

Research Article

Classification on Digital Pathological Images of Breast Cancer Based on Deep Features of Different Levels

Xin Li , HongBo Li , WenSheng Cui, ZhaoHui Cai, and MeiJuan Jia

School of Computer Science and Information Technology, Daqing Normal University, Daqing 163712, China

Correspondence should be addressed to HongBo Li; lhb20210401@163.com

Received 22 November 2021; Accepted 15 December 2021; Published 30 December 2021

Academic Editor: Xiangtao Li

Copyright © 2021 Xin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Breast cancer is one of the primary causes of cancer death in the world and has a great impact on women's health. Generally, the majority of classification methods rely on the high-level feature. However, different levels of features may not be positively correlated for the final results of classification. Inspired by the recent widespread use of deep learning, this study proposes a novel method for classifying benign cancer and malignant breast cancer based on deep features. First, we design Sliding + Random and Sliding + Class Balance Random window slicing strategies for data preprocessing. The two strategies enhance the generalization of model and improve classification performance on minority classes. Second, feature extraction is based on the AlexNet model. We also discuss the influence of intermediate- and high-level features on classification results. Third, different levels of features are input into different machine-learning models for classification, and then, the best combination is chosen. The experimental results show that the data preprocessing of the Sliding + Class Balance Random window slicing strategy produces decent effectiveness on the BreakHis dataset. The classification accuracy ranges from 83.57% to 88.69% at different magnifications. On this basis, combining intermediate- and high-level features with SVM has the best classification effect. The classification accuracy ranges from 85.30% to 88.76% at different magnifications. Compared with the latest results of F. A. Spanhol's team who provide BreakHis data, the presented method shows better classification performance on image-level accuracy. We believe that the proposed method has promising good practical value and research significance.

1. Introduction

In recent years, the global prevalence of breast cancer (BC) has been gradually increased, and the affected organisms tend to be younger, gender-neutral, and have racial ambiguity, which has posed a huge threat to human beings' normal life. In 2018, the World Health Organization's International Agency estimated that there were 2.1 million new female cases of BC among women about 25 percent of all cancers. The number of female cases was far greater than any other cancers in both developed and developing countries [1]. BC is the leading cause of cancer deaths among women between 20 and 60 years old. Early diagnosis and treatment can effectively reduce the risk of diseases and prevent the progression of cancers [2]. The traditional diagnosis of BC includes breast mammography, breast B ultrasound, dynamic enhanced magnetic resonance, and pathological biopsy [3]. Normally, pathologists need to combine the

feedback from medical equipment with their own diagnostic experience to test and analyze sample information for cancer diagnosis and treatment strategy. The diagnosis process is inefficient, costly, and subjective. Furthermore, due to the uneven distribution of pathologists and medical resources around the world, it is difficult to ensure timely and effective treatment for patients in remote areas and underdeveloped countries [4]. Therefore, an efficient, low-cost, and objective diagnosis method has important social significance and research value.

Digital pathological slicing scanner converts the pathological slicing of substance from photoelectric signal to digital signal and finally generates full-information and high-resolution digital image, namely whole slide imaging (WSI) [5]. Compared with traditional pathological slicing, it has the advantages of convenient preservation, less damage, remote diagnosis, and so on. WSI provides a basis for automatic classification and quantitative analysis of medical

pathological slicing [6, 7]. In pathological WSI, different diseases have different spatial pixel arrangements and cellular features. Cells of different shapes, sizes, and colors may also be of the same type. Therefore, automatic analysis of WSI by computer algorithm is a very challenging job. With the upgrade of software and hardware, machine-learning (ML) algorithm and deep learning (DL) algorithm have been widely used in the field of lesion region segmentation, location detection, and subtype classification as a digital pathology-aided diagnosis method. They also become one of the hot research fields in pattern recognition and artificial intelligence [8, 9]. In the automatic analysis of pathological images, the application of computer-aided diagnosis (CAD) technology has improved the accuracy and efficiency of disease diagnosis and grasps the development trend of disease more objectively. It also reduces the limitations of economic conditions, geographical environment, and medical foundation.

There are two types of BC: benign and malignant. The benign cancer is prone to be transformed into the malignant cancer in the early stage. The benign cancer mainly includes adenosis, phyllodes tumor, fibroadenoma, and tubular adenoma. The malignant cancer mainly includes papillary carcinoma, ductal carcinoma, lobular cancer, and mucinous cancer [10, 11]. Among them, ductal carcinoma and lobular cancer accounted for more than 95%.

In the traditional classification of the BC, doctors need to combine feedback from medical equipment with their own experience to diagnose the body, which is somewhat subjective and inefficient. In addition, with the increasing prevalence of BC and the lack of medical resources, problems such as missed diagnosis and misdiagnosis are prone to occur. Some ML algorithms can solve the above problems, but the model performance is not satisfactory and cannot provide scientific guidance and suggestions for clinical treatment. Therefore, researchers have conducted in-depth exploration and research on the CAD method based on the DL algorithm. Although great progress has been made, there are still some shortcomings in data preprocessing, feature extraction, unbalanced data analysis, and the selection of benchmark dataset. To solve the above problems, combining the correlation and difference between the ML and DL algorithms, we propose a novel BC's benign and malignant classification scheme based on deep features of different levels. Firstly, image preprocessing is performed on the BC's pathological tissue images, including normalization and window slicing. Secondly, the input data images are trained based on the AlexNet model. Then, we discuss how to train the network model and choose appropriate nodes by extracting features. Finally, the trained network model extracts features by choosing node. Then, the extracted features are input into the ML model to classify. The experiment is performed using a fivefold cross-validation approach, the same folds released with the BreaKHis dataset. The experimental results show that the average classification accuracy and standard deviation of pathological tissue images at different magnifications (40 \times , 100 \times , 200 \times , and 400 \times) are 87.85%, 86.68%, 87.75%, and 85.30%, respectively. The average classification accuracy and standard deviation of patients at different

magnifications (40 \times , 100 \times , 200 \times , and 400 \times) are 87.93%, 87.41%, 88.76%, and 85.55%, respectively. In summary, our approach has higher classification accuracy on the image-level accuracy and model stability than the latest classification results of Spanhol team from the BreaKHis image dataset in literature [12]. The following is a summary of the contribution of this article:

- (1) To improve the generalization of the model, four data preprocessing strategies are used in this study. The four window slicing strategies are Sliding, Random, Sliding + Random, and Sliding + Class Balance Random. The experimental results show that the preprocessing strategy of Sliding + Class Balance Random window slicing has the best classification results.
- (2) As a second step, the intermediate- and high-level features are extracted by the AlexNet model. The experiments show that the intermediate- and high-level features have a better classification effect.
- (3) Based on the previous steps, extracted features are input into a variety of ML models for training and evaluation. We classify BC pathological tissue images (BreaKHis), and SVM has great potential for classifying.

2. Related Work

The histopathological analysis is a highly specialized task. The effectiveness of diagnosis depends on pathologists' experience, attention, and fatigue. Traditional techniques such as MRI, ultrasound, and biopsy technology are used to detect and grade BC's lesions. Although biopsy techniques are time-consuming to diagnose, it is still one of the gold standards on diagnosis [13]. Common biopsy techniques include skin biopsy, fine-needle biopsy, core biopsy, and surgical biopsy. In the biopsy assay, biopsy tissue slicing samples are first obtained, and then, hematoxylin and eosin (H&E) staining is performed. At last, the pathologists use the equipment to analyze texture, morphology, and histological characteristics of the biopsy tissue to give the corresponding diagnostic results [14]. The pathologists focus, magnify, and scan the entire tissue under a high-power microscope. The procedure is time-consuming, repetitive, and subjective, so the diagnosis result may be very different.

With the advent of computer era, how to use computer algorithms to better assist pathologists in the simple and repeatable process on the pathologic diagnosis has always been a challenging task [15]. CAD allows pathologists to devote more energy and time on dealing with difficult disease. Therefore, a great deal of intensive research emerges from the CAD field, especially for the analysis of BC. Khamparia et al. [16] use a modified VGG model as a pretraining model on the DDSM dataset of X-ray, and the classification accuracy is 94.3%. Sharma et al. [17] use various evolutionary algorithms as the feature extractor for the Wisconsin breast cancer dataset on the UCI database. Then, they use an ML model as a classifier and get a maximum accuracy of 96.45% by the combination of BPSO

and SVM algorithms. Ha et al. [18] classify benign and malignant MRI images of 216 BC patients, which are provided by Columbia University Medical Center. In the study, they present a convolutional neural network (CNN) based on MRI features and finally achieve 70% accuracy. Cruz-Roa et al. [19] study the WSI tissue region in 162 patients with invasive ductal breast, which is provided by Pennsylvania University and New Jersey Cancer Institute. At first, the team transforms the WSI detection problem into a classification problem, so they reduce the difficulty on detection of disease areas. Then, a multilayer CNN is designed to classify the tissue images after slicing. Finally, the evaluation results achieve F1 score of 71.80% and a balanced accuracy of 84.23%. Fuzzy color histogram is manually extracted features and is input into random forest classifier for training. The results achieve 67.53% for F1 score and 78.74% for balanced accuracy. In the same way, the results of RGB histogram achieve 66.64% for F1 score and 77.24% for balanced accuracy. Anuranjeeta et al. [20] use morphological features to detect BC lesions and classify cells in pathological tissue images. Finally, the results achieve 85.7% accuracy and 0.884 AUC value on rotating forest classification model.

Spanhol et al. propose a series of studies on benign and malignant BC classification. The experiment is performed using a fivefold cross-validation approach, and the specific research contents are as follows:

- (1) At the beginning of the series, Spanhol et al. [12] provide 7909 BC digital pathological tissue images from 82 patient cases (BreKHis). They also design six different feature extractors to extract low-level features such as texture and edge. After then, ML algorithms such as decision tree (DT), random forest (RF), and support vector machines (SVMs) are used to classify for extracting features. Finally, a better classification result is the combination of the feature extractors PFTAS [21] and SVMs. The average accuracy of classification at different magnifications (40 \times , 100 \times , 200 \times , and 400 \times) ranges from 80% to 85%, and the standard deviation is about 5%.
- (2) To further improve the classification performance, Spanhol et al. [22] use a novel strategy to obtain small pieces from the BreKHis image data. After obtaining the data, these small pieces are input into CNN model for training. Image-level accuracy and patient-level accuracy are given as classification evaluation indicators. The final classification is decided by voting mechanism of ensemble learning. At last, on the patient-level accuracy and image-level accuracy, the average accuracy of classification ranges from 80% to 90%, with a large standard deviation.
- (3) In [23], Spanhol uses the AlexNet model for the pretraining on BreKHis dataset and extracting deep features. The features in different depths of layers are obtained and all are input into the logistic regression (LR) classifier to train and evaluate. Finally, the decaffeinated coffee method named by Spanhol et al. is combined with the method in literature [12]. The experiments show that the average accuracy at 200 \times magnifications achieves 86.3% on the patient-level accuracy and 84.2% on the image-level accuracy. However, compared with literature [12], the evaluation results of other magnifications significantly decreased.
- (4) Recently, Spanhol et al. [24] apply the multi-instance principle to classify BC pathological cancer images. On the patient-level accuracy, the average accuracy is 92.1%, 89.1%, 87.2%, and 82.7%, respectively. Similarly, on image-level accuracy, the average accuracy at different magnifications (40 \times , 100 \times , 200 \times , and 400 \times) achieves 87.8%, 85.6%, 80.8%, and 82.9%, respectively. Compared with literature [23], the image-level accuracy decreases at 200 \times and 400 \times magnifications, while the standard deviation increases.

Though researchers have made great progress in the image classification problem of BC, there are still many deficiencies. The studies from the literature [16–20] use a small BC dataset. As a result, it leads to the inability for fitting model and objectively evaluating generalization ability. In [12], it is hard to obtain the effective classification features using the traditional manual feature extraction method. In [22], researchers adopted the random data acquisition method. This method leads to model instability. Because 1000 data pieces (64*64*3) are obtained from each data image (350*230*3), the model training is difficult and the data are redundant. It is difficult to provide a large enough global view of the data.

3. Materials and Methods

3.1. Dataset. In our work, we use the publicly available BreKHis image dataset as the benchmark for classifying BC. The dataset contains 7909 pathological tissue images of benign and malignant BC in 82 groups of patients. The patients were invited to participate in research activities in the P&D Laboratory in Brazil during January to December in 2014. All information is anonymous to ensure patient privacy. Pathologists use H&E to stain and label the pathological tissue images. Samples are prepared using a standard paraffin method, including fixation, dehydration, removal, and correction. In the process of the digital transformation of BC tissue images, a Samsung digital color camera SCC-131AN is used to obtain digital tissue images of RGB channels with the magnifications of 40 \times , 100 \times , 200 \times , and 400 \times . Pathologists remove the worthless areas such as the black border and text annotation from the original image by clipping operation. Finally, the digital pathological tissue image of BC with 700 \times 460 \times 3 pixels is obtained. Sample tissue images of magnifications (40 \times , 100 \times , 200 \times , and 400 \times) are shown in Figure 1.

BreKHis dataset is divided into benign and malignant BC. Table 1 shows the distribution of them. Besides, the benign cancer and malignant cancer have different subtypes

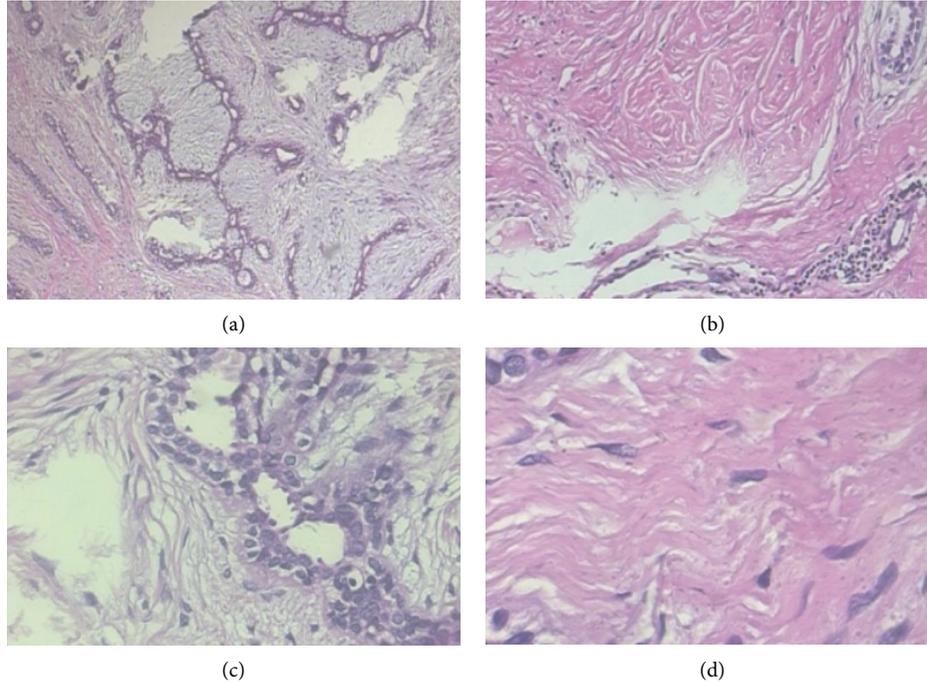


FIGURE 1: Digital pathological tissue images of breast cancer at different magnifications. (a) 40×, (b) 100×, (c) 200×, and (d) 400×.

TABLE 1: Distribution of benign cancer and malignant cancer.

Magnifications	Benign	Malignant	Total
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total	2480	5429	7909
Patients	24	58	82

under a high-power microscope. The benign cancer includes four types: adenosis (A), fibroadenoma (F), tubular adenoma (TA), and phyllodes tumor (PT); the malignant cancer includes four types: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). The class distribution of benign and malignant subtypes is shown in Tables 2 and 3.

3.2. Methods. In our work, the normalization and window slicing operations are firstly used to preprocess the dataset. We design the training model based on the AlexNet model. The features of nodes in different depths are extracted by the model. The extracted features are input into the ML classifier for classification. Finally, we choose the best combination of feature extraction node and ML classifier. Four window slicing strategies (Sliding, Random, Sliding + Random, and Sliding + Class Balance Random) are used to obtain small pieces of data. Next, small pieces are input into the model for training and evaluation. The optimal window slicing strategy is proved by experimental results. At last, the best strategy is used as the basis of subsequent research. The method structure diagram is shown in Figure 2.

TABLE 2: Distribution of subtypes of benign cancer.

Magnifications	A	F	TA	PT	Total
40×	114	253	109	149	625
100×	113	260	121	150	644
200×	111	264	108	140	623
400×	106	237	115	130	588
Total	444	1014	453	569	2480
Patients	4	10	3	7	24

TABLE 3: Distribution of subtypes of malignant cancer.

Magnifications	DC	LC	MC	PC	Total
40×	864	156	205	145	1370
100×	903	170	222	142	1437
200×	896	163	196	135	1390
400×	788	137	169	138	1232
Total	3451	626	792	560	5429
Patients	38	5	9	6	58

When the trained model is used to predict the whole image, the data are split into 15 small image sets with $128 \times 128 \times 3$ pixel size by 128 pixel step length. The small image set is input into the model for prediction, and 15 decision results are obtained for each whole pathological image. The classification with the most votes is chosen by voting as the final prediction result.

3.2.1. Data Preprocessing. In this study, the normalization method is adopted to preprocess data, which can compress the data range from -1 and 1. This data processing method is not only simple and easy to implement, but also can improve the training efficiency and generalization of the model. The mathematical calculation formula is shown in the following equation:

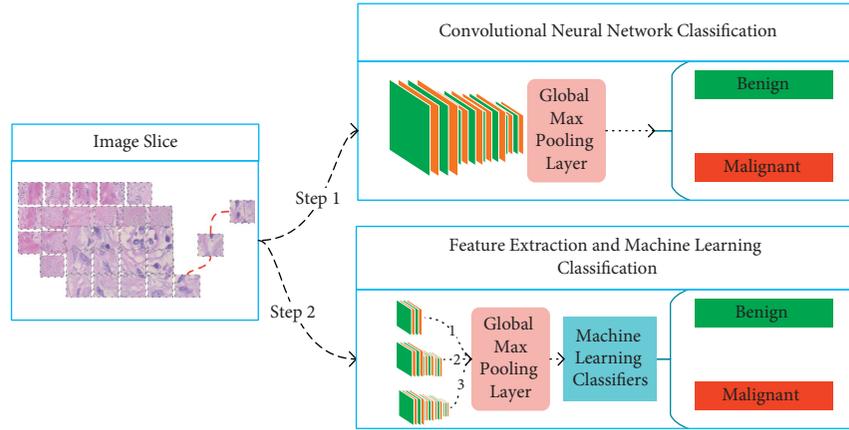


FIGURE 2: Method structure. Step1 is that the slicing images are input into CNN for benign-malignant classification. Step 2 obtains three different models for feature extraction according to the trained model of Step1 and combines the extracted features with the machine-learning model to classify benign cancer and malignant cancer.

$$X' = \left(\frac{X}{255} - 0.5 \right) * 2. \quad (1)$$

Because there is no area of interest in the BreakHis dataset, the experiments can be designed according to their own needs during the preprocessing by window slicing. In [22], the original data image is scaled from $700 \times 460 \times 3$ to $350 \times 230 \times 3$ in [25]. Four strategies are used to obtain small images. The specific division is shown in Table 4.

In Table 4, 32×32 and 64×64 pixel windows are used to obtain data images, and the sliding step is half of the corresponding window size, but it is hard to obtain a large global view of the data. When acquiring 1000 pieces of data by Random window slicing strategy, there is a lot of repeated information. Obviously, this preprocessing method can hardly provide effective data. It leads to underfitting and overfitting problems when the model is trained. To overcome this problem, we try four different window slicing strategies, including Sliding window slicing, Random window slicing, Sliding + Random window slicing, and Sliding + Class Balance Random window slicing. The experimental results show that Sliding window slicing strategy can guarantee model’s fitting training, but cannot guarantee generalization training. Oppositely, Random window slicing can guarantee model’s generalization training, but cannot guarantee the fitting training effectively. In our work, we have fully considered the advantages of the two strategies and proposed the Sliding + Random window slicing strategy. This idea not only guaranteed the model’s fitting training, but also improved the generalization ability.

Due to the unbalanced distribution of classes, we improve the Random window slicing strategy by obtaining random minority class data with directivity. The specific window slicing strategy of the data is shown in Table 5. The size of various window slicing strategies is $128 \times 128 \times 3$ pixels, and the sliding step is 128 pixels. This mode can ensure a large enough global view of data. Then, we use four data strategies to generate data for the model training. Sliding + Random window obtains 45 small pictures and can

TABLE 4: Data window slicing method of [22].

#	Patch size	Strategy	Number of patches
1	32×32	Sliding window	260
2	64×64	Sliding window	54
3	32×32	Random window	1000
4	64×64	Random window	1000

TABLE 5: Data window slicing mode.

#	Strategy	Number of patches
1	Sliding window	15
2	Random window	30
3	Sliding + Random window	45
4	Sliding + Class Balance Random window	Benign 45, malignant 22

improve model fitting and generalization ability. To solve the problem of class imbalance, Sliding + Class Balance Random window gives a random data acquisition method based on the first three and transforms unbalanced data into balanced. Therefore, it improved the classification performance of the model by randomly acquiring benign minority classes with directivity.

3.2.2. Network Model Design and Construction. We use the Keras framework to design CNN based on the AlexNet model. In the 2012 CVPR Competition, Krizhevsky et al. [26] proposed the AlexNet model, which consists of multilayer convolution, pooling, and nonlinear mapping layers. The AlexNet model’s network structure is shown in Figure3(a). After a large number of experiments, the network model with the best classification performance contains the following layers and parameters:

Input layer: the layer is the input of the network model and the output of the back propagation algorithm. The input layer image size is $128 \times 128 \times 3$.

Convolution layer: as an important structure for model feature extracting and learning, it is responsible for

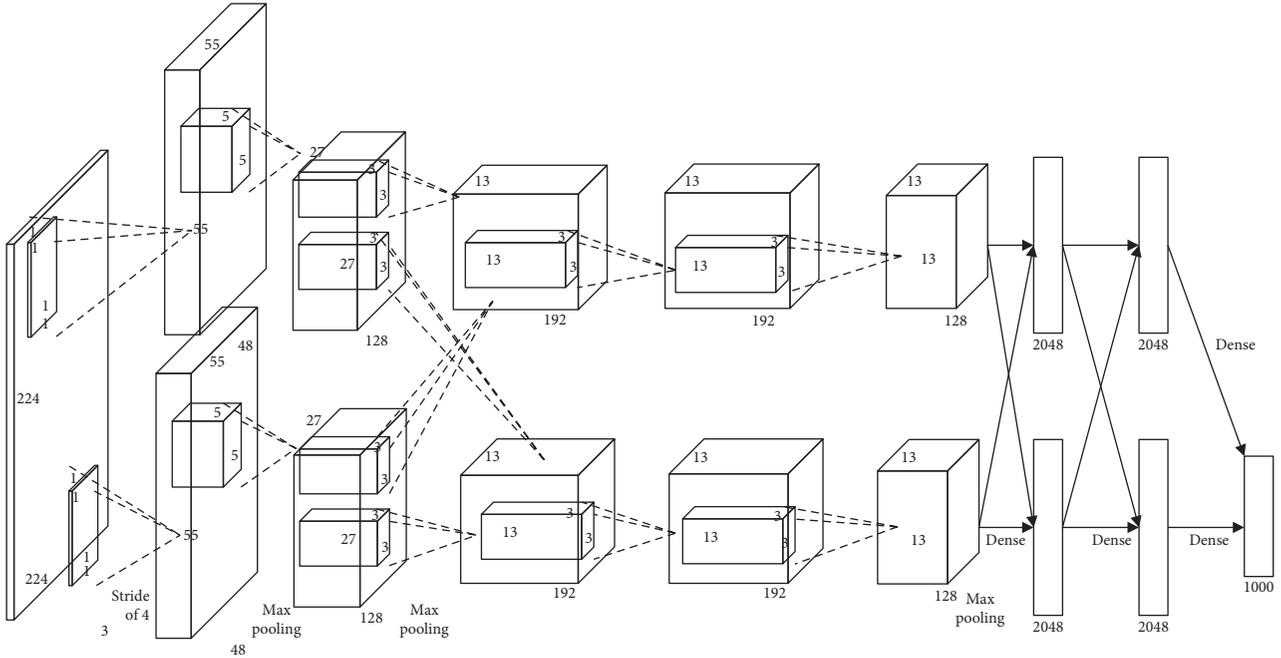


FIGURE 3: AlexNet model structure in [26].

gradually abstracting low-level features into high-level features. The model consists of three convolutional layers; the kernel size is 5×5 , 5×5 , and 3×3 ; the number of convolutional kernels is 16, 32, and 72; and the order of step is 3, 1, and 1. The zero filling method is adopted, and parameters of convolution kernel are initialized with the Gaussian distribution, and the bias is 0.

Pooling layer: the layer is responsible for reducing dimensionality of input features through down-sampling. Each convolutional layer is followed by a pooling layer. All pooling sizes are 2×2 ; step size is 2.

Nonlinear mapping layer: in this layer, low-level features are mapped to the high-level features to facilitate model classification. ReLU is used as the activation function in this process. Behind each convolution layer, a ReLU layer is added. The calculation formula of the ReLU function is shown as follows:

$$f(x) = \max(x, 0). \quad (2)$$

Global max pooling layer: in this layer, the maximum value of each feature map is output. The global max pooling layer can reduce fitting parameters and improve the generalization ability of the model.

Output layer: the layer is output of the model and input of the back propagation algorithm. The number of neurons in the output layer is set to 2, and softmax is used as the activation function.

The weight of the network layer is initialized by the Gaussian distribution, and the bias is initialized to 0. The structure of the CNN model and the configuration of the model's hyperparameter are shown in Tables 6 and 7.

TABLE 6: CNN model structure.

Layer	Type	Num kernels	Kernel size	Stride	Activation
0	Input	3	128×128	—	—
1	Convolution	16	5×5	3	ReLU
2	Max pool	—	2×2	2	—
3	Convolution	32	5×5	1	ReLU
4	Max pool	—	2×2	2	—
5	Convolution	72	3×3	1	ReLU
6	Max pool	—	2×2	2	—
7	Global max pool	—	—	—	—
8	Output	2	—	—	Softmax

TABLE 7: Hyperparameter settings for the CNN model.

Type	Values
Initial learning rate	$1e-4$
Learning rate schedule	Adam
Weight decay	$1e-5$
Train or test batch size	128
Loss function	Cross-entropy loss function
Number of epochs	100

3.2.3. Deep Feature Extraction and Machine-Learning Model Classification. According to the working principle of CNN, different convolutional layers have different feature extraction tasks. With the deepening of network layers, the low-level features are gradually abstracted to the high-level features. To fit data, CNN uses the high-level features to finish the corresponding learning tasks. Researches and experiments show that the classification effect of intermediate- and high-level features extracted by CNN is not worse than the high-level features; adopting an appropriate ML

classifier can improve the classification results. Therefore, we present a fusion method based on deep features and ML classifier for malignant-benign classification.

In Table 6, the trained CNN takes convolution + pooling layer as the splitting node. The models are split into three groups, and the global max pooling layer is added behind them to generate three new CNN models. Then, the data are input into the new CNN model for feature extraction, and the extracted features are input into ML classifiers for training. ML classifiers include SVM, LR, Gaussian naive Bayes (GNB), DT, RF, and multilayer feedforward neural network (MFNN). The feature extraction structure diagram is shown in Figure 4.

4. Results and Discussion

Firstly, patients used a fivefold stratified shuffle split cross-validation as the data splitting benchmark. This data splitting method is consistent in F. A. Spanhol team from literature [12, 22–24]. Secondly, the preprocessing operation normalizes the data and uses four window slicing strategies. Then, the preprocessed data are input into CNN for training and evaluation to obtain the best strategy. Once the optimal strategy is determined, the best feature extraction model nodes and choice should be discussed. Finally, after performing feature extraction, the extracted features are input into the ML model to classify. Figure 5 shows the main structure diagram of the experiment.

4.1. Experimental Environment. In this study, the experimental equipment is based on the Windows 10 system. The main hardware devices are 4-Core Xeon(R) W-2104 CPU @3.20 GHz and NVIDIA QuADro P2200 GPU with 4G video memory. Main software environments are PyChram 2021, Python 3.6.0, CUDA 9.0, CUDNN 7.0, TensorFlow GPU 2.2.0, Anaconda 3.5.0, and Keras GPU 2.4.3 version.

4.2. Evaluation Indicators. Two evaluation indicators, image-level accuracy and patient-level accuracy, are used to evaluate the classification performance of BC at different magnifications. The calculation formula of the image-level accuracy is shown as follows:

$$\text{image - level Accuracy} = \frac{\text{Number}_{\text{correct}}}{\text{Number}_{\text{image}}}, \quad (3)$$

where $\text{Number}_{\text{image}}$ represents the total number of all BC images. $\text{Number}_{\text{correct}}$ represents the number of correctly classified BC images. In the course of routine diagnosis, the pathologists need to evaluate the patient's overall cancer images to confirm health status. Therefore, patient-level accuracy is considered as an important evaluation indicator. The mathematical calculation formula is shown as follows:

$$\text{patient - level accuracy} = \frac{\sum_{p=1}^{\text{Number}_p} (\text{Number}_{\text{correct}}^p / \text{Number}_{\text{image}}^p)}{\text{Number}_{\text{patient}}}, \quad (4)$$

where $\text{Number}_{\text{patient}}$ represents the total number of patients. $\text{Number}_{\text{correct}}^p$ represents the number of correctly classified

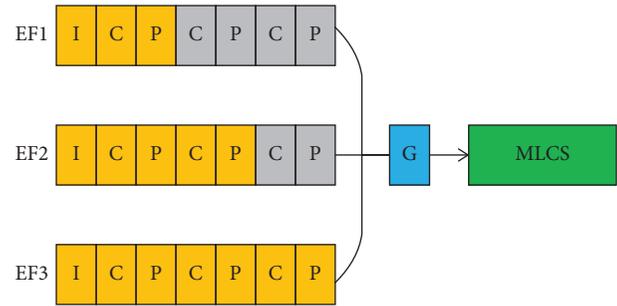


FIGURE 4: Feature extraction structure. EF1, EF2, and EF3, respectively, represent three different feature extraction methods. The orange area represents the effective feature extraction area, and the gray area represents the invalid area. I represents the input layer, C represents the convolution layer, P represents the max pooling layer, G represents the global max pooling layer, and MLCS represents the ML classifier set.

BC images of patient P, and $\text{Number}_{\text{image}}^p$ represents the total number of BC images of patient P. Due to the unbalanced distribution of the benign and malignant data, balanced accuracy (BAC) and F1 score are adopted to objectively evaluate the classification performance of the model. The mathematical calculation formulas of the two are shown as follows:

$$\text{balanced accuracy} = \frac{\text{recall} + \text{specificity}}{2}, \quad (5)$$

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \quad (6)$$

4.3. Experiment

4.3.1. Data Splitting. In this study, the dataset is split by the method of fivefold stratified shuffle split cross-validation. The cross-validation method can objectively evaluate the performance of the model and effectively avoid the evaluation result falling into the local optimal state. In the process of splitting the training set and the test set, 82 groups of patients are used as the splitting benchmark. 57 patients are in the training set of about 70%, while others are in the test set of about 30%. This data splitting method is the same as that of F. A. Spanhol team. When comparing the algorithm performance with it, the contribution of this study can be evaluated more objectively. The number of benign and malignant patients is 17 : 40 in the training set, while 7:18 in the test set. Both sets have similar class distribution, which is beneficial to objectively evaluate the model performance. The training set and the test set are shown in Table 8.

4.3.2. Convolutional Neural Network Classification. The data are normalized according to calculation formula (2). Then, we try to use four window slicing strategies to preprocess training data in Table 5. The preprocessed data are input into the model based on the CNN model for training and evaluation. As shown in Table 5, the data in Table 8 have

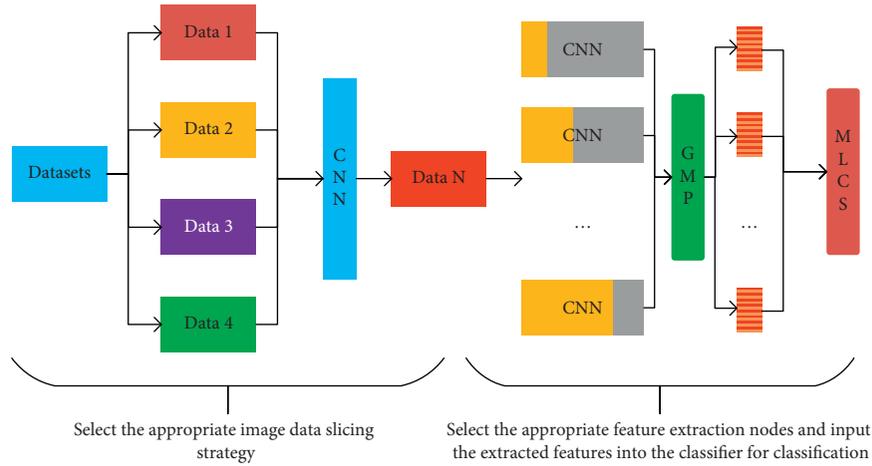


FIGURE 5: Experimental structure. Data 1 are obtained by Sliding window slicing. Data 2 are obtained by Random window slicing. Data 3 are obtained by Sliding + Random window slicing. Data 4 are obtained by Sliding + Class Balance Random window slicing. Data N are obtained by the best slicing strategy. The orange region represents the effective feature extraction region. The gray region represents the invalid feature extraction region. GMP represents the global max pooling layer. MLCS represents the abbreviation of ML classifier set.

been increased several times on different levels. The model structure of the CNN is shown in Table 6.

In the process of model evaluation, the small image of the test set is obtained by sliding in the test set image. The Sliding window is $128 \times 128 \times 3$ size, and the step is 128 pixels. The class of the large image is decided by voting. The voting rule is that the maximum number of image classes serves as the final prediction result for the whole image. Though other literature studies have not focused on classification problems from an imbalance view, we use F1 scores and BAC to evaluate model performance. The pathological tissue image classification results of BC at different magnifications are shown in Tables 9 and Tables 10.

As shown in Table 9 and Table 10, Sliding + Class Balance Random window slicing strategy has an excellent overall result on accuracy at different magnifications. Regarding the performance at image-level accuracy, the performance is more stable and has better evaluation results in F1 score and BAC accuracy. This experimental result shows that the presented method can alleviate the class imbalance problems. Regarding the performance at the patient-level accuracy, Sliding + Class Balance Random window slicing strategy is superior to other strategies in all magnifications. In a word, the Sliding + Class Balance Random window slicing strategy is more suitable for the following studies.

4.3.3. Feature Extraction and Machine-Learning Model Classification. For feature extraction, a well-trained CNN model takes convolution + pooling layer as feature extraction nodes after preprocessing of Sliding + Class Balance Random window slicing. For the ML model classification, the extracted features are input into the ML model for training and testing. Except for the MFNN, the hyperparameters of other classifiers (SVM, LR, GNB, DT, and RF) are all the default parameters in the Python-SkLearn ML Toolkit.

The combination of convolution and pooling layer is set standard for feature extraction nodes. Then, the global max

pooling layer is added after it. There are three cases in the number of input layer neural networks in Figure 4(i). In the MFNN training process, the three cases share the model structure and the hyperparameter configuration as shown in Tables 11 and 12. The features are extracted from the CNN of different nodes on the trained model and input into ML for classification. The accuracy, F1 score, and BAC are shown in Tables 13–15.

Table 13 shows that most ML classifiers have a better classification effect at the second node's position at different magnifications. Compared with other ML classifiers, SVM has the best classification results. Regarding the performance at image-level accuracy, the mean and standard deviation at different magnifications achieve 87.85%, 86.68%, 87.75%, and 85.30%, respectively. The results are 2.81%, 1.02%, 0.78%, and 1.01% higher than the pure CNN (Sliding + Class Balance Random) in Table 9. Besides, regarding the performance at patient-level accuracy, the mean and standard deviation at different magnifications are 87.93%, 87.41%, 88.76%, and 85.55%, respectively. The results of SVM are 1.87%, 1.12%, 0.07%, and 1.98% higher than the pure CNN (Sliding + Class Balance Random) in Table 9.

In Table 14, compared with other classifiers, SVM has the highest F1 score at the second node. The mean and standard deviation at different magnifications achieve 91.12%, 93.30%, 92.54%, and 90.45%, respectively. The results of SVM are 1.13%, 1.37%, 0.51%, and 1.04% higher than the pure CNN (Sliding + Class Balance Random) in Table 10.

In Table 15, compared with other classifiers, SVM has the highest BAC score at the second node. The mean and standard deviation at different magnifications achieve 86.57%, 87.64%, 89.42%, and 85.31%, respectively. The results of SVM are 1.08%, 1.62%, 1.42%, and 1.56% higher than the BAC at different magnifications of image classification by a pure CNN (Sliding + Class Balance Random) in Table 10.

In summary, the intermediate- and high-level features are very important for SVM classification. By this means, the

TABLE 8: Number distribution of the dataset with different magnifications.

Fold	Data type	Magnifications				Count
		40×	100×	200×	400×	
1	Train	1365 (392:973)	1396 (410:986)	1394 (414:980)	1248 (381:867)	5403 (1597:3806)
	Test	630 (233:397)	685 (234:451)	619 (209:410)	572 (207:365)	2506 (883:1623)
2	Train	1405 (436:969)	1449 (450:999)	1417 (432:985)	1291 (425:866)	5562 (1743:3819)
	Test	590 (189:401)	632 (194:438)	596 (191:405)	529 (163:366)	2347 (737:1610)
3	Train	1304 (378:926)	1383 (389:994)	1343 (389:954)	1180 (348:832)	5210 (1504:3706)
	Test	691 (247:444)	698 (255:443)	670 (234:436)	640 (240:400)	2699 (976:1723)
4	Train	1439 (441:998)	1553 (473:1080)	1448 (439:1009)	1291 (418:873)	5731 (1771:3960)
	Test	556 (184:372)	528 (171:357)	565 (184:381)	529 (170:359)	2718 (709:1469)
5	Train	1422 (455:967)	1530 (491:1039)	1458 (452:1006)	1339 (440:899)	5749 (1838:3911)
	Test	573 (170:403)	551 (153:398)	555 (171:384)	481 (148:333)	2160 (642:1518)

"Train" represents the training dataset, and "Test" represents the test dataset, and the number in braces represents the number of benign cancer and malignant cancer, respectively.

TABLE 9: Image-level accuracy and patient-level accuracy evaluation results at different magnifications (%).

Accuracy	Strategy	Magnifications			
		40×	100×	200×	400×
Image-level	Sliding	84.19 ± 2.27	83.80 ± 2.60	86.48 ± 2.80	83.93 ± 3.93
	Random	82.54 ± 2.81	84.41 ± 1.53	86.72 ± 2.77	82.74 ± 3.16
	Sliding + Random	84.50 ± 1.26	84.70 ± 1.25	86.95 ± 1.52	83.25 ± 2.47
	Sliding + class Balance Random	85.04 ± 0.87	85.66 ± 1.07	86.97 ± 0.77	84.29 ± 2.36
Patient-level	Sliding	85.70 ± 1.90	83.93 ± 2.12	87.73 ± 3.32	83.70 ± 4.04
	Random	83.73 ± 1.76	84.46 ± 2.52	88.31 ± 2.47	82.75 ± 3.35
	Sliding + Random	86.05 ± 2.17	86.03 ± 1.92	88.18 ± 3.31	83.31 ± 2.43
	Sliding + Class Balance Random	86.06 ± 1.62	86.29 ± 2.85	88.69 ± 4.17	83.57 ± 5.04

The optimal value of the evaluation result has been bolded.

TABLE 10: F1 score and BAC evaluation results at different magnifications (%).

Evaluation	Strategy	Magnifications			
		40×	100×	200×	400×
F1 score	Sliding	88.61 ± 1.64	88.69 ± 0.63	90.30 ± 2.15	88.07 ± 3.03
	Random	87.38 ± 1.95	89.04 ± 0.96	91.13 ± 3.01	87.28 ± 2.40
	Sliding + Random	89.31 ± 2.23	90.97 ± 0.82	90.97 ± 2.50	87.76 ± 1.45
	Sliding + Class Balance Random	89.99 ± 1.86	91.93 ± 0.71	92.03 ± 2.77	89.41 ± 4.31
BAC	Sliding	83.58 ± 2.04	83.75 ± 2.02	86.47 ± 4.37	82.55 ± 5.24
	Random	81.45 ± 2.45	83.92 ± 3.56	86.99 ± 4.48	81.14 ± 4.33
	Sliding + Random	84.73 ± 3.00	85.86 ± 2.24	87.05 ± 4.50	82.07 ± 3.43
	Sliding + Class Balance Random	85.49 ± 2.07	86.02 ± 3.31	88.00 ± 5.66	83.75 ± 6.28

TABLE 11: MFNN model structure.

Layer	Layer type	Number of neurons	Activation
1	Input	(16, 32, 72)	—
2	Fully connected	128	ReLU
3	Dropout	—	—
4	Fully connected	64	ReLU
5	Dropout	—	—
6	Output	5	Softmax

TABLE 12: Hyperparameter settings for the MFNN model.

Type	Value
Initial learning rate	1e-3
Learning rate schedule	SGD
Weight decay	1e-5
Train or test batch size	512
Loss function	Cross-entropy loss function
Number of epochs	500

classification effect of BC pathological tissue images has been improved as a whole.

Literature [23] uses a fully connected layer to extract features and input them into LR. We compare the combination, which is different levels of features and SVM from Table 13 to the former. In Table 16, the results of location points 1, 2, and 3

are better than fc6, fc7, and fc8 on image-level accuracy. Similarly, the results perform better than, or at least comparable to, the former on patient-level accuracy. In particular, Location Point 2 has the best overall results. The experimental results show that the convolutional layer is better than the fully connected layer on feature extraction. Our experimental method is consistent with the comparative literature.

TABLE 13: Results of feature extraction nodes on image-level accuracy and patient-level accuracy (%).

Accuracy	Point location	Classifier	Magnifications			
			40×	100×	200×	400×
Image-level	1	SVM	83.98 ± 2.90	84.99 ± 2.91	84.66 ± 2.41	82.78 ± 4.89
		LR	78.37 ± 2.72	81.03 ± 5.28	81.40 ± 7.19	80.04 ± 5.00
		GNB	61.96 ± 4.48	74.67 ± 6.82	78.38 ± 4.16	69.93 ± 6.29
		DT	80.60 ± 1.93	80.59 ± 3.29	84.02 ± 4.16	81.56 ± 3.73
		RF	80.43 ± 2.83	81.29 ± 2.93	84.48 ± 1.35	82.47 ± 3.61
		MFNN	79.56 ± 3.20	80.39 ± 4.76	81.76 ± 4.76	79.81 ± 7.76
	2	SVM	87.85 ± 2.69	86.68 ± 2.28	87.75 ± 2.37	85.30 ± 4.41
		LR	82.60 ± 3.34	82.48 ± 3.01	85.31 ± 2.20	82.89 ± 2.35
		GNB	81.17 ± 5.00	81.28 ± 3.30	85.91 ± 2.56	77.34 ± 6.61
		DT	80.88 ± 3.92	81.41 ± 2.33	85.33 ± 2.39	82.76 ± 2.90
		RF	80.60 ± 1.93	82.09 ± 2.82	85.68 ± 1.95	82.80 ± 3.9+
		MFNN	80.69 ± 3.15	82.94 ± 3.45	86.71 ± 3.14	83.34 ± 3.09
	3	SVM	84.91 ± 1.58	84.60 ± 1.78	87.38 ± 3.52	84.78 ± 4.89
		LR	81.00 ± 2.05	81.30 ± 1.30	85.23 ± 3.38	82.39 ± 4.94
		GNB	80.89 ± 3.88	81.36 ± 0.85	85.00 ± 3.15	80.48 ± 6.72
		DT	81.28 ± 2.40	80.94 ± 2.40	84.83 ± 2.46	81.89 ± 4.47
		RF	80.60 ± 1.93	81.31 ± 1.92	84.75 ± 2.51	82.43 ± 4.21
		MFNN	80.49 ± 2.85	84.69 ± 0.86	87.18 ± 3.63	82.80 ± 5.03
Patient-level	1	SVM	81.93 ± 3.92	80.96 ± 3.61	82.40 ± 2.37	81.17 ± 3.04
		LR	79.10 ± 3.91	80.81 ± 4.68	80.44 ± 5.33	78.21 ± 5.42
		GNB	68.16 ± 5.94	75.19 ± 3.35	78.94 ± 3.05	70.64 ± 6.80
		DT	79.81 ± 2.61	79.47 ± 3.72	83.63 ± 2.67	80.08 ± 3.13
		RF	80.67 ± 3.46	80.43 ± 3.17	83.76 ± 4.16	81.50 ± 2.91
		MFNN	80.34 ± 4.64	81.06 ± 4.53	81.92 ± 5.05	78.84 ± 3.72
	2	SVM	87.93 ± 3.91	87.41 ± 3.26	88.76 ± 2.50	85.55 ± 4.03
		LR	82.73 ± 2.84	82.76 ± 3.36	85.85 ± 2.90	81.60 ± 2.99
		GNB	78.65 ± 4.36	82.11 ± 3.92	84.47 ± 2.59	77.66 ± 5.07
		DT	81.57 ± 4.50	79.48 ± 3.32	85.15 ± 2.49	81.58 ± 2.84
		RF	80.96 ± 2.98	81.37 ± 3.55	84.54 ± 2.62	82.88 ± 2.90
		MFNN	83.32 ± 4.67	82.93 ± 4.72	87.64 ± 3.85	83.95 ± 2.63
	3	SVM	83.91 ± 0.96	83.49 ± 2.65	87.89 ± 3.46	82.28 ± 4.11
		LR	82.10 ± 0.97	82.44 ± 2.62	84.77 ± 3.75	81.10 ± 4.37
		GNB	82.35 ± 3.48	84.64 ± 1.39	83.93 ± 2.38	82.44 ± 5.14
		DT	80.82 ± 2.52	79.98 ± 2.54	84.97 ± 2.69	81.40 ± 4.12
		RF	80.72 ± 2.63	81.00 ± 2.77	84.42 ± 2.32	82.29 ± 3.92
		MFNN	82.58 ± 1.27	82.98 ± 2.83	87.95 ± 3.93	82.73 ± 3.88

TABLE 14: Extracted features from different feature extraction nodes on image-level F1 score (%).

Point location	Classifier	Magnifications			
		40×	100×	200×	400×
1	SVM	86.33 ± 2.05	87.17 ± 2.28	87.76 ± 2.05	84.28 ± 1.92
	LR	82.80 ± 2.41	85.21 ± 5.32	84.93 ± 7.62	83.67 ± 4.42
	GNB	74.81 ± 3.25	80.91 ± 2.67	82.54 ± 4.03	73.57 ± 4.86
	DT	85.04 ± 1.42	85.69 ± 2.43	88.20 ± 1.27	86.08 ± 3.13
	RF	83.69 ± 3.59	86.19 ± 2.16	88.52 ± 0.89	82.71 ± 6.64
	MFNN	84.07 ± 2.74	85.15 ± 5.0	85.91 ± 4.43	84.62 ± 4.00
2	SVM	91.12 ± 3.77	93.30 ± 4.71	92.54 ± 2.77	90.45 ± 3.41
	LR	87.97 ± 2.10	88.00 ± 2.93	88.73 ± 1.85	86.21 ± 1.51
	GNB	86.61 ± 3.68	86.86 ± 2.95	89.62 ± 2.42	86.05 ± 5.27
	DT	85.84 ± 2.71	86.74 ± 1.60	89.45 ± 1.86	86.23 ± 2.08
	RF	84.21 ± 2.87	87.45 ± 2.00	90.76 ± 1.35	88.23 ± 5.47
	MFNN	86.58 ± 2.40	87.33 ± 3.15	90.19 ± 2.41	87.67 ± 2.74
3	SVM	89.21 ± 0.95	90.98 ± 1.50	92.79 ± 2.66	89.83 ± 3.80
	LR	87.34 ± 1.27	87.55 ± 1.10	88.69 ± 2.54	86.56 ± 3.85
	GNB	85.12 ± 2.58	87.68 ± 1.61	89.55 ± 2.45	86.01 ± 5.00
	DT	85.08 ± 1.71	86.27 ± 2.12	89.62 ± 3.91	86.23 ± 3.40
	RF	84.80 ± 3.06	87.09 ± 1.48	90.04 ± 1.64	86.67 ± 3.21
	MFNN	87.71 ± 1.40	88.98 ± 0.68	90.80 ± 2.75	85.91 ± 2.40

TABLE 15: Extracted features from different feature extraction nodes are input into the image-level BAC evaluation results in the ML model (/%).

Point location	Classifier	Magnifications			
		40×	100×	200×	400×
1	SVM	80.07 ± 4.28	80.90 ± 3.45	81.77 ± 3.31	80.24 ± 3.98
	LR	76.80 ± 3.66	79.74 ± 4.90	80.44 ± 4.88	78.75 ± 4.74
	GNB	51.83 ± 8.18	69.25 ± 4.68	82.54 ± 4.03	69.74 ± 5.36
	DT	77.64 ± 3.93	78.07 ± 5.01	82.27 ± 2.84	79.47 ± 3.78
	RF	79.20 ± 3.99	78.91 ± 4.52	82.65 ± 2.23	80.39 ± 3.95
	MFNN	77.64 ± 3.86	78.98 ± 5.46	80.18 ± 5.30	77.52 ± 4.37
2	SVM	86.57 ± 4.06	87.64 ± 3.74	89.42 ± 3.31	85.31 ± 5.10
	LR	82.68 ± 4.60	84.05 ± 3.79	86.82 ± 3.46	79.05 ± 3.46
	GNB	73.84 ± 6.93	82.81 ± 2.31	88.62 ± 2.42	80.74 ± 5.36
	DT	80.87 ± 5.54	79.50 ± 4.10	83.68 ± 3.51	79.63 ± 3.88
	RF	80.40 ± 4.40	80.67 ± 4.80	84.04 ± 2.97	80.81 ± 4.38
	MFNN	80.33 ± 4.56	81.54 ± 5.39	87.16 ± 5.02	80.93 ± 3.75
3	SVM	83.21 ± 1.49	84.74 ± 3.22	87.88 ± 4.76	83.88 ± 5.56
	LR	81.49 ± 1.73	82.80 ± 2.93	85.75 ± 4.96	80.37 ± 5.46
	GNB	76.45 ± 5.73	87.68 ± 1.61	87.54 ± 2.45	79.95 ± 7.65
	DT	79.66 ± 3.51	78.85 ± 3.60	82.76 ± 3.71	79.91 ± 5.17
	RF	79.65 ± 3.97	79.27 ± 3.16	83.29 ± 3.15	80.52 ± 4.96
	MFNN	81.97 ± 2.03	83.29 ± 3.11	86.94 ± 5.15	82.11 ± 5.14

TABLE 16: Combination with SVM is compared with [23] (/%).

Accuracy	Location point	Magnifications				
		40×	100×	200×	400×	
Image-level	1	83.98 ± 2.90	84.99 ± 2.91	84.66 ± 2.41	82.78 ± 4.89	
	2	87.85 ± 2.69	86.68 ± 2.28	87.75 ± 2.37	85.30 ± 4.41	
	3	84.91 ± 1.58	84.60 ± 1.78	87.38 ± 3.52	84.78 ± 4.89	
	fc6	83.00 ± 2.60	84.60 ± 5.00	84.00 ± 2.80	81.10 ± 3.90	
	fc7	83.10 ± 2.10	83.30 ± 4.60	84.10 ± 1.50	81.60 ± 3.70	
	fc8	83.20 ± 2.40	84.00 ± 4.90	83.40 ± 1.10	80.90 ± 3.70	
	Patient-level	1	81.93 ± 3.92	80.96 ± 3.61	82.40 ± 2.37	81.17 ± 3.04
		2	87.93 ± 3.91	87.41 ± 3.26	88.76 ± 2.50	85.55 ± 4.03
3		83.91 ± 0.96	83.49 ± 2.65	87.89 ± 3.46	82.28 ± 4.11	
fc6		82.50 ± 8.60	83.60 ± 8.50	85.40 ± 5.20	81.10 ± 9.00	
fc7		83.40 ± 6.70	83.10 ± 8.40	86.00 ± 3.70	81.60 ± 8.60	
fc8		83.40 ± 6.90	83.80 ± 8.50	85.80 ± 3.50	80.70 ± 9.10	

The fc6, fc7, and fc8 are the three top-most layers in [23].

4.3.4. *Comparison with the Literature.* To verify the performance of the proposed method, we compare the optimal combination with relevant literature. To prove the validity of the comparative data, the experimental methods are the same as those in the compared literature. We use a fivefold stratified shuffle split cross-validation method, and the ratio of the training set and the test set is 7:3. Table 17 and Table 18 show the results of the comparison on image-level accuracy and patient-level accuracy.

In Table 17, regarding the performance at image-level accuracy the classification results are higher than those

models in the recent literatures at different magnifications. In Table 18, regarding the performance at patient-level accuracy, the results indicate that the proposed method performs better than, or at least comparable to, methods in terms of the quality of the classification from literature studies. Our method is not all superior to the optimal values of different magnifications compared with literature studies, but it is more accurate and stable in terms of overall model performance classification. In particular, the classification results on the patient-level accuracy are better than others at 200× magnification.

TABLE 17: Literature comparison results on image-level accuracy (%).

Magnifications	40×	100×	200×	400×
Literature [12]	—	—	—	—
Literature [22]	85.60 ± 4.80	83.50 ± 3.90	82.70 ± 1.70	80.30 ± 2.90
Literature [23]	84.60 ± 2.90	84.80 ± 4.20	84.20 ± 1.70	81.60 ± 3.70
Literature [24]	87.80 ± 5.60	85.60 ± 4.30	80.80 ± 2.80	82.90 ± 4.10
Literature [27]	—	—	—	—
Our method	87.85 ± 2.69	86.68 ± 2.28	87.75 ± 2.37	85.30 ± 4.41

TABLE 18: Literature comparison results on patient-level accuracy (%).

Magnifications	40×	100×	200×	400×
Literature [12]	81.60 ± 3.00	79.90 ± 5.40	85.10 ± 3.10	82.30 ± 3.80
Literature [22]	90.60 ± 6.70	88.40 ± 4.80	84.60 ± 4.20	86.10 ± 6.20
Literature [23]	84.60 ± 2.90	84.80 ± 4.20	84.20 ± 1.70	81.60 ± 3.70
Literature [24]	92.10 ± 5.90	89.10 ± 5.20	87.20 ± 4.30	82.70 ± 3.00
Literature [27]	83.08 ± 2.10	83.17 ± 3.50	84.63 ± 2.70	82.10 ± 4.40
Our method	87.93 ± 3.91	87.41 ± 3.26	88.76 ± 2.50	85.55 ± 4.03

5. Conclusions

In this study, we propose a novel classification method based on deep features of different levels to solve the BC classification problems. In the stage of data preprocessing, we present four slicing methods: Sliding window slicing, Random window slicing, Sliding + Random window slicing, and Sliding + Class Balance Random window slicing. The experimental results show that Sliding window slicing strategy can guarantee model's fitting training, and Random window slicing can enhance model's generalization training. Then, we combine the characteristics of both strategies and propose Sliding + Random window slicing strategy. This strategy performs well on model fitting ability and generalization ability. To overcome the classification problem caused by the unbalanced distribution of classes, we propose a Sliding + Class Balance Random window slicing. By comparing model performance, Sliding + Class Balance Random window slicing is the best data preprocessing strategy. In the stage of deep feature classification, features of different levels are combined with ML classifier, and the combination of intermediate- and high-level features with SVM has the best classification performance. The proposed method and some state-of-the-art method experiment on the BreaKHis breast cancer dataset. The experimental result shows that the proposed method can obtain better results than the ones reported in the relevant literature. It has been concluded that our method can be efficiently used for solving these problems due to its simplicity, reliability, and robustness. Although the method of benign and malignant BC has been presented, the classification of BC's subtypes has not been studied, and each subtype has a large imbalance ratio. In future research work, we aim to provide a DL algorithm for subtype classification.

Data Availability

The authors train and evaluate the method based on the BreaKHis datasets provided by F. A. Spanhol's team. This dataset is available at <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the submission and publication of this manuscript.

Acknowledgments

This work was substantially supported by the Daqing Normal University Land Enterprise Application Cultivation Project under no. 19ZR16.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] K. P. Traves and S. E. Cokenakes, "Breast cancer treatment," *American Family Physician*, vol. 104, no. 2, pp. 171–178, 2021.
- [3] G. Murtaza, L. Shuib, A. W. Wahid et al., "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges," *Artificial Intelligence Review*, vol. 53, no. 3, pp. 1655–1720, 2020.
- [4] J. Huang, P. S. Chan, fnm Lok et al., "Global incidence and mortality of breast cancer: a trend analysis," *Aging*, vol. 13, no. 4, pp. 5748–5803, 2021.
- [5] M. G. Hanna, V. E. Reuter, J. Samboy et al., "Implementation of digital pathology offers clinical and operational increase in efficiency and cost savings," *Archives of Pathology & Laboratory Medicine*, vol. 143, no. 12, pp. 1545–1555, 2019.
- [6] J. A. Retamero, J. Aneiros-Fernandez, and R. G. del Moral, "Complete digital pathology for routine histopathology diagnosis in a multicenter hospital network," *Archives of Pathology & Laboratory Medicine*, vol. 144, no. 2, pp. 221–228, 2020.
- [7] G. Litjens, T. Kooi, fnm Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [8] D. Li, D. Wang, fnm Dong et al., "False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: role of deep-

- learning-based CT diagnosis and insights from two cases,” *Korean Journal of Radiology*, vol. 21, no. 4, pp. 505–508, 2020.
- [9] L. Cong, W. Feng, Z. Yao, X. Zhou, and W. Xiao, “Deep learning model as a new trend in computer-aided diagnosis of tumor pathology for lung cancer,” *Journal of Cancer*, vol. 11, no. 12, pp. 3615–3622, 2020.
- [10] A. Ruiz-Casado, A. Álvarez-Bustos, C. G. de Pedro, M. Méndez-Otero, and M. Romero-Elías, “Cancer-related fatigue in breast cancer survivors: a review,” *Clinical Breast Cancer*, vol. 21, no. 1, pp. 10–25, 2021.
- [11] M. Maleki, A. Mardani, M. Ghafourifard, and M. Vaismoradi, “Qualitative exploration of sexual life among breast cancer survivors at reproductive age,” *BMC Women’s Health*, vol. 21, no. 1, pp. 1–10, 2021.
- [12] F. A. 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, 2015.
- [13] A. Alba-Bernal, R. Lavado-Valenzuela, M. E. Domínguez-Recio et al., “Challenges and achievements of liquid biopsy technologies employed in early breast cancer,” *EBioMedicine*, vol. 62, Article ID 103100, 2020.
- [14] A. E. Giuliano, “The evolution of sentinel node biopsy for breast cancer: personal experience,” *Breast Journal*, vol. 26, no. 1, pp. 17–21, 2020.
- [15] J. Bai, R. Posner, T. Wang, C. Yang, and S. Nabavi, “Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: a review,” *Medical Image Analysis*, vol. 71, Article ID 102049, 2021.
- [16] A. Khamparia, S. Bharati, P. Podder et al., “Diagnosis of breast cancer based on modern mammography using hybrid transfer learning,” *Multidimensional Systems and Signal Processing*, vol. 32, no. 9, pp. 747–765, 2021.
- [17] M. Sharma, S. Gupta, P. Sharma, and D. Gupta, “Bio-inspired algorithms for diagnosis of breast cancer,” *International Journal of Innovative Computing and Applications*, vol. 10, no. 3/4, pp. 164–174, 2019.
- [18] R. Ha, S. Mutasa, N. Gupta et al., “Predicting breast cancer molecular subtype with MRI dataset utilizing convolutional neural network algorithm,” *Journal of Digital Imaging*, vol. 32, no. 2, pp. 276–282, 2019.
- [19] A. Cruz-Roa, A. Basavanhally, F. González et al., “Automatic detection of invasive ductalcarcinoma in whole slide images with convolutional neural networks,” *Proceedings of SPIE-The International Society for Optical Engineering*, vol. 9041, pp. 139–144, 2014.
- [20] A. Anuranjeeta, K. 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.
- [21] L. P. Coelho, A. Ahmed, A. Arnold et al., “Structured literature image finder: extracting information from text and images in biomedical literature,” *Linking Literature, Information, and Knowledge for Biology*, vol. 6004, pp. 23–32, 2010.
- [22] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural networks,” in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2560–2567, Vancouver, Canada, July 2016.
- [23] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, “Deep features for breast cancer histopathological image classification,” *IEEE International Conference on Systems*, pp. 1868–1873, 2017.
- [24] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, “Multiple instance learning for histopathological breast cancer image classification,” *Expert Systems with Applications*, vol. 117, pp. 103–111, 2019.
- [25] L. G. Hafemann, L. S. Oliveira, and P. Cavalin, “Forest species recognition using deep convolutional neural networks,” in *Proceedings of the International Conference on Pattern Recognition(ICPR)*, pp. 1103–1107, Stockholm, Sweden, August 2014.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [27] 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(ICPR)*, pp. 2441–2446, Cancun, Mexico, December 2017.