Journal of Radiation Research and Applied Sciences 16 (2023) 100628

Contents lists available at ScienceDirect

Journal of Radiation Research and Applied Sciences

journal homepage: www.journals.elsevier.com/journal-of-radiation-research-and-applied-sciences

Breast ultrasound image classification and physiological assessment based on GoogLeNet

Shao-Hua Chen [a,*], Yan-Ling Wu [b], Can-Yu Pan [c], Luo-Yu Lian [d], Qi-Chen Su [a]

[a] Department of Ultrasound Medicine, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, 362000, China
[b] Respiratory Medicine, Quanzhou First Hospital of Fujian Medical University, Quanzhou, 362000, China
[c] Medical Imaging Department, Quanzhou First Hospital of Fujian Medical University, Quanzhou, 362000, China
[d] Department of Thoracic Surgery, Quanzhou First Hospital of Fujian Medical University, Quanzhou, 362000, China

ARTICLE INFO

Keywords:
Breast cancer diagnosis
Transfer learning
Deeping learning
Image classification
GoogLeNet

ABSTRACT

Background: Medical ultrasound image classification based on convolutional neural network is the mainstream breast cancer classification model, but its limited perceptual ability limits its ability to obtain global information.
Methodology: A total of 880 breast ultrasound images were collected from 700 patients including 103 normal images, 467 malignant tumor images and 210 benign tumor images. In this paper, the diagnosis of breast ultrasound images was realized by constructing the CNN model of GoogLeNet. Firstly, the breast images were preprocessed based on the TV model. After that, the CNN model was trained and a more accurate model with a wider range of application was obtained based on improved Inception. Then we extract image features of different sizes; Then, feature classification was completed in different model classifiers to realize benign and malignant detection of breast cancer. Meanwhile, comparative analysis was performed to verify the excellence of the GoogLeNet model.
Results: The training time for classification was effectively reduced, and the classification accuracy rate was improved, reaching 96.37% combined with transformer learning. The loss value is down to 0.3492. Then, the classification accuracy of different structural models is discussed in the two models. The results show that GoogLeNet has great advantages in the detection of breast cancer ultrasound images. The influence of migration on experimental results is further discussed. Finally, combined with transfer learning, the three models were tested separately. The results show that transfer learning can improve system performance.
Conclusion: Based on the TV model and the GoogLeNet model, this paper designs a method of breast cancer detection and classification based on the combination of TV and GoogLeNet model and transfer learning. Experiments show that this method can repair part of the texture damaged by markers in ultrasonic images, effectively extract the features of thyroid nodule, and accurately judge whether the breast is diseased, which greatly improves the diagnostic efficiency of doctors.

1. Introduction

Breast cancer is the most common disease among women and the main cause of death due to cancer in women (Bray et al., 2018). Medical ultrasound images are widely used due to their advantages such as low cost benefit, fast speed and non-invasive, and it has also become a common method to detect breast cancer (Chotai & Kulkarni, 2020; Nothacker et al., 2009). The computer-aided diagnosis system analysis of ultrasound images can effectively address the shortage and imbalance of physician resources in medical settings. Its ultimate goal is to achieve automatic classification of breast tumors based on computer-aided ultrasound images. Abnormal classification of breast is a crucial link in Computer Aided Diagnosis (CAD), and its classification results can assist experts to make a rough judgment of the patient's condition, which can quickly and directly relieve the pressure of doctors and reduce the rate of misdiagnosis. Wong et al. (Wong et al., 2012) used medical image diagnosis based on magnetic resonance imaging and added deep learning algorithm into CAD. In terms of cardiovascular diagnosis, Wong et al. (Wong et al., 2020) applied artificial intelligence in deep learning-based CAD. At present, there are two mainstream methods for classifying breast abnormalities. The first method is to first pretreat, manually locate and extract the general area of the lump, and then

* Corresponding author.
E-mail address: huashaochen@fjmu.edu.cn (S.-H. Chen).

https://doi.org/10.1016/j.jrras.2023.100628
Received 17 March 2023; Received in revised form 26 June 2023; Accepted 7 July 2023
Available online 20 July 2023

predict the obtained area of interest, which is more accurate but requires high labor cost; the second method is to predict on the whole image, which is cheaper and more flexible. The primary challenge lies in the fact that the affected area occupies a relatively minor portion of the overall image. Therefore, the key research focus is on extracting the lesion area's features accurately. In recent years, there have been many related studies on both methods.

Sehrawat et al. (Sehrawat et al., 2017) used wavelet transform to transform breast molybdenum target images into different frequency domains to extract more features so as to enhance the diversity of tumor characteristic information. The features that were extracted were classified by a Support Vector Machine (SVM) to improve the precision of the outcomes. Jadoon et al. (Jadoon et al., 2017) manually segment the focal area in breast molybdenum target image to obtain patch image, and then apply 2D Discrete Wavelet Transform to the pretreated image and Discrete Cosine Transform (DCT) obtained the characteristics of multiple frequency dimensions of the lesions, and then introduced a simple convolutional neural network to classify the extracted features directly to get the final results.In a similar vein, Ghiasi et al. (Ghiasi & Zendehboudi, 2021) employed Random Forest (RF) and Extremely Randomized Trees (ET) as classifiers for classification. Statistical analysis showed that the proposed method can accurately classify different types of breast cancer. Moon et al. (Moon et al., 2020), with the help of experts, manually extracted the region of interest in ultrasound breast images and marked the actual edge mask of the lump, so as to extract the characteristic information of the lump more accurately. The initial image, the region of interest image and the mass mask image were all fused to obtain the fourth type of data image, and then the four types of data were respectively put into four convolutional neural networks to extract the features, and finally very high accuracy was obtained. Mohammed et al. (Mohammed et al., 2018) also manually extracted the region of interest of ultrasound breast images, and did median filtering and image enhancement on it. Then, artificial neural network was designed to extract multi-scale features of the processed images, and finally sent to the classifier for classification. Ragab et al. (Ragab et al., 2021) used different networks such as AlexNet, GoogLeNet and ResNet to extract tumor features of different sizes. After feature extraction, the extracted features were integrated, and Principal Component Analysis (PCA) was employed to reduce data complexity. Finally, the feature dimensionality was reduced using the support vector machine for classification results. Medical images are often different from natural images. Natural images contain more objects and changes of various spatial positions, while breast images are gray maps, which contain fewer objects and changes. Therefore, in traditional machine learning methods, there are quite a few studies using frequency analysis to transfer signals to different frequency domains. Features are extracted from these different frequency components and used in SVM classifier (Jadoon et al., 2017; Sehrawat et al., 2017; Talha, 2016) to finally obtain the classification results of breast images. These previous studies proved that it is feasible to analyze breast image classification in frequency dimension.

Deep learning has become one of the main research tools in the field of computer vision. Tajbakhsh et al. (Tajbakhsh & Suzuki, 2017) used deep convolutional networks and traditional neural networks to detect and classify pulmonary nodules, and found that convolutional networks performed better. Setio et al. (Setio et al., 2016) proposed a multi-view convolutional network that integrates three candidate detectors to generate candidate regions. Li et al. (Li et al., 2020) introduced an interpretable deep learning framework that incorporates an "attention module" into convolutional neural networks (CNN), enabling doctors to better understand the decision-making process of CNN classification. CNN is the most common network in the fields of medical image classification, detection and segmentation in recent years, but its limited local receptive field restricts its ability to obtain global information. Qiu et al. (Qiu, Zheng, et al., 2021, pp. 200–208) propose a multi-improved residual network for super resolution of medical images. Cheng et al. (Qiu et al., 2021b, 2022) proposed a dual U-Net residual network and a multi-channel residual attention network for cardiac MRI image super resolution. In 2012, Szegedy et al. (Szegedy et al., 2015) proposed the GooLeNet network. Compared with AlexNet and VGG networks, GooLeNet has deeper network structure and fewer parameters. By adjusting the network structure and designing Inception layer, GooLeNet greatly reduces the number of references, thus improving the training efficiency. The performance of these convolutional networks in obtaining global structure information is poor, which limits their visual recognition ability. In this paper, we combine convolutional neural networks and TV model to classify breast ultrasound images.

2. Related work

2.1. Convolutional neural network

Convolutional neural network (CNN) is a multilayer perceptron for supervised learning, which is a popular technology in the current AI-based image recognition technology. It imitates the multi-layer process of human image recognition, similar to a biological neural network, extracts the features of the original image, thereby reducing the complexity of the model and reducing the parameters to be optimized, and is an effective method for extracting image features (Wang et al., 2018). Then, the extracted features are fed into the fully connected network. CNN includes input layer, hidden layer, and output layer, among which the common construction methods of hidden layer include convolutional layer, pooling layer, and fully connected layer.

The image input into the convolutional layer is mainly used to extract features, extract a feature in each feature map, complete the extraction of features through the convolutional layer, and update the weight through continuous back propagation during the training process. The convolutional layer is calculated as:

$$X_j = f\left(\sum_i x_i k_{ij} + b\right) \tag{1}$$

In Equation (1), f is the activation function, and k_{ij} is the convolution kernel and b is the offset.

Once the convolutional layer has extracted the features, the resulting feature map is passed to the pooling layer for feature selection and information filtering. The pooling operation is defined by Equation (2).

$$X_j = f\left(\beta_j L(x_j) + b\right) \tag{2}$$

In Equation (2), β_j is the weight coefficient and L is the downsampling function.

After aggregation, the information is fed into the densely connected layer, which corresponds to the implicit layer in the conventional feedforward neural network, and the architecture of the convolutional neural network that propagates the signals is illustrated in Fig. 1.

Equation (3) displays the cross-entropy formula, where y denotes the observed value and y' represents the anticipated value. Equation (4) illustrates the utilization of exponential decay in the learning rate, where r is the learning rate and f' signifies the gradient of the loss function.

$$H = -\sum y \times \log y'. \tag{3}$$

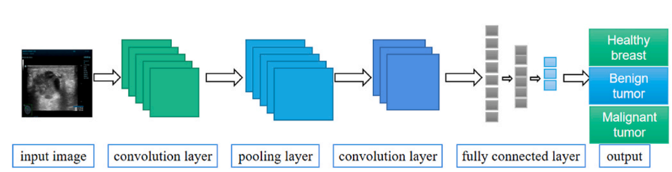

Fig. 1. CNN network structure diagram.

$$w_{n+1} = w_n - r \times f'. \tag{4}$$

2.2. Transfer learning

In order to solve the problem of neural network parameter mismatch caused by transfer learning between two completely unrelated domains, parameter adjustment is a very important step. Transfer learning can be divided into the following steps (Ao et al., 2021): (1) Select a large number of datasets with complete labels and associated with the target domain, use them to train parameters on the feature extraction layer of the neural network model, and use the trained model to effectively extract features from the target sample; (2) Effectively customize the remaining fully connected layer and classification layer of the neural network in line with the actual task; (3) Finally, according to the actual situation, the final task sample is used to train and fine-tune the parameters of the final reconstructed network model to complete transfer learning. Because the experimental dataset is relatively small, and the selected images are different from various images in the source domain, the model is fine-tuned by freezing the model.

Transfer learning can be categorized into three types: direct transfer learning, inductive transfer learning, and self-supervised transfer learning according to the relationship between the two types of samples and the purpose of the final learning (Pan & Yang, 2010). Transfer learning can use completely different types of source domain samples and task domain samples, but they are not irrelevant, and its basic requirements and principles are pre-trained models based on source domain training and learning, which also have a good test effect on the target task. Transfer learning is widely present in people's lives, the more the intersection of two different fields, the better the difficulty of transfer learning, on the contrary, the difficulty and effect of transfer are less ideal, and even produce the phenomenon of "negative transfer" (Dai et al., 2009), which is contrary to the ideal effect. In recent years, transfer learning has not only blossomed in theoretical research, but has also been frequently applied to different types but related industries. Yang et al. (Yang et al., 2020) In order to overcome the shortcomings of deep learning in breast cancer sentinel lymph node metastasis CT image data, the CNN-Fast network model trained on the ImageNet dataset was used for the classification of breast cancer sentinel lymph nodes through transfer learning, and the optimal results of current CT-based research were obtained. Lévy used the pre-trained AlexNet (Krizhevsky et al., 2012) deep learning model for the identification of benign and malignant breast cancer, and achieved diagnostic results that were superior to existing research results (LévyJain, 2016). The above studies show that transfer learning has been widely used in many tumor types and has good results, and it is now a feasible solution for breast cancer ultrasound image recognition. Transfer learning can solve the problem of insufficient training samples due to small datasets in deep learning.

2.3. GoogLeNet neural network model

GoogLeNet is a deep learning structure proposed by Google in 2014 (Cao et al., 2020). The general CNN structure only purely enlarges the network, which has two disadvantages: overfitting and increasing the amount of computation. The core idea of the Inception module introduced by GoogLeNet is to extract features from input samples through convolutional kernels of different sizes, and then combine the results obtained by each part, and finally input them into the next layer of the network, which can make full use of computing resources, extract richer image features under the same conditions, and obtain higher accuracy. The solution to both problems is to increase the depth and width of the network while reducing the parameters. To decrease the number of parameters, the complete linkage should transform into a scattered linkage. However, during implementation, the computation won't be significantly enhanced after the transformation due to the fact that most hardware is optimized for dense matrix calculations. Even though the data in the scattered matrix is minimal, the time consumed remains difficult to reduce. The inception structure can maintain the network structure's sparsity while also utilizing the dense matrix's high computing performance. Fig. 2 shows the original module diagram of Inception.

3. Methods

3.1. Construction of the dataset

3.1.1. Dataset amplification

This experiment collected a small number of datasets and required amplification of the datasets. This article amplifies the dataset using rotation and translation only. Image rotation refers to the process of turning an image at a certain angle according to a certain position, and image translation is to move all pixels in the image horizontally (in the X direction) or vertically (in the Y direction) according to a given amount of translation.

3.1.2. Dataset imbalance

Data imbalance is a common occurrence in datasets and can take many different forms. In medical applications, data imbalance is often observed in image datasets used for classification tasks. Imbalanced data can cause the learning algorithm and network to prioritize learning from the majority class samples, thus ignoring the minority class samples. This can result in incorrect classifications, such as misjudging a cancer patient as a non-cancer patient, which can have severe consequences for the patient's health and well-being. Hence, the positive samples in this paper's dataset, although limited in number, hold crucial information for the ultimate classification task, and any misclassification of these samples would result in significant costs (He & Garcia, 2009). Data imbalance in classification tasks can lead to two false negative rate. Therefore, it is necessary to process unbalanced data sets before the experiment to achieve balance. Unbalanced data processing methods include downsampling, upsampling, data synthesis, and cost-sensitive learning. To preprocess the unbalanced data set, the random upsampling algorithm (Liu et al., 2007, pp. 66–72) was chosen. This algorithm preserves the information in a few sample sets and ensures zero information loss (Longadge & Dongre, 2013).

3.1.3. Image preprocessing

Noise often exists in acoustic image data sets, and the type of noise is generally speckled noise. Median filter can deal with speckle noise well in spatial filtering.

The median filter can eliminate the speckle noise and retain the edge information in the image. However, the template of median filtering should not be too large, otherwise the image is prone to blur. In this experiment, 3×3 median filtering is adopted, and the results are shown in Fig. 3. It can be seen that the speckle noise is removed from the ultrasound image after median filtering, and the edge of the tumor becomes more obvious.

3.1.4. Image restoration based on the TV model

Experts perform manual annotation, which involves marking the affected region in ultrasound images, thereby compromising the texture

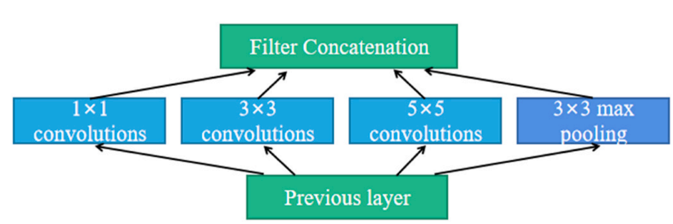

Fig. 2. The original module diagram of Inception.

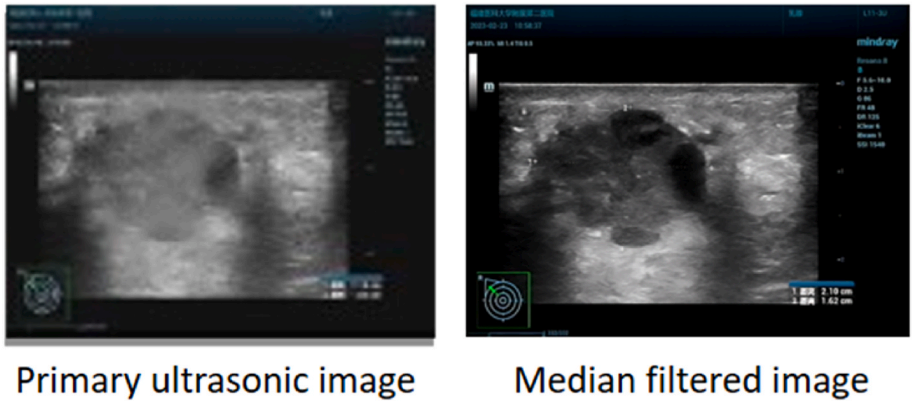

Fig. 3. Median filter preprocessing of breast ultrasound images.

and precision of the analyzed area's image. The data at hand is labeled by individuals. Experts manually label the data, which entails identifying the damaged area in ultrasound images, leading to the loss of some grain and reducing the accuracy and completeness of the analyzed area's image. The current dataset is annotated by human annotators. Professionals carry out manual annotation, which involves indicating the affected area in ultrasound images, resulting in the distortion of some texture and impacting the image's accuracy and entirety in the analyzed region. This has an impact on subsequent training, so the image needs to be repaired. In this study, we used the adaptive image restoration of the TV model for the images. The equation for estimating the value of the pixel after repair is shown in Equation (5), where G_O represents the pixel of point O to be repaired and G_p represents the pixels of the four neighborhood points of the current point O. The weight coefficient is denoted by H_{OO} which is determined by W_p and its calculation method is shown in Equation (6), Equation (7) and Equation (8), where $p \in A$, and ∇g_p represents divergence while $\lambda(O)$ is the λ parameter at O.

$$G_O^{(n)} = \sum_{p\in A} H_{O_p}^{(n-1)} G_p^{(n-1)} + H_{OO}^{(n-1)} G_O^{0}. \tag{5}$$

$$W_p = \frac{1}{\sqrt{a^2 + |\nabla g_p|^2}} \tag{6}$$

$$H_{O_p} = \frac{w_p}{\sum_{p\in A} w_p + \lambda(O)} \tag{7}$$

$$H_{OO} = \frac{\lambda(O)}{w_p + \lambda(O)} \tag{8}$$

3.2. The development of inception model and improvement of GoogLeNet model

Since AlexNet won the ImageNet competition in 2012, this model has been successfully applied in a large number of computer vision tasks, corresponding to AlexNet is based on LeNet network deepening and corresponding ReLu function and corresponding dropout layer. At the same time, another VGGNet was proposed and started to be used. This model has better generalization ability, smaller convolution kernel and deeper level compared with AlexNet. On the other hand, the model has good generalization ability, especially for image feature extraction target detection candidate box generation. But the biggest problem of this model is the large number of parameters. The inception model, which focused on this problem at the time it was proposed, did not make extensive use of fully connected networks as in VGGNet (Srivastava et al., 2014), so the final number of parameters was relatively small. This progressive feature has contributed to the widespread application and further development of the inception model. Since 2014, CNN has improved the quality of its network by increasing its depth and width. However, this method of purely increasing the depth of the network tends to cause the following shortcomings: (1) Overfitting phenomenon may occur when the number of network layers is deepened but the corresponding training set is relatively small; (2) With the deepening of the corresponding depth and width of the network, the calculation becomes more and more complex and difficult to apply. Inception networks, although designed to execute under strict memory and computational conditions, have very large parameter reductions and much smaller computational costs compared to simpler VGG and AlexNet before them. This advantage makes it suitable for many big data scenarios. In response to the above problems, Inception model design principles are as follows: (1) For early networks, avoid representation bottlenecks. The information flow corresponding to the input is obtained from the segmentation input and output input. The law of representation should be from the input to the output slowly decline, get the final result, and for some important factors of correlation structure abandoned network can not be maintained by the representation characteristics; Dimensions can only be used as initial estimates of content information; (2) The network achieves the purpose of accelerating the training speed by acquiring non-entanglement features; (3) Aggregation can be carried out in low-dimensional space, which can preserve the performance of the network; (4) Balancing the width and depth of the network can improve the quality of the network, but for the width, it can only be effective if the parallel is improved. Therefore, using this method requires a reasonable balance between the depth and width of the network.

However, there are still many problems in practical applications, so the Inception module has been improved, as shown in Fig. 4. The more times the network has been convolved, the more abstract the features will be, and the larger the sensitivity field involved in each feature will be. Therefore, with the increase in the number of layers, the ratio of 3 × 3 and 5 × 5 convolution will also increase correspondingly. The use of 5 × 5 convolution kernel will bring about a huge amount of computation, so 1 × 1 convolution kernel is adopted for dimension reduction. The

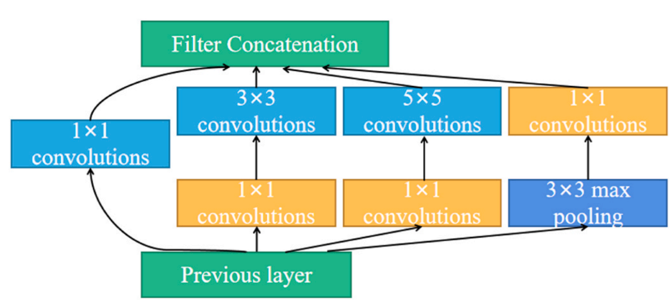

Fig. 4. Improved Inception model.

improved Inception structure not only greatly reduces the number of parameters, but also adds 1 × 1 convolution for feature dimension reduction. At the same time, followed by nonlinear activation functions, the generalization ability of neural networks has been improved. GoogLeNet uses InceptionModule, uses auxiliary classification nodes, and replaces large convolution kernel with multiple small convolution kernel, which improves the utilization rate of parameters and has good classification performance.

When using a deep neural network, the gradient can have difficulty propagating effectively through all layers. To address this issue, the success of a shallower network implies that the middle layers of the network generate highly distinctive features. One solution is to incorporate additional classifiers in these intermediate layers, which can enhance the discriminative capabilities of the lower-level classifiers. This is supposed to overcome the gradient disappearance problem while providing regularization. These classifiers take the form of smaller convolutional networks placed on top of the inception module's output. During the training period, their losses are added to the total losses of the network with a discount weight (the auxiliary classifier loss weight is 0.3). These secondary networks are discarded when inference is made. Later control experiments show that the effect of the auxiliary network is relatively small (about 0.5), and only one of them is needed to achieve the same effect. Finally, the network adopts average pooling to replace the full connection layer. Meanwhile, in order to avoid the gradient disappearing, two auxiliary softmax are added to the network to conduct the gradient forward.

4. Experiments

4.1. Datasets

A total of 880 breast ultrasound images were collected from 700 patients aged 25–75 years, including 103 normal images, 467 malignant tumor images and 210 benign tumor images. Among them, some images also have multiple lumps, that is, one image corresponds to multiple masks. If images with multiple masks are regarded as multiple data pairs, a total of 630 pairs of ultrasound-mask image pairs can be used for segmentation in this data set. The production tool of the data set is Labelme image labeling software written in python language. The production of each label is carried out under the guidance of professional doctors. The original data set is shown in Fig. 5. Fig. 5(a) to Fig. 5(b) are self-made breast ultrasound images, respectively. The size of the original image and label is 768 × 1024.

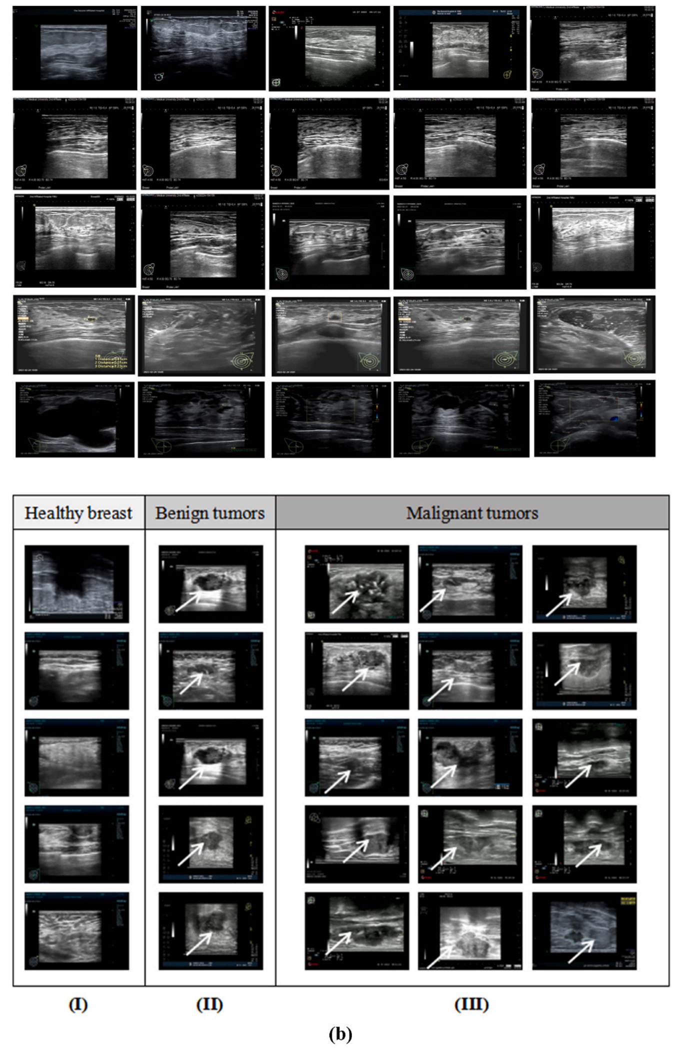

Fig. 5. The original image of the data set depicting (a) the ultrasonic image of normal breast; (b) the ultrasonic image of breast cancer (images with arrows indicate the presence of tumor and those images without arrows indicate healthy breasts. The I part indicates healthy breasts; the II part indicates benign tumors; the III part indicates malignant tumors).

4.2. Experiments details

4.2.1. Dataset processing

Existing data contains human labeling. This has an impact on subsequent training, so the image needs to be repaired. In this study, we used the adaptive image restoration of the TV model for the images. In order to verify the accuracy of the GoogLeNet model trained from scratch and compare it with AlexNet, we divided the data set into three parts: 60% was used as the training set, 20% as the verification set and 20% as the test set. The model performance was tested and compared separately. When comparing and analyzing the classification characteristics of GoogLeNet model combined with transfer learning, the amplified case charts were used to train the three models based on the improved LeNet5, VGG16 and GoogLeNet. Then 600 plots were divided into 5 groups as test data sets to test the above three models. Then LeNet5, VGG and GoogLeNet models were used to conduct experiments on the training set and the test set respectively.

4.2.2. Breast ultrasound images classification based on GoogLeNet neutral network

The whole network is divided into 5 blocks, each of which has 2 layers of convolution and 1 layer of maximum pooling layer. For the first module, the size of the volume convolution kernel of the corresponding two convolutional layers is set to 32, and the convolution is carried out in a dimensionally-invariant way. Then, relu activation function is used to increase the nonlinearity of the model, and l2 regularization method is adopted for the convolutional layer to reduce the possible overfitting problem of the whole model. After two layers of convolution, the corresponding output is transferred to a maximum pooling layer, and the corresponding pool_size is set to (2,2), so that the output of the whole network is 32*50*50. Then, with the same parameters as the first module, the number of convolutional kernels is 48. After the same structure as the last module, the output of the whole network is converted to 48*25*25. Then input the corresponding output into the next block, and set the size of the convolution kernel of the corresponding convolution layer to 64. After passing the corresponding four convolution layers and one maximum pooling layer, the output of the whole model is converted to 64*12*12. After passing these convolution and pooling layers, then through a similar module, However, the number of convolutional nuclei in the convolutional layer is set to 96, so that the output of the entire network is 96*6*6. Then, through a similar module, the number of corresponding convolutional nuclei is set to 128. Finally, after passing the entire output into the full connection layer, the overfitting of the network is reduced by using dropout and L2 regularization methods.

4.3. Comparative analysis of traditional GoogLeNet and AlexNet

The basic framework of deep learning in this experiment uses domestic PaddlePaddle2.1.2 and python3.8 as programming tools. The main reason for choosing PaddlePaddle is that all kinds of classical neural network models with optimized structure can be loaded in the framework, such as LeNet, AlexNet, GoogLeNet, etc. The activation function used in this experiment network is ReLU function, and the equations is shown as Equation (9), where x is the input variable.

$$f(x) = \max(0, x). \tag{9}$$

Compared with the traditional activation functions Sigmoid function and Tanh function, ReLU has higher efficiency in gradient descent and back propagation. Rectified Linear Unit (ReLU) is an unbounded activation function, which means that its derivative is non-zero when the input is positive. This property ensures that the gradient won't vanish, preventing the network from overfitting or getting stuck in a local minimum. The loss function in the experiment is the cross entropy function, as shown in Equation (10).

$$L = \frac{1}{N}\sum_{i} - [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]. \tag{10}$$

In Equation (10), y_i represents the label of the sample, with positive class 1 and negative class 0 and p_i represents the probability that the sample prediction is positive.

The cross-entropy function is not only simple in the back propagation, but also can improve the learning speed of deep learning with the activation function of the output layer.

According to the above, the two models are trained separately, and the loss value of the training set, the loss value of the verification set and the loss value of the test set of the two models are simply compared, and then the accuracy of the verification set and the test set are combined as a comprehensive evaluation method.

As shown in Fig. 6, the loss value comparison of training set shows that the loss value of GoogLeNet is relatively lower and declines gently than that of AlexNet.

As can be seen from the comparison diagram of loss value of verification set in Fig. 7, the GoogLeNet loss value starting from zero is relatively lower and the convergence value is more ideal in the same iteration times. As for the loss value of the test set, it can be seen from the comparison diagram of the loss value of the two model test sets in Fig. 8 that the loss value of GoogLeNet is still low and the convergence rate is more stable.

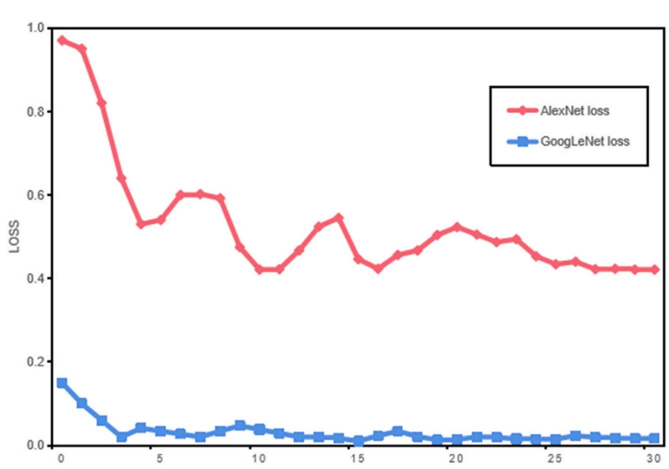

Fig. 6. Model training set loss value.

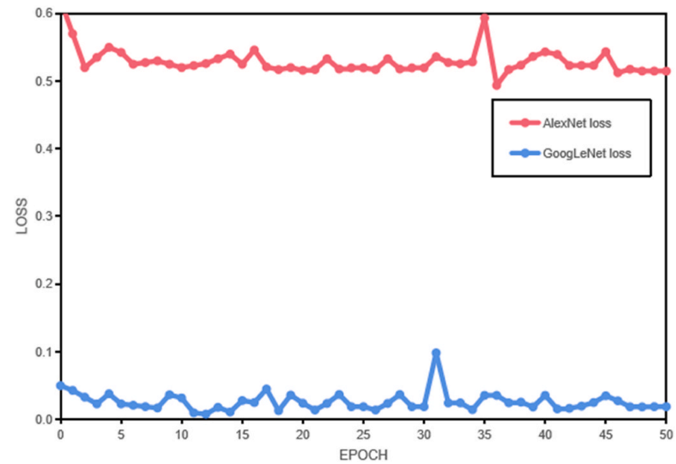

Fig. 7. Model validation set loss value.

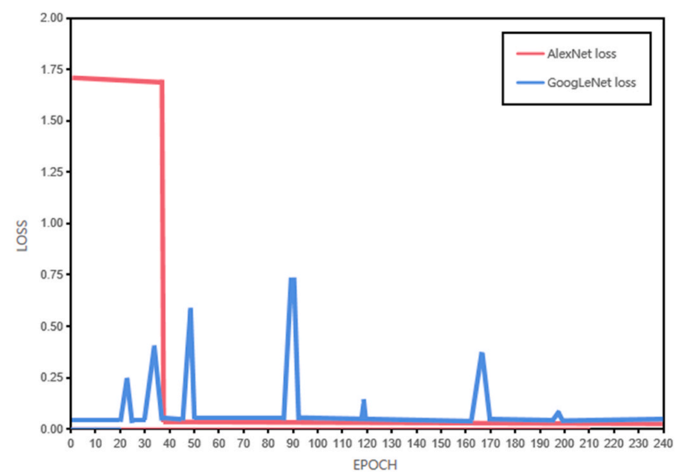

Fig. 8. Model testing set loss value.

The training accuracy of the two models is shown in Table 2. Through observation of Table 1 and the figure above, it is found that the GoogLeNet model has a good recognition effect, and the accuracy of verification set and test set is as high as 96.34% and 93.23%, which is obviously better than the accuracy of AlexNet model trained from scratch, which is 82.21%.

4.4. Comparison of the classification characteristics of GoogLeNet, VGG16 and LeNet5 combined with transfer learning

After training with 1605 breast ultrasound images and testing with 5 groups of approximately 120 images each, totaling 600 images, the discriminant accuracy of the three models is shown in Table 2.

The loss values of the three models are shown in Table 3, and the changing trend of the loss values of the three models in continuous iteration is shown in Fig. 9.

Table 2 shows that the GoogLeNet model has the highest accuracy in determining whether thyroid nodules are diseased. As can be seen from Table 3, the loss value of GoogLeNet model is relatively minimum.

Table 1
The accuracy of the training of the two models.

Model	Validation set accuracy	Test set accuracy
AlexNet	82.21	87.02
GoogLeNet	96.34	93.23

Table 2
Accuracy of different CNN models in the diagnosis of breast cancer ultrasound images/%.

Sample number	LeNet5	VGG16	GoogLeNet
1	81.76	92.45	97.82
2	85.34	90.34	94.55
3	86.23	89.88	97.07
4	84.98	92.65	96.41
5	85.32	88.97	95.99
Average	84.73	90.86	96.37

Table 3
The loss value of the training of different CNN models.

Sample number	LeNet5	VGG16	GoogLeNet
1	0.6274	0.5293	0.3985
2	0.7248	0.3794	0.2984
3	0.6298	0.4974	0.3023
4	0.7093	0.4297	0.3435
5	0.7322	0.5084	0.4017
Average	0.6847	0.4688	0.3492

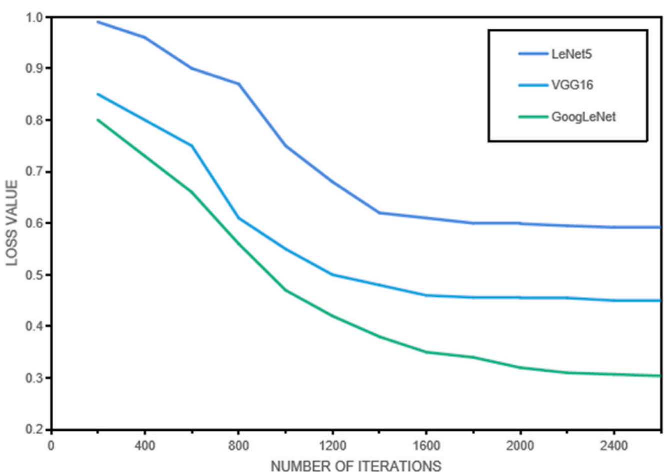

Fig. 9. The loss value of the training of LeNet5, VGG16 and GoogLeNet.

5. Discussion

Breast cancer is a kind of clinical disease with high incidence, and its incidence is increasing year by year. At present, the most common means to diagnose breast cancer is ultrasound (Xia et al., 2014). However, doctors' artificial discrimination is subjective to some extent and will be affected by fatigue and other factors. Therefore, the application of deep learning and the construction of convolutional neural network for breast cancer diagnosis has far-reaching practical significance for improving doctors' diagnostic efficiency. Image recognition and classification involve a variety of technologies, including image and processing, feature extraction, training model building, etc. (Hu et al., 2018). In recent decades of research by domestic and foreign experts and scholars, image classification has changed from manual classification to machine learning classification. As a new machine learning algorithm, deep learning plays an important role in science and technology, biology, machinery and other fields with its powerful learning efficiency and objective modeling speed (Tao & Wang, 2019). Well-known deep learning models include LeNet5 model (Le Cun et al., 1998), AlexNet (Krizhevsky et al., 2017) model, VGG16 model (Simonyan & Zisserman, 2014), GoogLeNet model (Rzegedy et al., 2015), ResNet model (He et al., 2016), etc.

Based on TV model and GoogLeNet model, this paper designs a method for breast cancer detection and classification based on TV and GoogLeNet model. Experiments show that this method can repair part of the texture damaged by markers in ultrasound images, and effectively extract the features of breast images for accurate judgment, which greatly improves the diagnostic efficiency of doctors.

Compared with other methods in recent years, the results show that the proposed method has the best performance in the end-to-end, fully automatic network. Although deep learning is widely used in the field of medical image processing, it still has many bottlenecks to be broken through and needs to be improved in the medical field.

Despite its strong generality, deep learning requires a large amount of data due to the privacy and professionalism of medical images, making its development in the field more challenging. When there is sufficient data, deeper networks generally perform better. However, as the complexity of the network increases, so does the number of parameters and computational consumption, which slows down the training speed and requires more advanced hardware equipment. The technology of deep and shallow feature fusion of images has gradually developed under the influence of deep learning technology and is suitable for medical diagnosis (Zhou et al., 2022). To apply deep learning algorithms to clinical practice, future studies need to focus on designing more efficient and lightweight networks. Deep learning's high data requirements, due to the privacy and professionalism of medical images, make it difficult to develop in the field. Deeper networks perform better with sufficient data, but their complexity increases parameters and computational consumption, slowing down training speed and requiring advanced hardware. More efficient and lightweight networks are needed for clinical practice. Not only ultrasound images, but also CT image processing is within the scope of research for various treatments (Zhao, Cheng, et al., 2023). Deep learning's strong generality is offset by its high data requirements due to the privacy and professionalism of medical images, making development in the field challenging. The technique of extracting features from images to determine tissue necrotic organ lesions is also reflected in MRI (Zhang et al., 2018). Deeper networks perform better with sufficient data, but their complexity increases parameters and computational consumption, slowing down training speed and requiring advanced hardware. Future studies should focus on designing more efficient and lightweight networks to apply deep learning algorithms to clinical practice. In practice and also in future implementation, state of the art three-dimensional visualization algorithms (Zhao, Lul, et al., 2023) can be utilized in the volumetric visualization of breast lesions so as to evaluate cancer conditions.

6. Conclusion

In this paper, the adaptive image repair method based on TV model was used to preprocess the breast ultrasound image. Subsequently, the GoogLeNet architecture was established, and the corresponding cost function, training rate, exponential moving average, and optimization technique were configured for the purpose of optimization, and the GoogLeNet improved based on transfer learning was trained for the diagnosis of benign and malignant breast cancer. After that, we obtained the accuracy of the traditional AlexNet model and GoogLeNet model, and the improved LeNet5 model and VGG16 model to distinguish benign and malignant breast diagnosis. When analyzing the classification characteristics of GoogLeNet model combined with transfer learning, we compared it with LeNet5 and VGG16, and compared and analyzed the classification accuracy of the three models in the training set and the test set.After training, all the three models can complete the recognition task. Through comparison, it can be seen that GoogLeNet has the highest accuracy in determining whether the breast is diseased, and the practical application level is also very high. The average accuracy of the improved GoogLeNet model can reach 96.37% and the loss value of 0.3492, which well indicates that the GoogLeNet model can diagnose patients with breast disease. The research presented in this paper employs deep learning techniques to aid in medical diagnosis. The subsequent phase will involve progressive model optimization and further research to ensure highly accurate classification diagnosis of breast ultrasound images. The deep learning-based image classification diagnosis method can serve as a valuable reference for physicians in diagnosing such ailments, enhancing diagnostic efficiency and accuracy, reducing labor costs, and offering novel approaches to future breast cancer ultrasonic diagnosis.

Ethical approval

The patients in this paper have given written consent to the study.

Funding

Authors acknowledge the help from technical reports provided by Deep Red AI system platform [https://www.deepredsys.com].

Declaration of competing interest

The authors declare no conflict of interest for this paper.

References

Ao, Z. L., Ma, R. Y., & Yang, X. W. (2021). Classification of engine borehole image based on DenseNet and ResNet fusion. *Computing Technology and Automation, 40*(3), 105–110.
Bray, F., Ferlay, J., Soerjomataram, I., et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca - a Cancer Journal for Clinicians, 68*(6), 394–424.
Cao, F. K., Bai, T., & Xu, X. L. (2020). Vehicle detection and classification based on highway monitoring video. *Computer Systems&Applications, 29*(10), 67–273.
Chotai, N., & Kulkarni, S. (2020). *General considerations in breast imaging.* Singapore: Springer.
Dai, W. Y., Jin, O., Xue, G. R., Yang, Q., & Yu, Y. (2009). Eigen transfer:a unified framework for transfer learning. *Morgan Kaufmann Publishers, 14*(6), 193–200.
Ghiasi, M. M., & Zendehboudi, S. (2021). Application of decision tree-based ensemble learning in the classification of breast cancer. *Computers in Biology and Medicine, 128*, Article 104089.
He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284.
He, K. M., Zhang, X. Y., Ren, S. Q., et al. (2016). Deep residual learning for image recognitioN[C]. In *Proc of IEEE conference on com puter vision and pattern recognition* (pp. 770–778). Piscataway, NJ: IEEE Press.
Hu, M. D., Zang, Y. D., & Xu, J. H. (2018). Texture image classification algorithm based on FABEMD. *Journal of Xi 'an University of Posts and Telecommunications, 23*(1), 53–58.
Jadoon, M. M., Zhang, Q., Haq, I. U., et al. (2017). Three-class mammogram classification based on descriptive CNN features. *BioMed Research International, 2017*.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neu- ral networks. *Advances in Neural Information Processing Systems, 25*(2), 1097–1105.
Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). Image Net classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90.
Le Cun, Y., Bottou, L., Bengio, Y., et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.
Lévy, D., & Jain, A. (2016). Breast mass classification from mammograms using deep convolutional neural net- works. *CoRR, 16*(12), 542–544.
Liu, A., Ghosh, J., & Martin, C. E. (2007). *Generative oversampling for mining imbalanced datasets.* DMIN.
Li, Z., Wang, Y., Li, X., Li, X., Zhang, Y., Li, J., … Li, J. (2020). An explainable deep learning framework for breast cancer diagnosis. *IEEE Transactions on Medical Imaging, 39*(9), 2668–2678.
Longadge, R., & Dongre, S. (2013). Class imbalance problem in data mining review. *arXiv preprint arXiv:1305, 1707*.
Mohammed, M. A., Alkhateeb, B., Rashid, A. N., et al. (2018). Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers & Electrical Engineering, 70*, 871–882.
Moon, W. K., Lee, Y. W., Ke, H. H., et al. (2020). Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer Methods and Programs in Biomedicine, 190*, Article 105361.
Nothacker, M., Duda, V., Hahn, M., et al. (2009). Early detection of breast cancer: Benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. *BMC Cancer, 9*(1), 1–9.
Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359.
Qiu, D., Cheng, Y., & Wang, X. (2021). Cardiac magnetic resonance images super-resolution via multi-channel residual attention networks. *Computational and Mathematical Methods in Medicine*, 1–8.
Qiu, D., Cheng, Y., & Wang, X. (2022). Dual U-net residual networks for cardiac magnetic resonance images super-resolution. *Computer Methods and Programs in Biomedicine, 218*, Article 106707.
Qiu, D., Zheng, L., Zhu, J., et al. (2021). *Multiple improved residual networks for medical image super-resolution* (pp. 200–208). Future Generation Computer Systems.
Ragab, D. A., Attallah, O., Sharkas, M., et al. (2021). A framework for breast cancer classification using multi-DCNNs. *Computers in Biology and Medicine, 131*, Article 104245.
Rzegedy, C., Liu, W., Jia, Y. Q., et al. (2015). Going deeper with convolu tion[C]. In *Proc of IEEE conference on computer vision and pattern recognition* (pp. 1–9). Piscataway, NJ: IEEE Press.
Sehrawat, D., Sehrawat, A., Jaiswal, D., et al. (2017). Detection and classification of tumor in mammograms using discrete wavelet transform and support vector machine. *International Research Journal of Engineering and Technology (IRJET), 4*(5), 1328–1334.
Setio, A. A. A., Ciompi, F., Litjens, G., et al. (2016). Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging, 35*(5), 1160–1169.
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large scale image recognition. *Computer Science, 23*(6), 1223–1229.
Srivastava, N., Hinton, G., Krizhevsky, A., et al. (2014). Dropout:A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.
Szegedy, C., Liu, W., Jia, Y. Q., et al. (2015). Going deeper with convolutions. *IEEE Computer Society*, 1–9.
Tajbakhsh, N., & Suzuki, K. (2017). Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification. *MTANNs vs. CNNs. Pattern recognition, 63*, 476–486.
Talha, M. (2016). Classification of mammograms for breast cancer detection using fusion of discrete cosine transform and discrete wavelet transform features. *Biomedical Research, 27*(2), 322–327.
Tao, X. J., & Wang, X. (2019). Image classification method based on deep learning algorithm. *Microcomputer application, 35*(3), 40–43.
Wang, X. X., Wang, M. N., Zhang, J. W., et al. (2018). Vehicle identification method based on improved convolutional neural network Le Net-5. *Computer application research, 35*(7), 2215–2218.
Wong, K. K. L., Fortino, G., & Abbott, D. (2020). Deep learning-based cardiovascular image diagnosis: A promising challenge. *Future Generation Computer Systems, 110*, 802–811.
Wong, K. K. L., Sun, Z., Tu, J. Y., Worthley, S. G., Mazumdar, J., & Abbott, D. (2012). Medical image diagnostics based on computer-aided flow analysis using magnetic resonance images. *Computerized Medical Imaging and Graphics, 36*(7), 527–541.
Xia, B. L., He, C. L., Song, B., et al. (2014). Diagnostic value of ultrasound-guided fine needle aspiration cytological examination of thyroid nodule. *Progress of modern general surgery in China, 17*(7), 540–542.
Yang, X., Wu, L., Ye, W., et al. (2020). Deep Learning signature based on staging CT for preoperative prediction of sentinel lymph node metastasis in breast cancer. *Academic Radiology, 27*(9), 8–10.
Zhang, Z., Yang, J., Ho, A., Jiang, W., Logan, J., Wang, X., Brown, P. D., Mcgovern, S. L., Guha-Thakurta, N., Ferguson, S. D., Fave, X., Zhang, L., Mackin, D., Court, L. E., & Li, J. (2018). A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. *European Radiology, 28*(6), 2255–2263.
Zhao, B., Cheng, T., Zhang, X., Wang, J., Zhu, H., Zhao, R., Li, D., Zhang, Z., & Yu, G. (2023). CT synthesis from MR in the pelvic area using Residual Transformer Conditional GAN. *Computerized Medical Imaging and Graphics, 103*, Article 102150.
Zhao, C., Lul, F. C., Du, S. C., Wang, D., & Shao, Y. P. (2023). An earth mover's distance based multivariate generalized likelihood ratio control chart for effective monitoring of 3D point cloud surface. *Computers & Industrial Engineering, 175*, Article 108911.
Zhou, H., Liu, Z., Li, T., Chen, Y., Huang, W., & Zhang, Z. (2022). Classification of precancerous lesions based on fusion of multiple hierarchical features. *Computer Methods and Programs in Biomedicine, 229*, Article 107301.