratio of every image are

Table 1: Sample input images of alopecia areata and healthy hairs.

Classes	Images
Alopecia areata	
Healthy hair	

Table 1: Continued.

Classes	Images
	

the same [24]. Table 1 lists a few healthy hair images from the Figaro1k dataset that have been utilized in this study.

3.1.2. Alopecia Areata Image Dataset. A total of 68 hair images of alopecia areata are retrieved from the Dermnet dataset. The dataset available on Dermnet comprises of twenty-three categories of dermatological diseases, including alopecia areata [25]. Another type of disease images includes that of eczema, seborrheic keratoses, tinea ringworm, bullous disease, poison ivy, and psoriasis [25]. Table 1 illustrates a few alopecia areata images that we utilized.

3.2. Proposed Framework with SVM and KNN. To ensure that our data comprising the sample input images is organized and error-free, the dataframe function from Pandas Python Library is utilized to eliminate unwanted rows and columns and to clean the images. The code is written using Python on a Linux workstation utilizing the TensorFlow package with NVidia Titan GPU. The classification technique is executed with the aid of two machine learning methods, support vector machine (SVM) and k-nearest neighbor (KNN). The proposed flow process of the classification framework is exhibited in Figure 1. It starts initially with the input sample

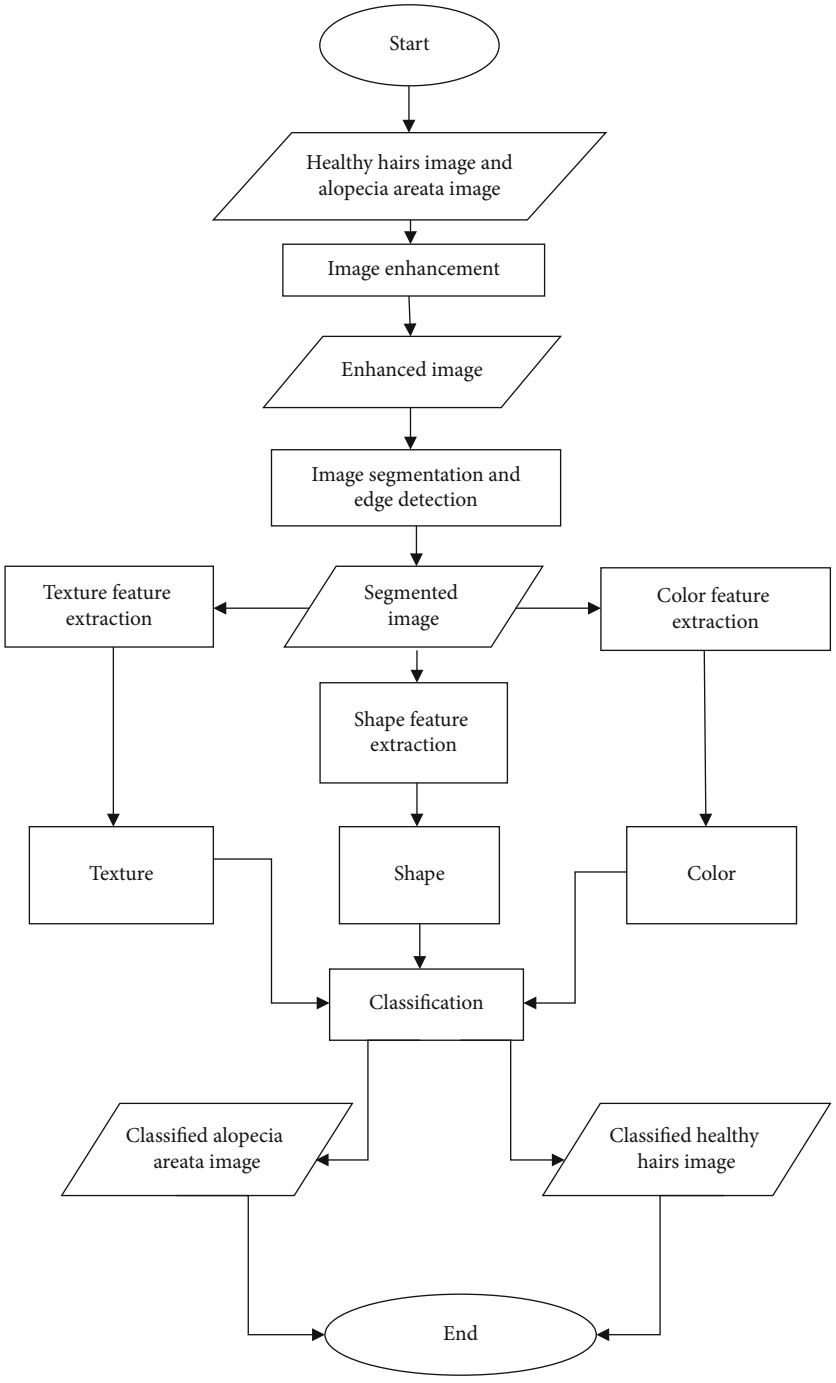


Figure 1: Proposed flow process of the classification framework.

images of healthy hairs and alopecia areata. This is followed by image enhancement process that permits getting rid of any unwanted deformation in the images. Following image enhancement, image segmentation and edge detection are carried out. Furthermore, three features, namely, color, texture, and shape, are extracted, and the classification process is executed. Empirical studies represent that more reliable results can be acquired if 20-30% of data is used for testing and 70-80% for training [26]. Hence, in this study, 70% of the images have been used for model training and the remaining 30% are utilized for testing. The end result is the classification of an image into alopecia areata (class 0) or healthy hairs (class 1).

3.3. Image Preprocessing

3.3.1. Image Enhancement. The technique of image enhancement relates to improving the contrast, brightness, and the pixel luminance values [27]. In this study, the sklearn.preprocessing library part of scikit-image processing that involves a vast array of techniques for image enhancement and image segmentation has been employed. The technique pertaining

to histogram equalization is used to enhance the sample input images. Histogram equalization (HE) tends to improve areas of lower local contrast and enhances the intensities that ultimately lead to increase in the global contrast of the sample input images [28]. In our study, histogram equalization (HE) is executed by converting the RGB image into an equivalent hue-saturation-value (HSV) image format. Histogram equalized intensity matrix is produced, and the image is enhanced. This is exhibited by Figures 2 and 3 demonstrating the sample image of alopecia areata and healthy hair before and after histogram equalization, respectively.

3.3.2. Image Segmentation and Edge Detection. The image segmentation operation pertains to dividing the constituents of an image into desired fragments or sections that have similar features like texture, intensity, and pixel values [29]. In this study, image segmentation has been carried out via the resize operation and edge detection. The resize operation resizes an image by a given scaling factor or dimension. In this study, the resize dimension was set to 64; hence, a segmented output image with a dimension factor of 64 was generated. A major technique of image segmentation is edge detection. Edge detection is used to identify curves in an image that follow a path pertaining to rapid change in the intensity of the image [30]. In this study, the antialiasing technique in relation to edge detection and as part of the scikit-image processing library of Python has been utilized. The antialiasing operation is set to true to denote that the rough edges in the images are smoothened. Figures 4 and 5 show sample images of alopecia areata and healthy hair before and after edge detection, respectively.

3.4. Feature Extraction. Our study involves the extraction of three features of color, texture, and shape from each input sample image. The libraries of Python used for color, shape, and texture feature extraction include cv2 and skimage.

3.4.1. Color Feature. In this study, the images have been converted to a NumPy array involving a list of color pixel values of RGB. The cv2 library is used to compute the mean of each of the three color channels including red, green, and blue. The first mean value that the cv2 library generates is of the blue channel, the second is of the green channel, and the third relates to the red channel. The cv2 library stores RGB images as NumPy array in a reverse order; hence, the first value corresponds to the blue channel, the second to the green channel, and the third value to the red channel.

3.4.2. Texture Feature. As part of this study, the skimage library and the cv2 library of Python have been imported to exploit scikit's image processing capability. Local Binary Patterns (LBPs) are employed as texture descriptors to compute the local representation of the texture feature. The local representation that aids in extracting the texture feature is constructed by comparing each pixel of the image with its surrounding neighborhood of pixels. The texture is extracted using LBPs, where gc is the intensity value of the central pixel and gp is the intensity of the neighboring pixel with index p

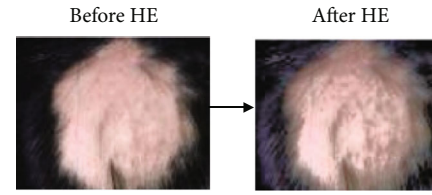


Figure 2: Sample outcomes of HE for alopecia areata.

as specified in the following equation:

$$\mathrm{LBP}(\mathrm{gp}x, \mathrm{gp}y) = \sum_{p=0}^{p-1} S(\mathrm{gp} - \mathrm{gc}) \times 2^p. \quad (1)$$

The function S can be expressed as

$$S(x) = \left\{ \begin{array}{l} 1 \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{array} \right\}. \quad (2)$$

3.4.3. Shape Feature. The OpenCV library of Python that has been imported in our study utilizes Hu moment shape descriptor to extract the shape feature, where h denotes the computed Hu moment and η represents the normalized central moment. The Hu moment shape descriptor is exhibited in Equation (3) and Equation (4). Central moment is involved in the computation of Hu moments as they shift the center of the image to the centroid region. Furthermore, Hu moments incorporating central moments tend to be invariant to translation, scale, and rotation that help in the extraction of the shape feature. Hu moments are able to extract shape features by quantifying the outline of the sample input images thus yielding the NumPy array form of the images. Finally, the flatten operation flattens the NumPy array to produce the shape feature vector.

$$h_O = \eta_{2O} + \eta_{O2}, \quad (3)$$

$$h_1 = (\eta_{2O} - \eta_{O2})^2 + 4\eta_{11}^2. \quad (4)$$

3.5. Classification. In this study, support vector machine (SVM) and k-nearest neighbor (KNN) have been utilized for classifying healthy and alopecia areatahair images into their accurate classes. Figure 6 depicts the architecture of our framework, including the training and testing phases. The initial steps in the framework are concerned with image preprocessing and feature extraction. This is followed by model training with machine learning algorithms and then execution of the testing phase.

3.6. Mathematical Operations of Classifiers

3.6.1. Mathematical Operations of SVM. Support vector machine (SVM) determines the linear and nonlinear separability with the aid of a hyperplane [31]. Its kernel method transforms two-dimensional nonlinearly separable data into higher dimensions that yield the optimal hyperplane to separate the data [32]. The kernel trick employs the multiplication of a kernel function k with the dot product $x_i \cdot x_j$ as

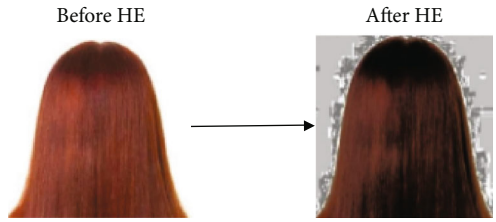


Figure 3: Sample outcomes of HE for healthy hair.

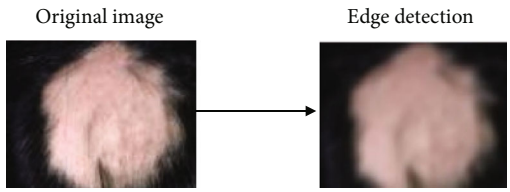


Figure 4: Sample outcomes of edge detection for alopecia areata.

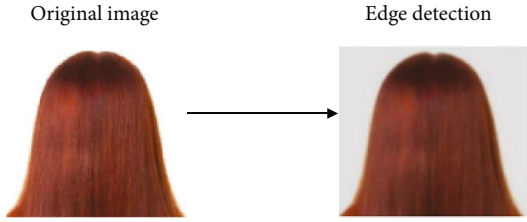


Figure 5: Sample outcomes of edge detection for healthy hair.

represented by

$$\max_{\propto} \sum_{i=1}^{m} \propto_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \propto_i \propto_j y_i y_j K\left(x_i \cdot x_j\right). \tag{5}$$

In this study, the radial basis function (RBF) has been used as the kernel function. It is also referred to as the Gaussian kernel and contains a parameter γ as shown in the following equation:

$$K\left(x_i.x_j\right) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right). \tag{6}$$

3.6.2. Mathematical Operations of KNN. The k-nearest neighbor algorithm pertains to finding the nearest neighbors. The process involves finding the nearest point that lies close to the input point in a given dataset [33]. In this study, the neighbors are specified as three which denotes that for every new input data, the three closest neighbors will be evaluated for classification. The algorithm initially analyzes the Euclidean distance that transforms data points into mathematical values.

The Euclidean distance formula in Equation (7) finds the distance between two points in a plane with coordinates (x, y) and (a, b).

$$\text{dist}((x, y), (a, b)) = \sqrt{(x-a)^2 + (y-b)^2}. \tag{7}$$

3.7. SPSS Analysis. The paired sample T-test was performed via Statistical Package for Social Sciences (SPSS), IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp., on the accuracies generated from both SVM and KNN. The number of accuracy samples for both the algorithms was thirty.

4. Results and Evaluation

4.1. Performance Evaluation. The performance evaluation of support vector machine (SVM) and k-nearest neighbor (KNN) is evaluated using confusion matrices. The confusion matrix exhibited in Figure 7 demonstrates the predicted outcomes for the two classes. The two classes, alopecia areata and healthy hairs, have been denoted by 0 and 1, respectively. When the actual value is 1 and the predicted value is also 1, then the outcome is true positive (TP); otherwise, the outcome is false negative (FN). On the contrary, when the actual value is 0 and the predicted value is also 0, then the outcome is true negative (TN); otherwise, false positive (FP) is generated.

Figure 8 exhibits the confusion matrix formulated after application of support vector machine (SVM) and represents that out of the 81 images tested, 74 images were classified accurately, thus yielding an accuracy of 91.4%. 22 images were classified as alopecia areata, and 52 images were classified as healthy hairs.

Figure 9 exhibits the confusion matrix generated after the application of k-nearest neighbor (KNN) and shows that out of the total 81 images tested, 72 images were classified accurately, thus yielding an accuracy of 88.9%. 24 images were classified as alopecia areata, and 48 images were classified as healthy hairs. The reported accuracies were achieved after 10-fold cross-validation.

The performance and classification results for SVM and KNN are shown in Table 2. The accuracy for both SVM and KNN can be calculated by dividing the number of truly classified images by the total number of test images and multiplying the result with 100 as expressed in the following equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \times 100\%. \tag{8}$$

Equation (9) demonstrates the results obtained via SPSS analysis.

$$\begin{aligned} t\,(29) &= 4.744, \\ p &< 0.001. \end{aligned} \tag{9}$$

4.2. Performance Metrics. The sklearn library in Python helps to compute the true positives and false positives and true negatives and false negatives [34]. Classification techniques encapsulate classification metrics, namely, precision, recall, and F1 score. Precision relates to the ability of a classifier to be precise, i.e., its capacity to not to mark a positive circumstance that is actually negative. Recall is the ability of a classifier to identify all the true positives. Recall can be defined for each class as the ratio of true positives to the summation of true positives and false negatives [35]. F1 score can be

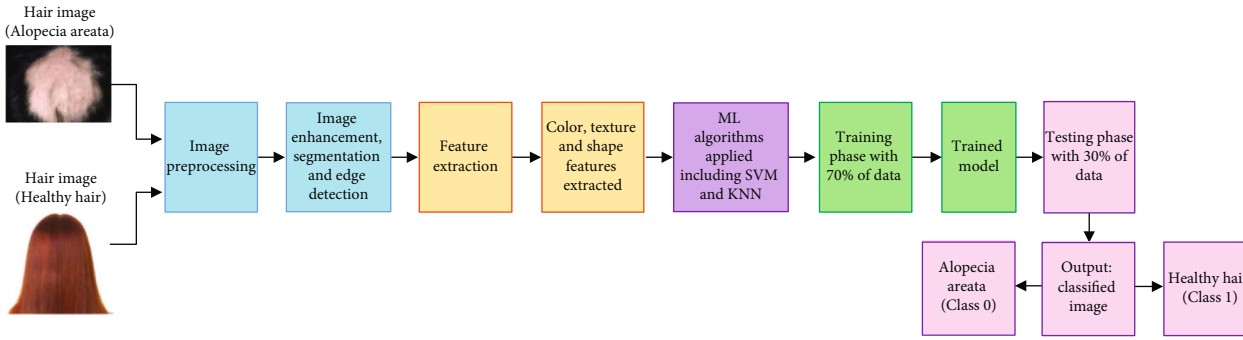


Figure 6: Architecture of the proposed framework. Top left: sample input images of alopecia areata and healthy hairs. Turquoise block: image preprocessing. Orange block: color, texture, and shape feature extraction. Purple block: SVM and KNN application. Green block: training phase. Pink block: testing phase.

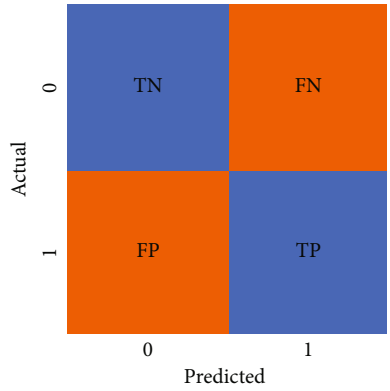


Figure 7: Confusion matrix.

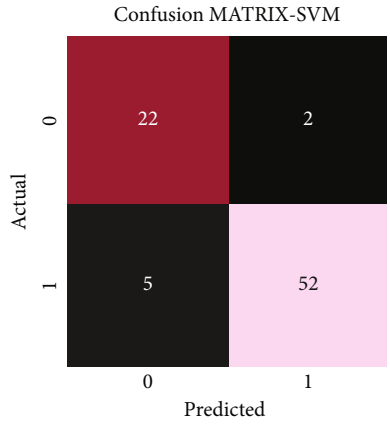


Figure 8: Confusion matrix of SVM illustrating the classified images of both alopecia areata (class 0) and healthy hairs (class 1).

defined as a score that exhibits the relationship between precision and recall [35]. In order to evaluate and analyze the effectiveness of the classification framework, the performance metrics shown in Equations (10), (11), and (12) including precision, recall, and F1 score were employed, where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. The values of these performance metrics have been computed as per their formulas and are represented in Table 3.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%, \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \tag{11}$$

$$\text{F1 score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \times 100\%. \tag{12}$$

5. Discussion

5.1. Principal Findings. The objectives of this study were to extract the color, texture, and shape features from healthy and alopecia areata hair images and apply machine learning algorithms including support vector machine (SVM) and *k* -nearest neighbor (KNN) to execute classification of the images. SVM can classify linear and nonlinear data by generating a line or a hyperplane [31]. The RBF kernel method employed in our study aids to transform the data that yields the optimal hyperplane and in turn generates higher accuracy. On the contrary, KNN uses the Euclidean distance function to compute the probability of the test inputs that are closer to the data points [36]. Our study demonstrates better performance of SVM as compared to KNN when classifying hair images into healthy versus alopecia areata.

Other systems that have used the same machine learning techniques have also shown higher accuracies. For example, a skin lesion classification system based on support vector machine (SVM) and *k*-nearest neighbor (KNN) resulted in the accuracies of 89.50% and 82.00% for support vector machine (SVM) and *k*-nearest neighbor, respectively [37]. In another study, dermoscopic images were utilized for the classification of skin cancer using support vector machine (SVM), *k*-nearest neighbor (KNN), and random forest. The results demonstrated that support vector machine (SVM) performed better than the other two classifiers [38]. Better performance of support vector machine (SVM) lies in its mathematical operations. Hence, it can be deduced that the higher accuracy of 91.4% by support vector machine (SVM) is due to the use of kernel function that transforms the data into higher dimensions and yields the optimal hyperplane.

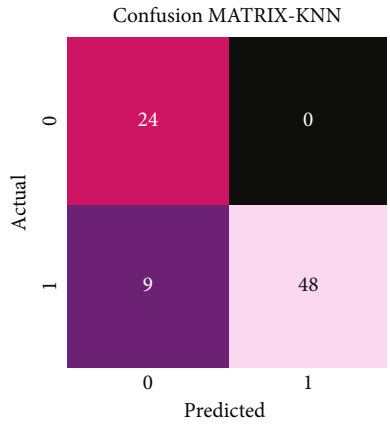


Figure 9: Confusion matrix of KNN illustrating the classified images of both alopecia areata (class 0) and healthy hairs (class 1).

Table 2: Classification results of SVM and KNN.

ML algorithm	Test images	TRUE classification	Validation	% accuracy
SVM	81	74	10-fold	91.4
KNN	81	72	10-fold	88.9

Table 3: Values of performance metrics including precision, recall, and F1 score for both alopecia areata (class 0) and healthy hair (class 1).

ML algorithm	Class	Precision (%)	Recall (%)	F1 score (%)
SVM	0	81.5	91.7	86.3
SVM	1	96.3	91.2	93.7
KNN	0	72.7	100	84.2
KNN	1	100	84.2	91.4

The strength of our proposed framework lies in being the first of its kind to classify alopecia areata and healthy hairs using hair images. The limitation of this study includes applying machine learning (ML) techniques on a limited dataset with no clinical data being collected. Nevertheless, it has been widely observed that deep learning techniques such as convolutional neural network (CNN) tend to generate high accuracies [39]. Furthermore, deep learning applications do not involve image preprocessing and feature extraction [39, 40]. An intelligent scalp analysis system was proposed employing convolutional neural network (CNN), and results exhibited accuracy of 89.77% [18]. Hence, future work can be carried out using CNN so that even better classification performance can be achieved.

6. Conclusion

This study proposed a classification framework for healthy hairs and alopecia areata using hair images with features including color, texture, and shape being extracted and support vector machine (SVM) and *k*-nearest neighbor (KNN) being applied. The application of support vector machine (SVM) and *k*-nearest neighbor (KNN) presented an accuracy of 91.4% and 88.9%, respectively. These accuracies exhibit that the proposed classification framework has been found to be successful and robust in classifying two different sets of hair images. However, future work with deep learning techniques such as convolutional neural networks (CNN) can also be carried out and integrated with the existing system.

Data Availability

The datasets used are available at the following links: (1) http://projects.i-ctm.eu/it/progetto/figaro-1k and (2) https://www.kaggle.com/shubhamgoel27/dermnet.

Conflicts of Interest

The authors have no conflict of interest in this study.

References

[1] N. S. Sadick, "New-generation therapies for the treatment of hair loss in men," *Dermatologic Clinics*, vol. 36, no. 1, pp. 63–67, 2018.

[2] C. H. Pratt, L. E. King, A. G. Messenger, A. M. Christiano, and J. P. Sundberg, "Alopecia areata," *Nature Reviews Disease Primers*, vol. 3, no. 1, pp. 1–17, 2017.

[3] A. A. Alshahrani, R. Al-Tuwaijri, Z. A. Abuoliat, M. Alyabsi, M. I. AlJasser, and R. Alkhodair, "Prevalence and clinical characteristics of alopecia areata at a tertiary care center in Saudi Arabia," *Dermatology Research and Practice*, vol. 2020, Article ID 7194270, 4 pages, 2020.

[4] A. Sterkens, J. Lambert, and A. Bervoets, "Alopecia areata: a review on diagnosis, immunological etiopathogenesis and treatment options," *Clinical and Experimental Medicine*, vol. 21, no. 2, pp. 215–230, 2021.

[5] T. Simakou, J. P. Butcher, S. Reid, and F. L. Henriquez, "Alopecia areata: a multifactorial autoimmune condition," *Journal of Autoimmunity*, vol. 98, pp. 74–85, 2019.

[6] N. Meah, D. Wall, K. York et al., "The alopecia areata consensus of experts (ACE) study part II: results of an international expert opinion on diagnosis and laboratory evaluation for alopecia areata," *Journal of the American Academy of Dermatology*, vol. 84, no. 6, pp. 1594–1601, 2021.

[7] H. Sahli, "An introduction to machine learning," in *TORUS 1–Toward an Open Resource Using Services: Cloud Computing for Environmental Data*, pp. 61–74, Wiley Online Library, 2020.

[8] G. Mathiyalagan and D. Devaraj, "A machine learning classification approach based glioma brain tumor detection," *International Journal of Imaging Systems and Technology*, vol. 31, pp. 1424–1436, 2021.

[9] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk, and H. Chen, "A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique," *IEEE Access*, vol. 9, pp. 71194–71209, 2021.

[10] Z. Yue, C. Sun, F. Chen et al., "Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis," *Biomedical Optics Express*, vol. 12, no. 5, pp. 2559–2574, 2021.

[11] S. Zhao, P. Wang, A. A. Heidari et al., "Multilevel threshold image segmentation with diffusion association slime mould

algorithm and Renyi's entropy for chronic obstructive pulmonary disease," *Computers in Biology and Medicine*, vol. 134, article 104427, 2021.

[12] R. Pangti, J. Mathur, V. Chouhan et al., "A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases," *Journal of the European Academy of Dermatology and Venereology*, vol. 35, no. 2, pp. 536–545, 2021.

[13] P. Wu, H. Ye, X. Cai et al., "An effective machine learning approach for identifying non-severe and severe coronavirus disease 2019 patients in a rural Chinese population: the Wenzhou retrospective study," *IEEE Access*, vol. 9, pp. 45486–45503, 2021.

[14] S. Seo and J. Park, "Trichoscopy of alopecia areata: hair loss feature extraction and computation using grid line selection and eigenvalue," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 6908018, 9 pages, 2020.

[15] I. Kapoor and A. Mishra, "Automated classification method for early diagnosis of alopecia using machine learning," *Procedia Computer Science*, vol. 132, pp. 437–443, 2018.

[16] S. Ibrahim, Z. A. Noor Azmy, N. N. Abu Mangshor, N. Sabri, A. F. Ahmad Fadzil, and Z. Ahmad, "Pre-trained classification of scalp conditions using image processing," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 1, pp. 138–144, 2020.

[17] E. Bernardis and L. Castelo-Soccio, "Quantifying alopecia areata via texture analysis to automate the salt score computation," *Journal of Investigative Dermatology Symposium Proceedings*, vol. 19, no. 1, pp. S34–S40, 2018.

[18] W.-C. Wang, L.-B. Chen, and W.-J. Chang, "Development and experimental evaluation of machine-learning techniques for an intelligent hairy scalp detection system," *Applied Sciences*, vol. 8, no. 6, 2018.

[19] K. Dubey, V. Srivastava, and D. S. Mehta, "Automatedin vivoidentification of fungal infection on human scalp using optical coherence tomography and machine learning," *Laser Physics*, vol. 28, no. 4, article 045602, 2018.

[20] S.-H. Lee and C.-S. Yang, "An intelligent hair and scalp analysis system using camera sensors and Norwood-Hamilton model," *International Journal of Innovative Computing Information and Control*, vol. 14, no. 2, pp. 503–518, 2018.

[21] R. Hoffmann, "TrichoScan: a novel tool for the analysis of hair growth *in vivo*," *Journal of Investigative Dermatology Symposium Proceedings*, vol. 8, no. 1, pp. 109–115, 2003.

[22] A. Esfandiari, K. R. Kalantari, and A. Babaei, "Hair loss diagnosis using artificial neural networks," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, 2012.

[23] H. Shih, "A precise automatic system for the hair assessment in hair-care diagnosis applications," *Skin Research and Technology*, vol. 21, no. 4, pp. 500–507, 2015.

[24] M. Svanera, U. R. Muhammad, R. Leonardi, and S. Benini, "Figaro, hair detection and segmentation in the wild," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 933–937, Phoenix, AZ, USA, 2016.

[25] T. A. Rimi, N. Sultana, and M. F. A. Foysal, "Derm-NN: skin diseases detection using convolutional neural network," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1205–1209, Madurai, India, 2020.

[26] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation," Technical Report: UTEP-CS-18-09, 2018.

[27] P. K. Verma, N. P. Singh, and D. Yadav, "Image enhancement: a review," in *Ambient Communications and Computer Systems*, pp. 347–355, Springer, 2020.

[28] H. Singh, *Practical Machine Learning and Image Processing*, Springer, 2019.

[29] A. Bali and S. N. Singh, "A review on the strategies and techniques of image segmentation," in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, pp. 113–120, Haryana, India, 2015.

[30] S. C. Shekar and D. Ravi, "Image enhancement and compression using edge detection technique," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 5, 2017.

[31] R. Liao, *Support Vector Machines*, CSC, 2015.

[32] K. Tajima, Y. Hirohashi, E. R. R. Zara, and T. Kato, "Frank-Wolfe algorithm for learning SVM-type multi-category classifiers," in *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pp. 432–440, Philadelphia, PA, USA, 2021.

[33] L.-Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, "The distance function effect on k-nearest neighbor classification for medical datasets," *Springerplus*, vol. 5, no. 1, pp. 1–9, 2016.

[34] R. Garreta and G. Moncecchi, *Learning Scikit-Learn: Machine Learning in Python*, Packt Publishing Ltd, 2013.

[35] J. Brownlee, *Machine Learning Mastery with Python: Understand your Data, Create Accurate Models, and Work Projects End-to-End*, Machine Learning Mastery, 2016.

[36] S. Zhang, "Challenges in KNN classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, 2021.

[37] M. Koklu and I. A. Ozkan, "Skin lesion classification using machine learning algorithms," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 4, no. 5, pp. 285–289, 2017.

[38] A. Murugan, S. A. H. Nair, and K. S. Kumar, "Detection of skin cancer using SVM, random forest and kNN classifiers," *Journal of Medical Systems*, vol. 43, no. 8, pp. 1–9, 2019.

[39] J. Ahmad, H. Farman, and Z. Jan, "Deep learning methods and applications," in *Deep Learning: Convergence to Big Data Analytics*, pp. 31–42, Springer, 2019.

[40] L. Peng, M. Peng, B. Liao, G. Huang, W. Li, and D. Xie, "The advances and challenges of deep learning application in biological big data processing," *Current Bioinformatics*, vol. 13, no. 4, pp. 352–359, 2018.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 4244157, 12 pages
https://doi.org/10.1155/2021/4244157

Research Article

A Computationally Virtual Histological Staining Method to Ovarian Cancer Tissue by Deep Generative Adversarial Networks

Xiangyu Meng,[1,2] **Xin Li**,[3] **and Xun Wang**[1,4]

[1]*College of Computer Science and Technology, China University of Petroleum, Qingdao, 266580 Shandong, China*
[2]*College of Computer and Information Science, Inner Mongolia Agricultural University, Huhhot, 010018 Inner Mongolia, China*
[3]*Department of Gynecology 2, Renmin Hospital of Wuhan University, Wuhan, 430060 Hubei, China*
[4]*China High Performance Computer Research Center, Institute of Computer Technology, Chinese Academy of Science, Beijing, 100190 Beijing, China*

Correspondence should be addressed to Xin Li; xin.li@whu.edu.cn and Xun Wang; wangsyun@upc.edu.cn

Received 20 April 2021; Accepted 10 June 2021; Published 2 July 2021

Academic Editor: Pan Zheng

Histological analysis to tissue samples is elemental for diagnosing the risk and severity of ovarian cancer. The commonly used Hematoxylin and Eosin (H&E) staining method involves complex steps and strict requirements, which would seriously impact the research of histological analysis of the ovarian cancer. Virtual histological staining by the Generative Adversarial Network (GAN) provides a feasible way for these problems, yet it is still a challenge of using deep learning technology since the amounts of data available are quite limited for training. Based on the idea of GAN, we propose a weakly supervised learning method to generate autofluorescence images of unstained ovarian tissue sections corresponding to H&E staining sections of ovarian tissue. Using the above method, we constructed the supervision conditions for the virtual staining process, which makes the image quality synthesized in the subsequent virtual staining stage more perfect. Through the doctors' evaluation of our results, the accuracy of ovarian cancer unstained fluorescence image generated by our method reached 93%. At the same time, we evaluated the image quality of the generated images, where the FID reached 175.969, the IS score reached 1.311, and the MS reached 0.717. Based on the image-to-image translation method, we use the data set constructed in the previous step to implement a virtual staining method that is accurate to tissue cells. The accuracy of staining through the doctor's assessment reached 97%. At the same time, the accuracy of visual evaluation based on deep learning reached 95%.

1. Introduction

Computer-aided medical diagnosis is a hot topic nowadays. In recent years, researchers devoted to the issue in this direction and achieved excellent research outcomes. Some studies [1–5] established spiking neural networks to simulate biological metabolic processes, infer the final physiological calculation results, and obtain the final diagnosis solution. Some studies [6–10] rely on the idea of deep learning to build a deep neural network model to complete the diagnosis of patients based on the characteristics of various medical data. The above methods have achieved extraordinary accuracy and efficient treatment for specific medical domain. However, few studies have focused on the preparation and labeling of medical data. At present, traditional medical data preprocessing methods can no longer meet the needs of intelligent diagnosis with large data volumes.

The ovarian cancer is a global problem, is typically diagnosed at a late stage, and has no effective screening strategy [11]. Microscopic imaging of tissue samples is the basis for subsequent diagnosis and prognosis of cancer. H&E staining and labeling of tissue samples can better help locate suitable cancer tissues and perform subsequent analysis, diagnosis, and prognosis. Therefore, strict control of tissue section staining standards will significantly enhance the final diagnosis and prognosis results. However, the traditional histopathological section staining process involves many standard operating steps, and each technician must strictly adhere to these gold standards. These methods are often time-consuming and laborious and often have higher requirements. The

histological tissue appearance may assume different color intensities depending on the staining process, operator ability, and scanner specifications [12]. The above problems will seriously harm the analysis of the disease pathology of the tissue and the events of the disease prognosis. Many coping strategies have been established to improve the success rate and quality of tissue staining in clinical trials. Massimo et al. [12] presented a novel fully automated stain separation and normalization approaches for Hematoxylin and Eosin stained histological slides to improve the contrast between histological tissue and background and preserve local structures without changing the color of the lumen and the background. Mario et al. [13] used experiments to clarify Eosin-based fluorescence spectroscopy can be used to directly examine H&E stained tissue slides. Relevant areas can be imaged and spectral analysis done to obtain objective data. The above method does solve the problems of low accuracy and poor effect in traditional staining to a certain extent. However, the first method still cannot avoid the strict and complicated processing steps of the traditional method. Although the second method uses a more novel method of labeling pathological tissues, the use of H&E staining analysis is still the mainstream processing method for pathology research today. The popularization and promotion of this program are still a long process.

We propose to construct a computationally staining and labeling algorithm for H&E staining of the ovarian cancer tissue sections. This method can effectively avoid the complicated steps of H&E staining of traditional ovarian cancer tissue sections and make up for the diagnosis tension caused by the lack of data. Rivenson et al. [14] proposed a virtual staining method for pathological sections based on deep learning. They placed the fresh tissue section on a fluorescence microscope to observe the autofluorescence imaging of the tissue sample and then stained and labeled the sample to obtain the corresponding stained image. After repeating the above work, a huge data set is constructed, and then, a deep learning model is performed to complete the feature learning from unstained samples to stained samples. However, the above method requires a large number of tissue samples from the patient and cannot be effectively implemented under multiple limited conditions. At the same time, the process of data construction still cannot completely get rid of the traditional H&E staining process.

To solve the problems above, we firstly proposed a weakly supervised image generation algorithm based on the CycleGAN model [15], which generate the corresponding unstained image for the stained ovarian sections. Figure 1(a) shows the overview of this domain translation method. We introduce the domain consistency loss based on the original CycleGAN model, to ensure that the results after the cycle generation are accurately matched to the specific domain. The introduction of input buffers can better magnify the effect of domain consistency loss. We only collected 400 H&E staining images of ovarian cancer and 80 autofluorescence images from other tissues as the experimental data set. From the experimental results, it can be seen that under such extremely inconsistent distribution conditions, the construction of H&E staining images to corresponding unstained images can be completed based on our method. The data set constructed by the above method can provide a good data guarantee for this goal. Figure 1(b) shows the overview of the virtual staining process. We analyzed whether the state-of-the-art image translation model can be effectively used in this experimental environment, but it is a pity that although these methods have some effects, they cannot meet our requirements for absolute fineness. Therefore, we made improvements on the traditional UNet basic framework [16] and proposed the Parallel Feature Fusion Network (PFFN). At the same time, we introduced a more superior training method to better fit the model to the optimal state. Compared with the traditional image translation method, the quality of the image generated by our improved image translation method is superior.

In this work, we mainly solved two problems. The first is that due to the limited number of pathological tissue samples, we provide a method for constructing a virtual data set composed of autofluorescence imaging of ovarian cancer pathological tissues and corresponding H&E staining imaging. Using our method, limited ovarian tissue images can be augmented with high quality in a short time. Next, based on our augmented data, we propose a virtual staining method. Using this method can swiftly and efficiently execute virtual staining of ovarian cancer pathological slices, and the quality of the generated virtual stained images is guaranteed. We have also compared with previous methods; the latter cannot exceed our proposed method in terms of image quality or evaluation accuracy. Our Code is available at https://github.com/menggerSherry/ImageStain.

2. Materials and Methods

2.1. Related Work

2.1.1. Generative Adversarial Network. Generative Adversarial Network was first proposed by Goodfellow et al. [17]. Different kinds of GANs model have shown its remarkable data generation especially in the computer vision domain. Recently, successful research such as image generation [18–20], image-to-image translation [15, 21, 22], and superresolution [23, 24] shown remarkable result. Traditional structure of GAN contains two networks: a generator and a discriminator. The generator learns from a random noise to images which is same as the train set. The discriminator learns to distinguish the real image in the data set and the fake image generated by the generator. The propose of the idea of GAN can produce better image results through the continuous adversarial training between the generator and the discriminator. However, there are still some problems in GAN training, such as unstable training and model collapse.

2.1.2. Conditional GANs. Traditional GAN model has shown very powerful data generation capabilities. However, we cannot artificially control the generation state of GAN and let the model generate the image we need. Mirza and Osindero [25] successfully solved this problem. Many researchers control GAN to generate data purposefully by imposing some conditions in training and introduce many conditional GAN models. Researchers have made many improvements to

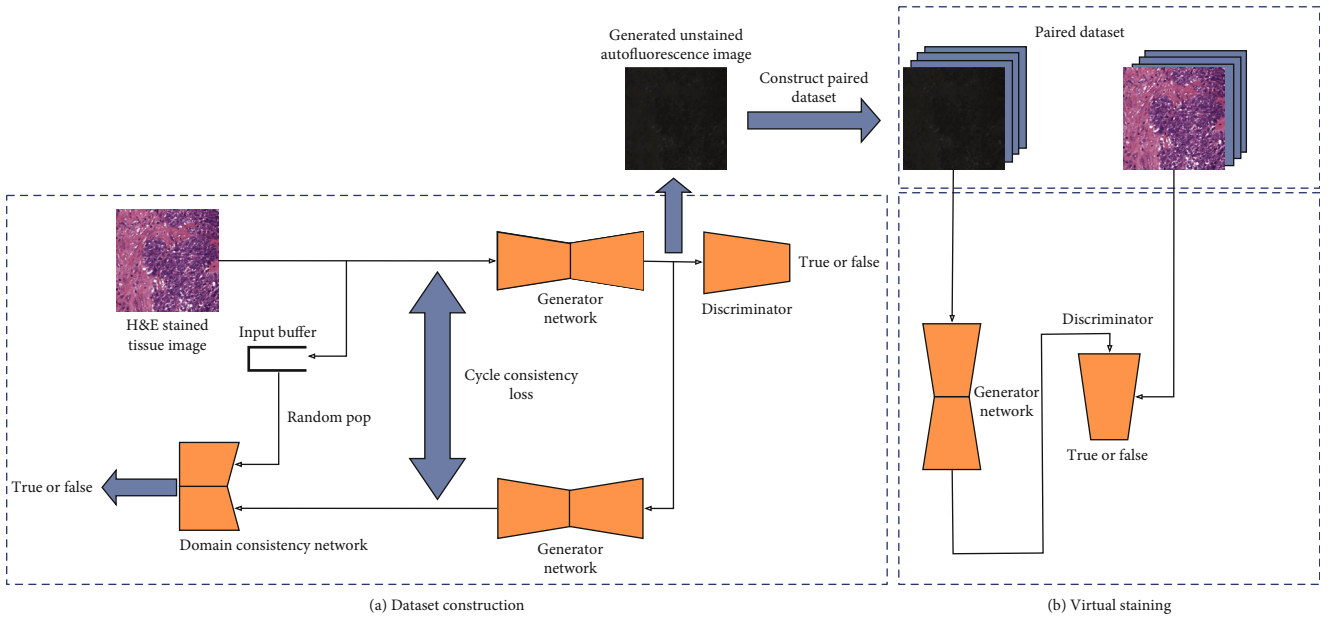


Figure 1: Overview of this virtual staining process: (a) the overview of this domain translation method; (b) overview of the virtual staining process.

conditional GAN, making conditional GAN widely used. Research on the conditional GAN is widely welcomed in the face editing [26], domain transfer [27], and photo editing [28]. Today's conditional GAN has not only been widely used in the direction of computer vision but also began to receive attention in audio and language processing. As conditional GAN has been widely used, its problems have gradually emerged. During the training process, it is easy to fit only to several optimal directions, which eventually leads to the collapse of the model and loss of model diversity. At the same time, in the training process, user cannot control the training progress of the generator and the discriminator, making GAN very unstable during the training process.

2.1.3. Improvement of GAN. Model collapse and training instability seriously affected the final experimental results. There are many aspects of research that have begun to solve these two fatal problems and have achieved good results. Raford et al. [20] use deep convolutional networks to design generators and discriminators and adopt batch normalization. The proposal of DCGAN solves the problem of unstable training and the model collapse. At the same time, applying CNN to the network structure can better adapt to the processing of images. Martin et al. [18, 19, 29] analyzed in detail the reasons for the collapse of the GAN model and the unstable training. They modified the loss of the original GAN to Wasserstein loss [18] and added a gradient penalty [19] so that the GAN model completely avoided these two problems.

2.1.4. Image-to-Image Translation. The image-to-image translation has been a hot topic since GAN was proposed. Because of the wide range of uses of this type of problem, many researchers have begun research in this field. This issue was first raised by Isola et al. [21]. They modified the conditional GAN and finally achieved excellent image translation results. Today, image translation has achieved remarkable results in domain translation [15, 26], superresolution image synthesis [22, 30], video synthesis [31], etc. The problem we are facing now is the H&E staining of ovarian cancer pathological slices. Inspired by the above successful cases, we used the idea of conditional GAN to improve a new network structure and training strategy and finally realized this virtual staining of ovarian cancer. Through the final verification stage, we found that significant success has been achieved in both efficiency and effectiveness.

2.2. Construction of the Paired Data Set. Our goal is to finish the accurate staining of ovarian cancer tissue. This means that every cell structure can be accurately stained. Therefore, we intend to build a fully supervised image to image translation model. With reference to the method of Rivenson et al. [14], the unstained image is obtained by placing a fresh tissue section of ovarian cancer in a fluorescence microscope for direct observation. Then, performing elaborate staining on this fresh tissue to get the stained image. In this way, each unstained image corresponds to a stained image as a label. It is indeed feasible to construct a perfect data set by repeatedly conducting the above steps. The work of Rivenson et al. did give marvelous results. But when we implement their idea, we found that collecting so many fresh sections in a short time is indeed not an easy task, and it is laborious and tiring to do these jobs repeatedly. There is also a problem that the data set is limited. We have a large number of H&E staining images of ovarian cancer, but the number of autofluorescence images of fresh slices is very rare, which makes the distribution of the two sets of data very uneven. Due to some of the above problems, we decided to abandon the method of Rivenson et al. and propose a deep learning method to complete the construction of the above-paired data set under limited supervision conditions.

2.2.1. CycleGAN Baseline. Unpaired image to image problem is an important problem. In many cases, building a paired data set takes a lot of time and work, but using unpaired image to image translation can avoid time-consuming data collection. The CycleGAN proposed by Zhu et al. [15] is a good model to solve such problems. Zhu et al. proposed a cycle consistency loss, $\|F_{X\longrightarrow Y}(G_{Y\longrightarrow X}(x)) - x\|_1$, $\|G_{Y\longrightarrow X}(F_{X\longrightarrow Y}(x)) - x\|_1$ finished the translation of the domain X and the domain Y. Using this unsupervised method can effectively achieve conversion between two distributed data. But as mentioned above, the data distribution between our existing two domains is extremely uneven. In our experiment, we used the CycleGAN model to complete the data construction, but the results were very disappointing.

Figure 2(a) is the result of training using the CycleGAN model. We can see that there are a large number of sharp holes in the image. The input stained image's positions corresponding to these holes have normal textures, indicating that this model cannot effectively learn the features corresponding hole domain. We then introduce a large number of data augmentation methods based on the original CycleGAN, such as random jitter, random horizontal and vertical flips, random jitter rotation (first interpolate and zoom, then randomly rotate a small angle, and finally crop to the original image size), and elastic deformation [16]. We can observe the results as shown in Figure 2(b). The number of cavities is obviously reduced, and the overall image quality has been slightly improved, but the existence of cavities is still not completely resolved.

We zoomed on the position of the hole as shown in Figure 2(c). It can be found that the generator and discriminator in these positions did not play their role at all. We hypothesis that the feature distribution of the unstained image domain is sparser than the stained image domain's. The essence of domain to domain translation is to learn the features of images in a domain and then translate these features based on the supervision of the image feature rules of the target domain. Based on the above assumptions, when converting from a stained domain to an unstained domain, the features of the stained domain learned by the generator may be difficult to be reasonably represented by the limited positive sample image features of the unstained domain distribution. This will affect the discriminator's training on the area where these features are located. The generator synthesized a black hole in this area, which can make the discriminator think it is true, resulting in a train mode collapse in this area. With continuous training, the effect in areas with sufficient supervised positive sample features is getting better and better; this area remains unchanged, and the black hole continues to become obvious.

2.2.2. Domain Consistency Network. When data augmentation is introduced, the number of samples theoretically increases, but with the increase in the number of data augmentation methods introduced, many augmented images may appear only a few times during training, which will cause the underfitting issue. We therefore introduce an input buffer to store the input image after a large amount of data augmentation and then randomly select the input image from the buffer as the input of the network. At the same time, in order to enhance the fitting ability of the network, we introduce the domain consistency loss. Specifically, a domain consistency discriminator is introduced to distinguish which domain the image belongs to. It participates in the training with the generator. In this way, through continuous training, the generator can synthesize images that are more accurate to a specific domain.

Figure 3 describes the structure of the domain consistency network and its training process. Where SepConv is the Deep Separable Convolution, LReLU is the LeakyReLU activation operation. The network first downsampling the input image on both sides three times to extract the effective features of the image, then fuses the two features and performs a series of convolution layers with 1×1 kernel to extract relevant information from the fused features and obtain a single-channel result. We ensure that the output dimension of the network is the same as the output dimension of the discriminator. In this way, real images with a large amount of data augmentation are first pushed into the buffer, and then, the buffer randomly selects two batches of images as the real image input of the domain consistency network. The domain consistency network learns that they belong to the same domain. The image generated by the generator and the real image randomly selected from the buffer are used as the input of the domain consistency network, and the network learns to distinguish that they belong to different domains. While learning to fool the discriminator, the generator also needs to fool the domain consistency network so that the domain consistency network thinks that the generated image and the real image are in the same domain.

We define the mapping $G_1 : X \longrightarrow Y$ as the conversion process from the stained image domain X to the unstained image domain Y, and its corresponding domain consistency network is C_1. We can describe the domain consistency loss as:

$$L(G_1, C_1, X, Y_1, Y_2) = \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\left(C_1(\text{Aug}(Y_1), \text{Aug}(Y_2))\right)^2\right] + \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\left(1 - C_1(\text{Aug}(Y_1), G_1(X))\right)^2\right]. \tag{1}$$

The Aug in Equation (1) represents the corresponding data augmentation; $\text{Aug}(Y_1)$ and $\text{Aug}(Y_2)$, respectively, represent the unstained image randomly selected from the input buffer after the augmentation transformation. Here, G_1 tries to generate an unstained image $G_1(x)$ that is very close to the representative $\text{Aug}(Y_1)$ of the unstained image domain, and C_1 tries to distinguish whether the two input images $\text{Aug}(Y_1)$ and $\text{Aug}(Y_2)$ are in the same domain. Like the idea of adversarial training, G_1 tries to minimize the objective function of equation; C_1 tries to maximize the objective function of Equation (1), which is expressed as $\min_{G_1}\max_{C_1} L(G_1, C_1, X, Y_1, Y_2)$. Similarly, to ensure the balance of training, we use $\min_{G_2}\max_{C_2} L(G_2, C_2, Y, X_1, X_2)$ to represent the domain consistency loss affecting the stained image.

The image shown in Figure 2(d) is the result of the unstained image synthesized by the generator after we introduce the domain consistency network. It can be found that

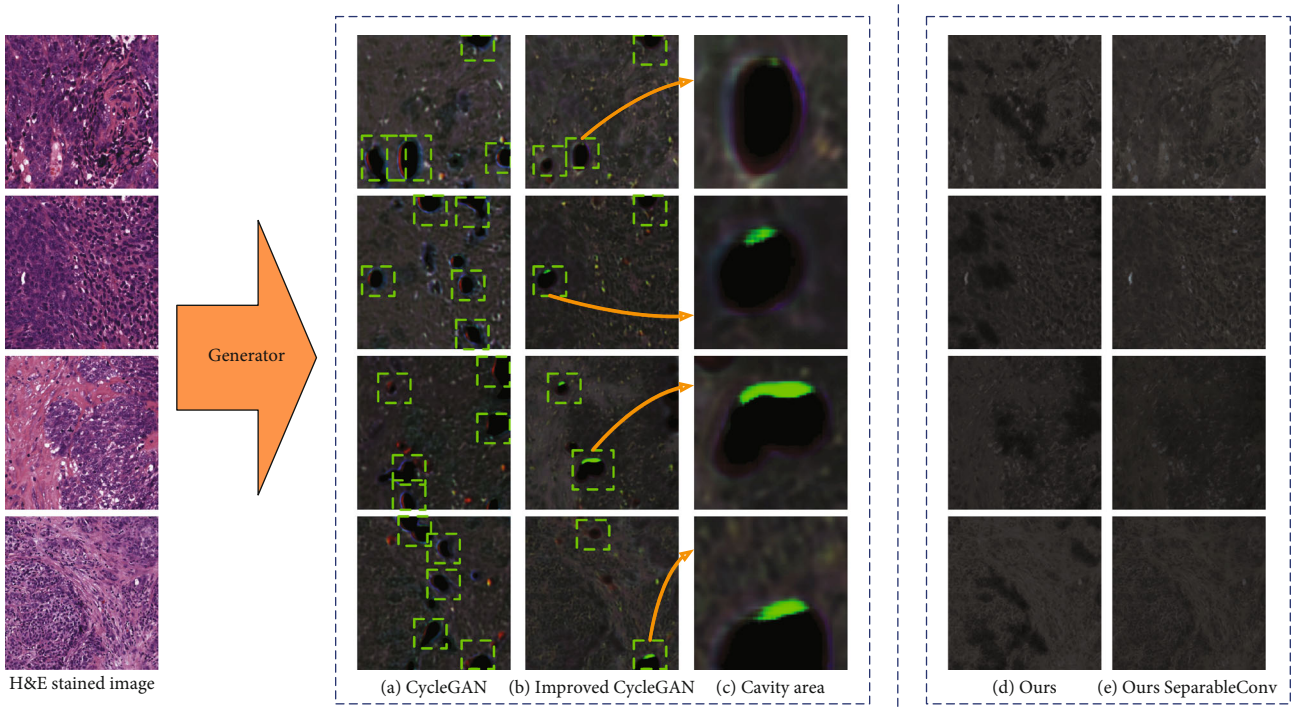


Figure 2: Comparison of the results of using different trained models to construct data sets: (a) results generated by CycleGAN model; (b) results synthesized by the improved CycleGAN model; (c) enlarge some cavity area of the synthesized image; (d) results after introducing domain consistency network training; (e) result of using our modified generator structure and domain consistency network.

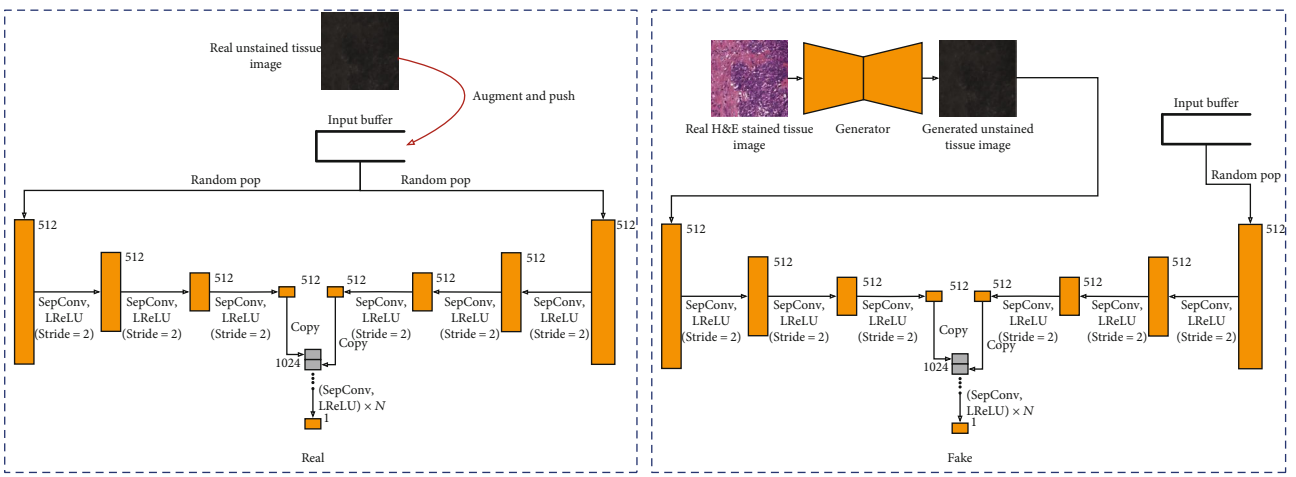


Figure 3: Training the domain consistency network.

compared with the results of the first two models, the result of Figure 2(d) is closer to the real image domain, the image quality has been significantly improved, and the sharp holes have become significantly smoother. Yet it is extremely frustrating that the issue of sharp holes has not been completely solved, and there are still unmatched black areas in the generated image.

2.2.3. Modification of the Generator Network Structure. Let us revisit the reasons for the formation of sharp holes. A similar problem also occurred in the experiment of Karras et al. [32–34]. They observed that most images generated by StyleGAN [33] exhibit characteristic droplet-like artifacts that resemble water droplets. They think it is caused by the AdaIN problem in StyleGAN, and then, they canceled the normalization operation in StyleGAN2 [34], so that droplet-like artifacts can be effectively solved. The generator of the traditional CycleGAN model adopts the network structure proposed by Johnson et al. [35]. The design idea of the network is to perform deep feature fusion of the sampled feature maps by stacking residual blocks. And CycleGAN uses Instance Normalization after each convolution operation. Combined with the hypothesis of Karras et al., we believe that the key issue lies in Instance

Normalization. Instance Normalization is to normalize each layer of feature maps separately, which may ignore the correlation between each layer of feature maps to a certain extent. Meanwhile, due to the structural particularity of the residual block, directly summing the residual and the result after convolution is likely to amplify the effect of Instance Normalization, thus creating this kind of hole.

We made a simple design on the basis of the original generator to completely solve this problem. Since Instance Normalization may affect the correlation between feature maps, we enhance the correlation of the network's feature channels. We first use the Xception block [36] instead of the original residual block. Xception block uses a depth-wise separable convolution, which is mainly composed of depth-wise convolution and point-wise convolution. The benefit of the depth-wise separable convolution is that the convolution's spatial correlation and the feature map channel correlation are operated separately, which reduces the number of training parameters and improves the influence of the convolution on the channel correlation. Next, in order to completely eliminate the problem of Instance Normalization, we replace the convolution operation in the upsampling process and downsampling with depth-wise separable convolution operation. Through the above modifications, we believe that the new generator can completely avoid the cavity problem.

Figure 2(e) is the inference result after training with the improved generator network. We can see that we have completely solved the hole problem, and the image quality was further enhanced. At the same time, through quantitative evaluation, we conclude that this method can construct a good unstained data set. We evaluated the quality of images generated by different methods, and the test results are shown in Table 1. From the evaluation results in the table, we can see that the quality of the images generated by our method far exceeds the state-of-the-art method. At the same time, we also submit the generated data to the doctor to judge, so that the doctor can distinguish the authenticity of the generated image. We have prepared 400 unstained images generated by different methods to allow doctors to judge the images within the specified time. According to the number of correct images, we can get the accuracy. The correct rate of each trial we recorded is shown in Table 2. It can be found that the accuracy of the images obtained by the previous methods is very low. Our analysis is due to the influence of the black holes in the generated images. When there are black holes in the generated image, the doctor will naturally distinguish the difference from the real image and consider the image to be a fake image. This may be why when we completely solve the hole problem, the accuracy of the image is doubled.

Table 1: Comparison of the quality of unstained images using different methods.

	CycleGAN	Improved CycleGAN	Ours	Ours (with separable Conv)
IS ↓	1.590	1.700	1.407	1.311
FID ↓	471.421	360.029	235.410	175.969
MS ↓	0.883	0.873	0.794	0.717

Table 2: Accuracy results of unstained images synthesized using different methods.

	CycleGAN	Improved CycleGAN	Ours	Ours (with separable Conv)
Doctor 1	12.50%	1.25%	24.25%	77.5%
Doctor 2	3.50%	5.50%	39.5%	86.5%
Doctor 3	0.00%	1.50%	55.50%	93.50%

2.3. A Virtual Staining Method Specific to Tissue Texture. Through the above methods, we successfully constructed the paired data set composed of unstained images and stained images. We thus can regard virtual staining as an image-to-image translation problem. Today, many mature algorithms in the field of image-to-image translation have produced amazing results, yet whether these algorithms can be directly applied to this special domain remains to be verified. What we want is an image translation that is accurate to the tissue, so we need to build a more accurate image-to-image translation algorithm. In order to achieve this goal, we have made many modifications to the loss function, network structure, and training strategy.

2.3.1. The Review of Image-to-Image Translation. Conventional image-to-image translation algorithms are usually based on the Pix2Pix baseline. Isola et al. [21] use UNet as the generator and use the patchGAN structure discriminator to discriminate images with accuracy to the patch. And the L1 loss is introduced on the basis of the conventional GAN loss to evaluate the pixel gap between the real image and the generated image. UNet was originally a dedicated network structure designed to handle cell structure segmentation tasks. It can effectively retain a lot of accurate and detailed feature information through layer-by-layer skip connections. At the same time, the patchGAN discriminator reduces the receptive field of the image to be determined, so that the discriminator has a stronger ability to distinguish the details of the image, which also promotes the quality of the generated image. We consider using the idea of Pix2Pix to perfect the model so that the model can better apply to the problems we are facing this time.

2.3.2. A UNet Structure-Based Generator. The essence of the image translation task is that we input an image into the network; the network can learn various features of the image and then convert the original features of the image into the target features. We can simply regard the feature as the information that people can obtain by observing the image, specifically, the information that can perceive after the pixel value of the image is saw by the person. Therefore, the translation process from an image with original features to an image with another type of features can be conceded as the pixel value

conversion of the image. From this perspective, image translation and semantic segmentation tasks are very similar, which is why the generator networks in the earliest image translation tasks (such as Pix2Pix [21]) are designed with the help of UNet networks. The UNet network was first used in the semantic segmentation task of the cell dimension. It uses skip connection to allow the network to effectively learn detailed features. Therefore, our task is also based on the UNet structure to design generator.

The process of using convolution to extract image feature values is a process of continuous dimensionality reduction of image feature data. In this process, the network must selectively learn to extract more representative features. On the contrary, some low-frequency, nonrepresentative features will be ignored in the feature extraction process. When we directly sample the dimensions of features, the network can only use these most representative features. In the process of directly decoding from high-frequency features to the target image, the network will ignore many low-frequency features, which makes the generated image quality very unsatisfactory. The introduction of skip connection is to fuse these low-frequency features with the features of the restoration process. This is why the image translation model using the UNet network as the generator can achieve great results.

So why do not we directly tell the network what characteristics we want to learn? We therefore deconstructed the original UNet network and designed a new generator. We call it the Parallel Feature Fusion Network (PFFN). Its network structure is shown in Figure 4. The stained image we input first enters the Average Pooling Sampling Block in Figure 4(a), and the network will perform the sampling work according to different sampling steps. Through the above operations, we obtain sampled images at different scales. Taking an input image with a size of 256×256 pixels as an example, after sampling with steps 6, 5, 4, 3, 2, and 1, respectively, the different scaled images with 8×8, 16×16, 32×32, 64×64, 128×128, and 256×256 are obtained. Through continuous downsampling operations, the image will lose many low-level aspects features but on the contrary can retain many high-level features. For example, after we sample a cell tissue image, we can see that the high-level aspect features of the image such as the shape of the cell, but we cannot see the low-level features that are lost after sampling, such as the detailed structure within the cell.

After receiving the input images of six scales, we input the images of each scale into the corresponding Parallel Feature Extraction Block in Figure 4(b). Each feature extraction network is designed based on the UNet structure, and the detailed network implementation is shown in Figure 5. The function of the FromRGB module in Figure 5(a) is to convert the image into a feature map of 512-dimensional channels. In the entire network, we stipulate that the feature map is 512 dimensions. We design three branches in FromRGB module, and each branch adopts different sampling methods. Finally, the feature maps of different sensory scales sampled by different sampling methods are deeply fused and used as the input of the improved UNet network in Figure 5(b).

Due to the different sampling scales of input images of different scales, we design three UNet structures as the feature extraction network, named UNet_2, UNet_4, and UNet_6, respectively. The subscripts indicate the feature sampling depth of the UNet. The network structure shown in Figure 5 is UNet_2. What differs from the traditional UNet is that we added the additional skip connections. Unlike the conventional UNet network, we have introduced additional skip connections from the upper layer to the lower layer on the UNet. As we mentioned, every time the network passes through a convolutional layer, some low-frequency features are lost. The original UNet skip connection only guarantees the low-frequency feature transfer to the same layer, but the lower layer may also need the low-frequency feature of the upper layer. We therefore introduced a skip connection from the upper layer to the lower layer so that the bottom layer can also learn effective low-frequency features. This design is very similar to the idea of UNet3+ [37], but we removed the skip connection to the deeper layer. First of all, the number of layers of our three UNet networks is not deep enough. The introduction of so many skip connections may not be significantly improved. On the contrary, it will bring a greater amount of calculation. This new structure diagram of UNet is shown in Figures 5 and 6 where the red arrows indicate the new skip connection we added. We used Unet_6 and UNet as the generator to train the models separately and evaluate the quality of the generated images. It can be seen from Table 3 that the image quality generated by UNet_6 is improved compared with the traditional UNet, where the FID decreased by about 2-3 and the Inception Score decreased by about 0.01-0.04. However, the network depth of UNet_6 is only 6 layers, and the depth of UNet reaches 8 layers. It can be proved that UNet_6 has a powerful feature learning ability within a limited sampling field.

We specify that 8×8 and 16×16 input images use UNet_2 network, 32×32 and 64×64 input images use UNet_4 network, and 128×128 and 256×256 input images use UNet_6 network. In the process of continuous sampling of images with 256×256 pixels to 8×8, the lower resolution image retains the higher-level aspect features, which can be learned by using the shallow network structure like UNet_2. As the image pixels increase, the UNet network structure continues to deepen, and the effect of UNet will continue to be highlighted. The low-level aspect features of high-resolution images can be learned through the deeper network like UNet_4 and UNet_6. In this way, the network can learn the image features of each level of the sampling module according to our wishes, and the resulting feature maps cover the feature values of the image from coarse to fine. We finally introduce an upsample process to continuously fuse these features to obtain the final output image. In the process of upsample, we also introduce skip connection to ensure the feature of high-level aspects to propagate down better. Compared with the traditional UNet network as the generator, the network designed by us has a wider reception field, and the image obtained has a stronger performance ability, while covering all the characteristics of UNet. From Table 3, compared with the image quality generated by UNet, the FID of the image generated by using our PFFN network as a generator was reduced by about 3-9, and the Inception Score was reduced by about 0.1-0.2. Compared with the image FID

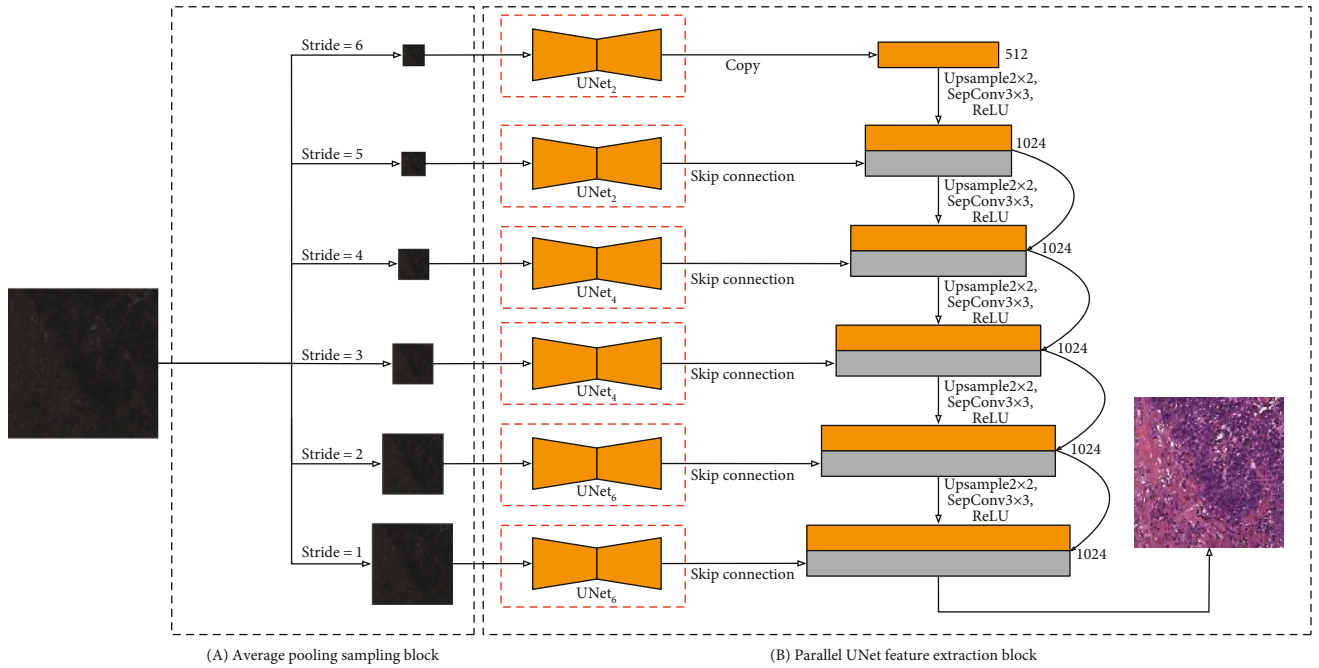


FIGURE 4: The structure of Parallel Feature Fusion Network.

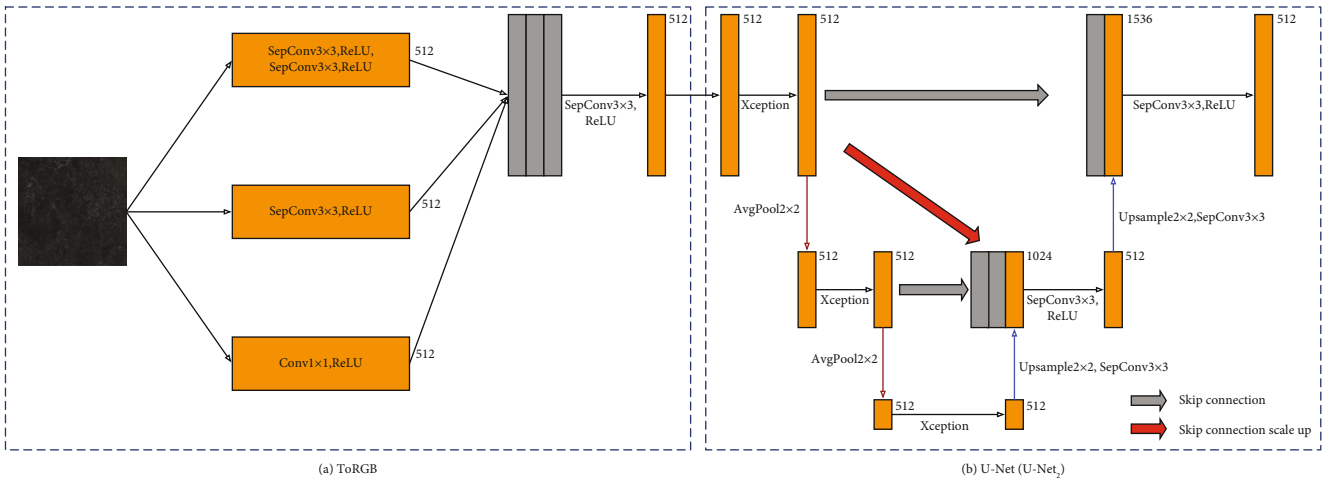


FIGURE 5: Network structure of feature extraction module ($UNet_2$). (a) FromRGB module. This module extends low-dimensional images to higher dimensional feature maps. (b) Feature extraction module based on $UNet_2$.

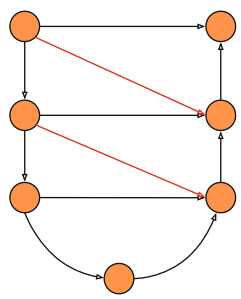


FIGURE 6: The UNet structure introducing additional skip connections.

generated by $UNet_6$, the FID was reduced by about 1-8, and the Inception Score was reduced by about 0.1-0.2.

2.3.3. Choice of Loss Function. The most commonly used loss functions of traditional GANs are cross-entropy loss and L_2 loss. These loss functions have proven their feasibility in GAN training through a large number of experiments. But there are also many studies show that using these two loss function optimization models in GAN's training will cause very terrible results. Moreover, many drawbacks of GAN, such as mode collapse and training instability, are caused by the use of these instable loss functions. We used this kind of loss function to train in the experiment, yet the result is not

Table 3: Comparison of image quality using different loss functions and generator network structures.

Network	Loss function	FID ↓	IS ↓
UNet	L_{l_2}	57.6092	1.3405
	L_{wgan}	54.1733	1.4687
	L_{logistic}	56.1733	1.4687
	$L_{l_2} + L_{\text{fm}}$	59.4684	1.4720
	$L_{\text{wgan}} + L_{\text{fm}}$	58.1196	1.4720
	$L_{\text{logistic}} + L_{\text{fm}}$	56.1639	1.4432
UNet_6	L_{l_2}	54.1436	1.3928
	L_{wgan}	50.8299	1.4256
	L_{logistic}	52.4708	1.3865
	$L_{l_2} + L_{\text{fm}}$	51.3790	1.3907
	$L_{\text{wgan}} + L_{\text{fm}}$	55.2754	1.4852
	$L_{\text{logistic}} + L_{\text{fm}}$	49.3387	1.4073
PFFN (ours)	L_{l_2}	54.8384	1.3835
	L_{wgan}	49.1167	1.3903
	L_{logistic}	49.6818	1.4124
	$L_{l_2} + L_{\text{fm}}$	49.3575	1.3238
	$L_{\text{wgan}} + L_{\text{fm}}$	47.0977	1.3505
	$L_{\text{logistic}} + L_{\text{fm}}$	48.8730	1.2158

what we expected. Therefore, choosing an appropriate loss function is very important in this experiment.

In this experiment, we consider using two robust loss functions: Wasserstein loss with gradient penalty [18, 19] and Logistic loss [33] with R1 regularization [38, 39] as the training loss function. Wasserstein distance is simple and direct compared to the original loss function and highly correlates with the quality of the synthesized image of the generator. Using Wasserstein loss may be a good choice. Logistic loss is applied in the StyleGAN paper by Karras et al. [33]. He used this loss as the adversarial loss and generated the high-resolution face images. Both of these two loss functions have very good performance. In order to verify which loss function can be better applied to our virtual staining experiment, we use these functions to train the model and evaluate the generator by the quality of the image. As shown the results in Table 3, we find that the image quality synthesized by the generator trained with the Wasserstein loss (Equation (2)) is better.

$$L_{\text{gan}}(G, D, x, y) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[D(x, G(x))] - \mathbb{E}_{y\sim p_{\text{data}}(y), x\sim p_{\text{data}}(x)} \cdot [D(x, y)] - \lambda_{gp}\mathbb{E}_{\hat{x}\sim p_{\text{data}}(\hat{x})}\left[\left\|\left(\nabla_{\hat{x}} D(\hat{x})\right\|_{l2}^2 - 1\right)^2\right]. \tag{2}$$

The traditional image translation model usually introduce an additional L1 loss (Equation (3)) based on adversarial loss; its function is to narrow the global gap between the real image and the generated image. But the introduction of this loss will make the image blurred and attenuate the quality of the generated image. To get a higher quality image, we need a sturdy loss function to act on the generator. Wang et al. [22] used the feature matching loss (Equation (4)) in the high-resolution image translation task of the Cityscapes data set and got flawless results. We introduced this loss into the training of the generator and found that the effect has been significantly improved through the final evaluation results. According to the results of our experiment, we choose Wasserstein loss in the adversarial loss, as shown in the Equation (2), where $\mathbb{E}_{\hat{x}\sim p_{\text{data}}(\hat{x})}[\|(\nabla_{\hat{x}} D(\hat{x})\|_{l2}^2 - 1)^2]$ represents the gradient penalty for the $\hat{x}$, which is the random interpolating of the positive sample and the generated sample. In order to reduce the global difference between the generated image and the real image, we retain the L1 loss as shown in Equation (3). In order to further improve the quality of the generated image, we increase the feature matching loss, as shown in the Equation (4), where N represents the number of layers of the discriminator.

$$L_{\text{pix}}(G, x, y) = \mathbb{E}_{x\sim p_{\text{data}}(x), y\sim p_{\text{data}}(y)}\left[\|y - G(x)\|_{l1}\right], \tag{3}$$

$$L_{\text{fm}}(G, D, x, y) = \mathbb{E}_{x\sim p_{\text{data}}(x), y\sim p_{\text{data}}(y)} \cdot \left[\frac{1}{N}\sum_{i=1}^{N}\left(\|D(x, y) - D(x, G(x))\|_{l1}\right)\right]. \tag{4}$$

The model loss we finally get is expressed by the following Equation (5). We use the gradient descent method to solve the following equation: $G_{\text{opt}}, D_{\text{opt}} = \arg\ \min_G \max_D L(G, D, x, y)$. Finally, we can get the optimal solution of the staining model.

$$L(G, D, x, y) = L_{\text{gan}}(G, D, x, y) + \lambda_{\text{pix}} L_{\text{pix}}(G, x, y) + \lambda_{\text{fm}} L_{\text{fm}}(G, D, x, y). \tag{5}$$

3. Results and Discussion

Through the above training, we have successfully generated H&E staining of ovarian cancer pathological slices. Figure 7 is the result comparison between the generated virtual stained image and the real H&E staining image. Intuitively, the gap between the real and the fake is quite hard to distinguish, but for this method to be better used in medical products, we need to compose a series of evaluations on these virtual stained images we generate.

3.1. Artificial Pathology Analysis. A successful virtual staining section can express the correct pathological characteristics; otherwise, it will seriously affect the doctor's pathological diagnosis. To evaluate the generated pathological slices more subjectively, we invited three professional doctors to analyze the difference between the image generated by our model and the real image and evaluate whether the staining for these images is successful. After evaluating each image, we can roughly get the staining accuracy of our model. Since it is a

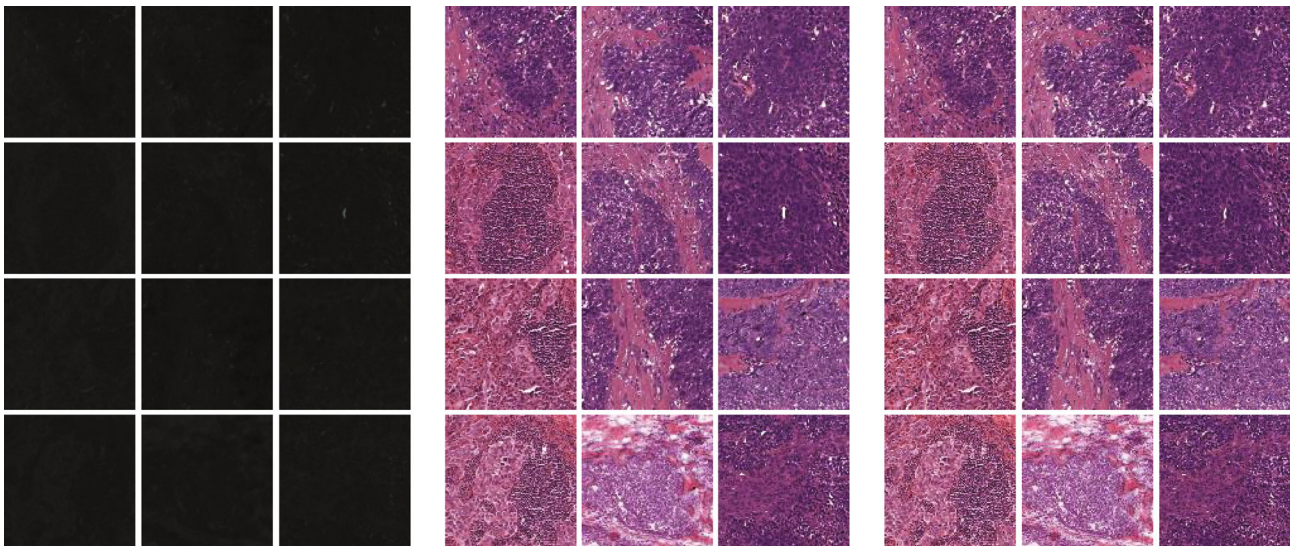


Figure 7: Virtual staining result display on pathological sections of ovarian cancer.

Table 4: Staining accuracy of our model analyzed by three doctors.

	Doctor 1	Doctor 2	Doctor 3
Samples with successful staining	190	196	194
Accuracy	95%	98%	97%

heavy task to evaluate a large number of images, we only randomly selected 200 images as the evaluation data, and the final results obtained are shown in Table 4. According to the doctor's evaluation results in Table 4, the staining accuracy of our method reaches 97%, which proves that our method has achieved a perfect staining effect. However, our verification case is only 200 cases, which cannot well represent the overall effect. We need to add evaluation cases in the next period of treatment.

3.2. Visual Simulation Analysis. It was mentioned in the previous section that direct subjective analysis by doctors can certainly get a good score result, but it will consume a lot of time and work, and the results of the evaluation using a small number of samples do not have a good overall representativeness. In particular, our model can generate a large number of H&E staining models in a short time, which is very unrealistic for doctors to perform analysis and evaluation.

To overcome this problem, we propose a method based on deep learning to simulate the visual analysis of doctors. The H&E staining sections of real ovarian cancer we selected have detailed pathological analysis results, and each section doctor clearly marked a cancer lesion and tumor type. We can train a classification network based on the data set composed of real images based on the above annotations. We can think that the trained classification network has the doctor's focus classification ability. Since the generated pathological stained slices should have the same pathological characteristics as the corresponding real stained slices, we use the pretrained classification network to make inference and prediction tumor type results for the corresponding generated virtual stained images. Finally, we calculate the difference between the result of the generated image and the real result to get the final accuracy. This accuracy can approximately represent the quality of the lesion features based on the image generated by our model.

First of all, each stained image in our data set is annotated by professional doctors according to the four types of tumors. Next, we use these labeled data sets to train a VGG16 classification network. The classification accuracy of the trained network reached about 97%. We can think that this VGG network has a strong ability to distinguish ovarian cancer tumor types. Then, we use the trained VGG network to predict the stained ovarian cancer sections we generated and calculate the accuracy. If the prediction of the generated image is correct, it can indicate that the virtual stained slice we synthesized expresses the correct feature of the lesion. We think it is reasonable to apply such images to pathological analysis. Finally, the accuracy of the proposed method reached 95%. We can conclude the final difference of virtual staining to be 2%. It can be proved that our method has reached the standard of pathological analysis of ovarian cancer.

4. Conclusions

We provide a more efficient solution for H&E staining of ovarian cancer pathological sections. Using our method can be very effective to save time and quickly assist the doctor in diagnosis. We have used many evaluation methods. From the results, the quality of the stained image generated by our method is very perfect. At the same time, we have also proposed an effective autofluorescence image generation algorithm in the absence of valid data, which can save time-consuming and laborious data preparation time in many cases. In the next research, we will carry out research on

virtual staining of more pathological tissues in order to realize a more extensive virtual staining technology.

Data Availability

The data we use is mainly composed of the TCGA ovarian cancer database and the clinical data. TCGA ovarian cancer data can be obtained from https://portal.gdc.cancer.gov/. Considering the privacy of patients, we cannot open access to our clinical data.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 61972416, 61873280, and 61873281) and the Natural Science Foundation of Shandong Province (No. ZR2019MF012).

References

[1] X. Li, X. Wang, H. Li, X. Shi, and P. Zheng, "A programming 20–30nm rectangular DNA origami for loading doxorubicin to penetrate ovarian cancer cells," *IEEE Transactions on Nanobioscience*, vol. 19, no. 1, pp. 152–157, 2020.

[2] T. Song, P. Zheng, M. L. Dennis Wong, and X. Wang, "Design of logic gates using spiking neural P systems with homogeneous neurons and astrocytes-like control," *Information Sciences*, vol. 372, pp. 380–391, 2016.

[3] Z. Li, L. Zhang, Y. Su, J. Li, and X. Wang, "A skin membrane-driven membrane algorithm for many-objective optimization," *Neural Computing and Applications*, vol. 30, no. 1, pp. 141–152, 2018.

[4] X. Wang, T. Song, F. Gong, and P. Zheng, "On the computational power of spiking neural P systems with self- organization," *Scientific Reports*, vol. 6, no. 1, pp. 1–16, 2016.

[5] X. Wang, P. Zheng, T. Ma, and T. Song, "Small universal bacteria and plasmid computing systems," *Molecules*, vol. 23, no. 6, article 1307, 2018.

[6] S. Pang, Y. Zhang, M. Ding, X. Wang, and X. Xie, "A deep model for lung cancer type identification by densely connected convolutional networks and adaptive boosting," *IEEE Access*, vol. 8, pp. 4799–4805, 2020.

[7] X. Wang, D. Liu, J. Zhu, A. Rodriguez-Paton, and T. Song, "CSConv2d: a 2-D structural convolution neural network with a channel and spatial attention mechanism for protein-ligand binding affinity prediction," *Biomolecules*, vol. 11, no. 5, p. 643, 2021.

[8] X. Wang, Y. Zhong, and M. Ding, "Repositioning drugs to the mitochondrial fusion protein 2 by three-tunnel deep neural network for Alzheimer's disease," *Frontiers in Genetics*, vol. 12, 2021.

[9] S. Wang, L. Dong, X. Wang, and X. Wang, "Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy," *Open Medicine*, vol. 15, no. 1, pp. 190–197, 2020.

[10] X. Wang, B. Liu, Y. Dong, S. Pang, and X. Tao, "Anthropometric landmarks extraction and dimensions measurement based on ResNet," *Symmetry*, vol. 12, no. 12, article 1997, 2020.

[11] U. A. Matulonis, A. K. Sood, L. Fallowfield, B. E. Howitt, J. Sehouli, and B. Y. Karlan, "Ovarian cancer," *Nature Reviews Disease Primers*, vol. 2, no. 1, pp. 1–22, 2016.

[12] M. Salvi, N. Michielli, and F. Molinari, "Stain Color Adaptive Normalization (SCAN) algorithm: separation and standardization of histological stains in digital pathology," *Computer Methods and Programs in Biomedicine*, vol. 193, article 105506, 2020.

[13] M. R. Castellanos, V. M. Nehru, E. C. Pirog, and L. Optiz, "Fluorescence microscopy of H&E stained cervical biopsies to assist the diagnosis and grading of CIN," *Pathology-Research and Practice*, vol. 214, no. 5, pp. 605–611, 2018.

[14] Y. Rivenson, H. Wang, Z. Wei et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nature Biomedical Engineering*, vol. 3, no. 6, pp. 466–477, 2019.

[15] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2223–2232, Venice, Italy, October 2017.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., pp. 234–241, Springer, Cham, 2015.

[17] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," 2014, http://arxiv.org/abs/1406.2661.

[18] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, http://arxiv.org/abs/1701.07875.

[19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," 2017, http://arxiv.org/abs/1704.00028.

[20] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, http://arxiv.org/abs/1511.06434.

[21] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1125–1134, Honolulu, HI, USA, July 2017.

[22] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8798–8807, Salt Lake City, UT, USA, June 2018.

[23] X. Wang, K. Yu, S. Wu et al., "Esrgan: enhanced super-resolution generative adversarial networks," in *Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science, vol 11133*, L. Leal-Taixé and S. Roth, Eds., Springer, Cham, 2018.

[24] C. Ledig, L. Theis, F. Huszár et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690, Honolulu, HI, USA, July 2017.

[25] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, http://arxiv.org/abs/1411.1784.

[26] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: unified generative adversarial networks for multi-domain image-to-image translation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797, Salt Lake City, UT, USA, June 2018.

[27] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning, PMLR*, pp. 1857–1865, Sydney, Australia, 2017.

[28] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras, "Neural face editing with intrinsic image disentangling," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5541–5550, Honolulu, HI, USA, July 2017.

[29] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," 2017, http://arxiv.org/abs/1701.04862.

[30] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1511–1520, Venice, Italy, October 2017.

[31] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu et al., "Video-to-video synthesis," 2018, http://arxiv.org/abs/1808.06601.

[32] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," 2017, http://arxiv.org/abs/1710.10196.

[33] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, Long Beach, CA, USA, June 2019.

[34] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8110–8119, Seattle, WA, USA, June 2020.

[35] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9906*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., pp. 694–711, Springer, Cham, 2016.

[36] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1251–1258, Honolulu, HI, USA, July 2017.

[37] H. Huang, L. Lin, R. Tong et al., "Unet 3+: a full-scale connected unet for medical image segmentation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059, Barcelona, Spain, May 2020.

[38] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?," in *Proceedings of the 35th International Conference on Machine Learning, PMLR*, pp. 3481–3490, Stockholm, Sweden, 2018.

[39] A. Ross and F. Doshi-Velez, "Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Hilton New Orleans Riverside, New Orleans, 2018.