IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 February 2024, accepted 16 February 2024, date of publication 20 February 2024, date of current version 26 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3367772

RESEARCH ARTICLE

Application of Semi-Supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data

SAI LI[1], (Member, IEEE), PENG KOU[2], MIAO MA[3], HAOYU YANG[4], SHUO HUANG[5], AND ZHENGYI YANG[3]

[1]College of Mechanical and Electrical Engineering, Zaozhuang University, Zaozhuang 277160, China
[2]Liupanshui Planning and Design Institute of Surveying and Mapping, Liupanshui 553000, China
[3]School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia
[4]College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA
[5]Key Laboratory of Infrared System Detection and Imaging Technology, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China

Corresponding author: Sai Li (lisaizzxy@163.com)

This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR202103050458.

ABSTRACT Deep learning has attracted wide attention recently because of its excellent feature representation ability and end-to-end automatic learning method. Especially in clinical medical imaging diagnosis, the semi-supervised deep learning model is favored and widely used because it can make maximum use of a limited number of labeled data and combine it with a large number of unlabeled data to extract more information and knowledge from it. However, the scarcity of medical image data, the vast image size, and the instability of image quality directly affect the model's robustness, generalization, and image classification performance. Therefore, this paper proposes a new semi-supervised learning model, which uses quadratic neurons instead of traditional neurons, aiming to use quadratic convolution instead of the conventional convolution layer to improve the feature extraction ability of the model. In addition, we introduce two Dropout layers and two fully connected layers at the end of the model to enhance the nonlinear fitting ability of the network. Experiments on two large medical public data sets - ISIC 2019 and Retinopathy OCT - show that our method can improve the model's generalization performance and image classification accuracy.

INDEX TERMS Quadratic neuron convolution, convolution neural network, semi-supervised learning, medical image classification.

I. INTRODUCTION

Since Hinton and other researchers introduced the concept of deep learning into machine learning, deep learning has made remarkable achievements in many fields with its powerful performance, wide open-source framework, and tools. For example: image classification and retrieval [1], [2], [3], [4], natural language processing [5], [6], face recognition [7], [8], [9], and object detection [10], [11], [12], etc. These accomplishments have captured widespread attention and found practical applications in scientific research and various industries [8], [12]. With the continuous improvement of deep learning technology, it has become urgent to deal with the challenge of obtaining tagged data in specific fields. This has prompted researchers in related fields to explore algorithms integrating semi-supervised learning and deep learning [13]. As a result, exploring novel methodologies at the intersection of semi-supervised learning and deep learning has become a focal point in the ongoing evolution of machine learning technologies.

As we all know, the performance of the traditional fully supervised medical image classification algorithm is closely related to the number of available labeled medical images. However, some factors, such as the complexity of the

The associate editor coordinating the review of this manuscript and approving it for publication was Turgay Celik.

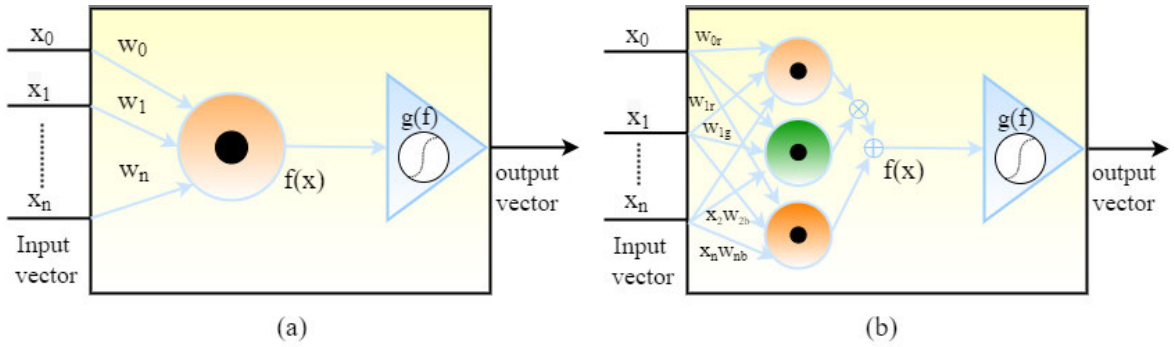


FIGURE 1. (a)The structure of traditional artificial neurons. (b)Quadratic neuron structure is composed of quadratic function and nonlinear excitation function.

labeling process, the time cost of professional doctors, and patients' privacy, limit the amount of data available for fully supervised model training [14]. With the continuous progress of medical technology, the amount of unlabeled medical image data has increased significantly [15]. The rapid growth of this data volume, combined with the rapid development of semi-supervised learning technology, provides a tremendous opportunity to utilize the untapped potential of unlabeled medical images effectively [16]. By combining semi-supervised learning with deep learning technology, researchers can effectively tap the potential of unlabeled medical image data and apply it to various medical image analysis tasks [17]. This progress shows great medical imaging potential and provides new opportunities for more accurate diagnosis, improved treatment strategies, and higher patient care.

Skin cancer is one of the most common cancers known at present, which usually occurs in people who have frequent outdoor activities or are exposed to the sun [18]. In recent years, the semi-supervised learning method, which skillfully uses a limited number of labeled and many unlabeled data, has shown excellent performance in predicting skin lesions. Specifically, Harangi [19] successfully constructed a model for the classification of skin lesions by combining four different deep neural network structures, effectively solving the challenge of scarce labeled data. Xue et al. [20] consider that the noise in the labeled data may interfere with the classifier results. Hence, they designed a sample mining method based on uncertainty to eliminate the noise interference in the image. Santos et al. [21] proposed a semi-automatic approach to cluster the extracted features and achieved excellent results on multiple skin injury data sets. In addition, Narayan et al. [22] presented a real-time deep medical image classification model (Enhance-Net-Net) to meet medical image classification needs better. Zhang et al. [23] put forward a semi-supervised learning framework called BoostMIS, which combines adaptive pseudo-tagging and active annotation of information to release the potential of the medical image SSL model fully. Zhou et al. [24] proposed ReFixMatch-LS by combining consistency regularization and pseudo-labeling and applied it to medical image classification. Zhou et al. [25] presented false label growth threshold (GTPL) and false label loss (PLD) and used them to FixMatch and CoMatch to improve their semi-supervised classification performance effectively.

Retinopathy is rising gradually in modern life, seriously threatening people's quality of life [26]. Using medical imaging technology to predict advanced lesions can prevent the further deterioration of retinopathy in time and effectively. Rasty et al. [27] used a multi-scale convolution mixed expert model to classify retinopathy. Schlegl et al. [28] proposed an automatic diagnosis and treatment method for macular degeneration based on deep learning and achieved the best results on multiple data sets. Rong et al. [29] proposed an alternative auxiliary classification method for automatically classifying retinal OCT images. While Huang et al. [30] improved the variety of macular diseases by the hierarchical guided convolutional neural network. In addition, Sun et al. [31] put forward an automatic recognition framework of maculopathy based on a 2D feature map and attention convolution, which achieved a Good result on public data sets. Arrieta Ramos [32] proposed a semi-supervised method to detect diabetic retinopathy through self-supervised pre-training. Then, they supervised fine-tuning and knowledge distillation with a small group of labeled images. Wang et al. [33] proposed a novel deep semi-supervised multi-instance learning framework, which played a good role in detecting diabetic retinopathy.

However, after careful observation, it is not difficult to find that the existing semi-supervised models are relatively weak in robustness, generalization ability, and classification performance due to the influence of multiple factors such as large medical image size, unbalanced number of various images, and unstable image quality. Therefore, a new semi-supervised medical image classification model is proposed in this study, aiming at coping with and alleviating these problems. Our contributions are as follows:

1) This study improves the DenseNet network and adopts the quadratic convolution method to enhance the model's ability to represent complex data and improve the model's ability to extract features.

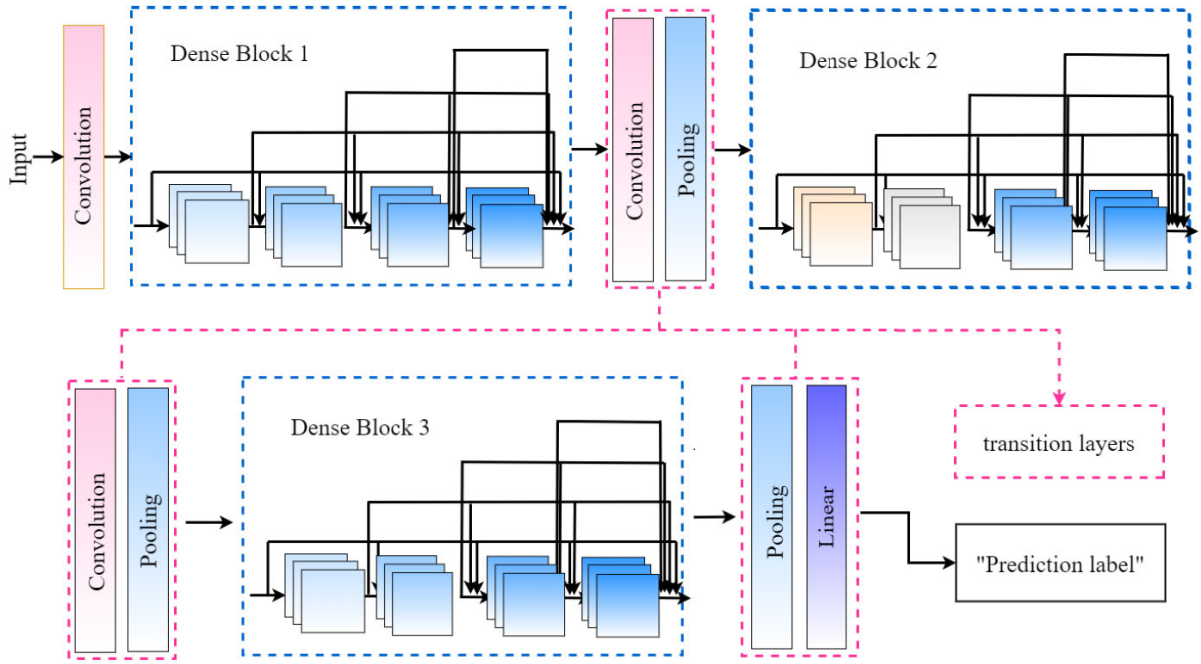


FIGURE 2. Detailed architecture of DenseNet network: Densely connected convolutional neural network.

2) In addition, it is worth noting that two new leakage layers and two fully connected layers are built at the end of the network structure. This design aims to strengthen the nonlinear fitting ability of the network.
3) This paper proposes a novel semi-supervised medical image classification model and compares it with some classical and state-of-the-art semi-supervised image classification models. The results indicate a significant improvement in robustness, generalization capability, and classification accuracy for this model.

II. RELATED WORK

A. QUADRATIC NEURON

As shown in Fig 1. (a), the traditional artificial neuron structure has a linear inner product and nonlinear excitation function [34]. Specifically, given the input vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$, the output $f(x)$ obtained by linear internal product operation is denoted as:

$$f(x) = \sum_{i=1}^{n} w_i x_i + b \tag{1}$$

Then, $f(x)$ is processed by a nonlinear excitation function (sigmoid function). This structure is suitable for separating two groups of linearly separable inputs. However, single neurons are prone to classification errors for linearly inseparable input groups.

Inspired by the diversity of neurons in the biological nervous system, Fan et al. proposed an innovative quadratic neuron [35] for extracting features from one-dimensional bearing fault signals, as illustrated in Fig 1. (b). In this design, the input vector goes through a series of transformation steps before being passed to the nonlinear excitation function, including two inner product operations and the sum of a norm term. Its mathematical expression is as follows:

$$f(x) = \left(\sum_{i=1}^{n} w_{ir} x_i + b_1\right)\left(\sum_{i=1}^{n} w_{ig} x_i + b_2\right) + \sum_{i=1}^{n} w_{ib} x_i^2 + c \tag{2}$$

The design of this new neuron is inspired by the diversity of neurons in the biological nervous system, aiming to simulate and introduce this diversity to improve the model performance. In fault diagnosis, this method is expected to enhance the interpretation and classification of one-dimensional signals [36], especially when dealing with linear inseparable inputs; traditional techniques are prone to classification errors. Therefore, we improve this new neuron and apply it to the two-dimensional convolutional neural network, aiming to bring more excellent classification performance and robustness to the semi-supervised medical image classification model.

B. DENSENET NETWORK

DenseNet(Densely connected convolutional networks) [37], as a convolutional neural network with a deeper structure, has a series of remarkable advantages:

- The number of parameters is relatively tiny: Compared with other convolutional neural networks, DenseNet has fewer parameters, which helps to reduce the complexity of the model and the risk of over-fitting.

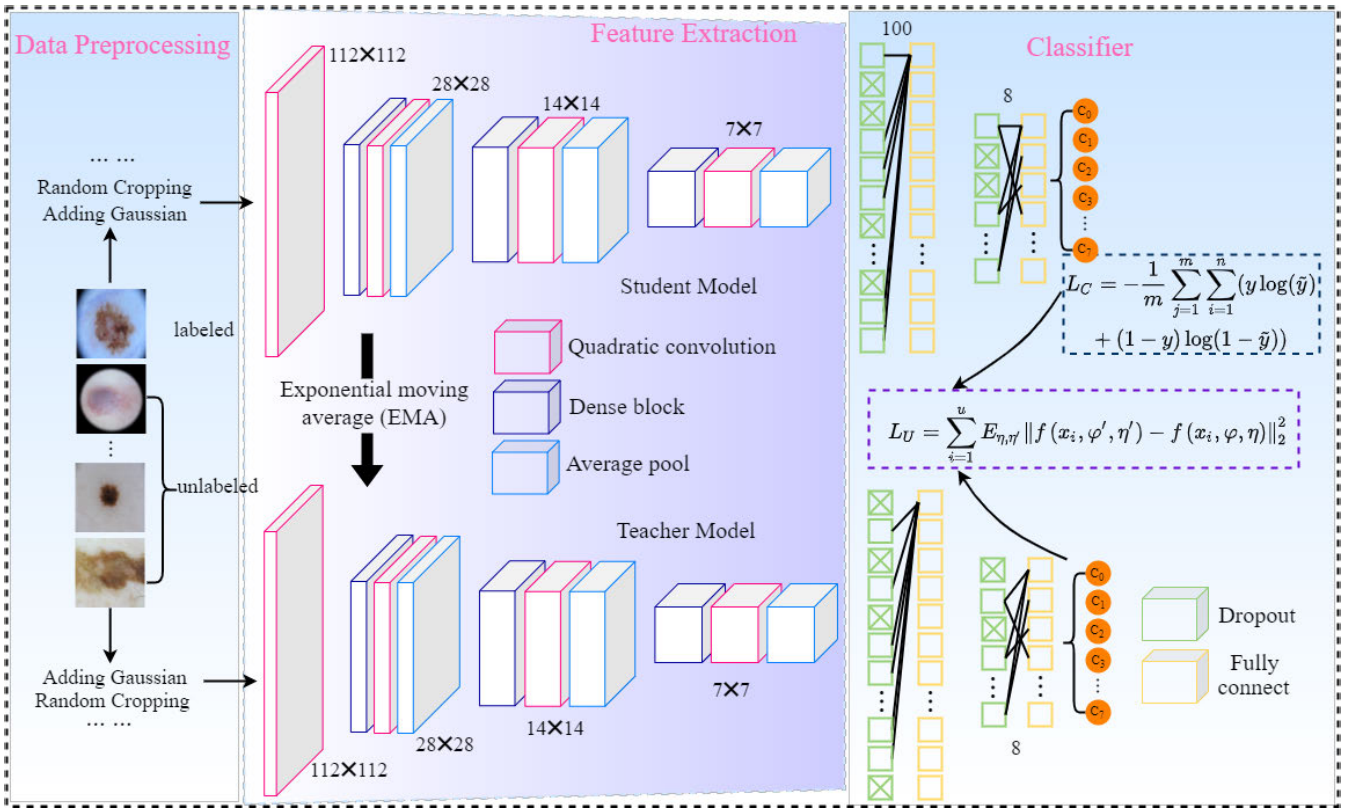


FIGURE 3. The general framework of semi-supervised medical image classification model.

- Bypass connection enhances the reuse of features: DenseNet adopts dense association, which promotes the comprehensive reuse of features and improves the model's efficiency.
- Easy to train and regularization effect: DenseNet's architecture makes training more accessible and has a particular regularization effect.
- Alleviate the problem of gradient disappearance and model degradation: The characteristics of Dense connection help to alleviate the pain of gradient disappearance, make the network more easily trained in the deep layer, and prevent model degradation.

Considering the problems of large image size and unstable image quality in medical image data sets, this paper chooses DenseNet-121 as the backbone network. This choice ensures that the model can effectively handle these large-scale data sets and give full play to the advantages of DenseNet interconnection architecture.

Fig 2 shows a specific DenseNet structure containing three Dense blocks and three transition layers. It is not difficult to see that the layers in the dense block are connected. The input of each layer comes from the characteristic diagram of all the layers in front of it, and the output of each layer is directly connected with the information of all the layers behind it. This dense connection realizes the efficient reuse of features and helps to improve performance. The transition layer includes 1x1 convolution for adjusting the number of channels and 2x2 average pooling for reducing the feature map size. Their main functions are to connect adjacent Dense blocks, integrate the features of the previous Dense block, and simultaneously downsample the feature map, reducing the number of parameters of the model, which is helpful to the efficiency of training.

III. PROPOSED METHOD

A. OVERALL ARCHITECTURE

As shown in Fig 3, the semi-supervised medical image classification model proposed in this paper utilizes DenseNet as its backbone network. Notably, we have introduced quadratic neurons to enhance the convolutional layers in DenseNet, forming quadratic convolutional layers. Additionally, two dropouts and two fully connected layers are incorporated at the network's end. The model consists of two parts: supervised training based on labeled samples and unsupervised training based on unlabeled samples. This unique architecture allows for better utilization of abundant unlabeled real-world data, closer to practical application scenarios, and achieves superior performance in medical image classification tasks.

Specifically, given a medical image dataset $S = S_L \cup S_U$, where S_L and S_U represent manually labeled medical images and unlabeled medical images, respectively, it is noteworthy that the quantity of S_L is significantly smaller than that of S_U.

The supervised training process based on manually labeled samples S_L is as follows:

1) Firstly, the marked image is preprocessed (adding noise or random cutting, etc.) to make it visually different from the original image as much as possible.
2) Next, the preprocessed image is sent to the feature extraction module of the student model to learn the features in the image that are more valuable to the current task.
3) Then, the above features are transmitted to the classifier module and further analyzed to generate pseudo-labels of the input image.
4) The cross-entropy function is used to calculate the difference between the pseudo label generated by the model and the actual label of the image, and the gradient descent algorithm [38] is used to reduce the above difference and update the parameters of the model to classify the medical image more accurately.

Mathematically, the cross-entropy function [39] is defined as follows:

$$L = -[y \log \tilde{y} + (1 - y) \log(1 - \tilde{y})] \tag{3}$$

Here, y represents the actual label of the image, and $\tilde{y}$ represents the probability that the model correctly predicts the category of the picture. Therefore, the supervision loss of supervised training with marked samples can be expressed as:

$$L_C = -\frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{n} (y \log(\tilde{y}) + (1 - y) \log(1 - \tilde{y})) \tag{4}$$

In the above formula, m is the number of labeled samples in the medical image dataset, and n is the total number of categories.

Then, we use the exponential weighted average moving technology (EMA) [40] to update the trained student model parameters and construct the teacher model. This process aims to prepare for the subsequent unsupervised training based on unlabeled samples. The formula of the exponential weighted average moving technique is as follows.

$$\varphi'_t = a\varphi'_{t-1} + (1 - a)\varphi_t \tag{5}$$

In this formula, φ'_t and φ_t represent the parameters of the teacher model and the student model at t, respectively, and a is a superparameter of the smoothing coefficient, which is used to adjust the update rate of weights. The unsupervised training steps based on unlabeled samples are as follows:

1) Firstly, each unlabeled image is preprocessed (adding noise or randomly cutting, etc.) to maximize the difference between two sub-images of the same image.
2) Next, the above two sub-images are input into the student and teacher models to learn the potential features of the two sub-images.
3) Subsequently, the features of the above two sub-images are passed to the classifier module and assigned pseudo-labels.
4) Based on the principle of consistent regularization [41], that is, if a slight disturbance is applied to the same unlabeled image, the prediction result should not change significantly. Finally, the mean square error function [42] measures the difference between the pseudo-labels of the two sub-images, and the gradient descent algorithm updates the parameters of the student and teacher models.

Unsupervised loss of unmarked samples is expressed as:

$$L_U = \sum_{i=1}^{u} E_{\eta,\eta'} \left\| f\left(x_i, \varphi', \eta'\right) - f\left(x_i, \varphi, \eta\right) \right\|_2^2 \tag{6}$$

Here, u represents the total number of unlabeled samples. η and η' represent two different ways of disturbing the same image. $f\left(x_i, \varphi, \eta\right)$ and $f\left(x_i, \varphi', \eta'\right)$, respectively, represent the prediction vectors obtained after processing by the student model and the teacher model. Therefore, the final loss function of the semi-supervised medical image classification model can be expressed as:

$$\text{Loss} = L_C + \lambda L_U \tag{7}$$

In the above formula, L_C and L_U respectively represent the supervised loss of labeled samples in the medical image data set and the unsupervised loss of unmarked pieces in the medical image data set, and λ is a hyperparameter for balancing supervised loss and unsupervised loss.

B. QUADRATIC CONVOLUTION

As shown in Fig 3, this study uses Quadratic neurons instead of traditional ones. This choice means we use quadratic convolution to replace the conventional convolution in the DenseNet structure. This quadratic convolution operation can be divided into two essential parts: autocorrelation inner product operation and standard convolution operation.

The nonlinear mapping of the standard convolution operation of traditional neurons mainly depends on the activation function. However, different from it, the Quadratic neurons use autocorrelation inner product operation to introduce additional nonlinear mapping, which is difficult for traditional neurons to achieve. As we all know, nonlinear mapping can approximate complex functions more effectively than linear mapping. Therefore, the quadratic convolution formed by Quadratic neurons in this paper can significantly enhance DenseNet's ability to represent features, which is not due to the increase in the number of parameters but to the nonlinear calculation involved.

During the training period, this paper adopts a particular parameter initialization strategy [36], in which the parameters of secondary neurons are initialized as $\boldsymbol{w}^g = 0$, $\boldsymbol{b}^g = 1$, $\boldsymbol{w}^b = 0, c = 0$, while $\boldsymbol{w}^r$ and $\boldsymbol{b}^r$ are initialized according to the standard initialization method. Therefore, in the initialization stage, each secondary neuron in the network behaves like a traditional neuron. Subsequently, we adopt the normal learning rate γ_r for parameter $(\boldsymbol{w}^r, b^r)$ and a smaller learning rate $\gamma_{g,b}$ for parameter $\left(\boldsymbol{w}^g, b^g, \boldsymbol{w}^b, c\right)$ in the training process. This strategy helps the network better adapt to the nonlinear characteristics of secondary neurons.

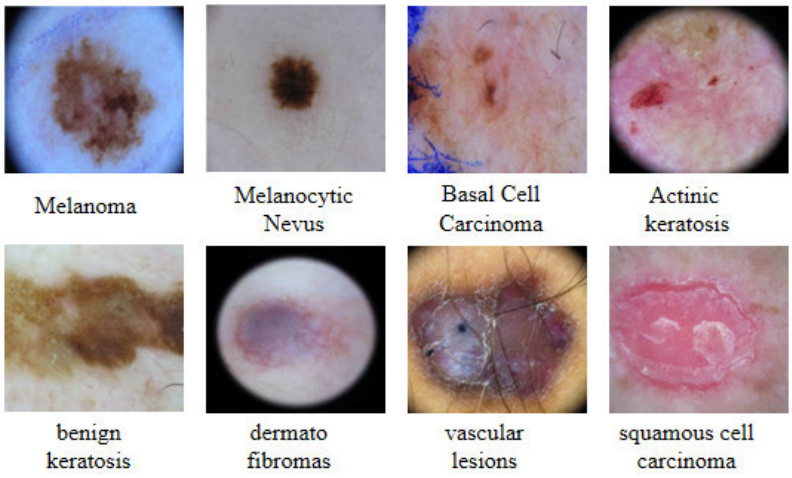


FIGURE 4. ISIC 2019 dataset: Analysis of the distribution of sample samples of various categories.

C. DROPOUT LAYER AND FULLY CONNECTED LAYER

As we all know, the fully connected layer plays a crucial role in the convolutional neural network, which filters the features captured by the feature extraction module and maps these abstract "image distributed feature representations" to the corresponding label space for image classification tasks. A typical fully connected layer has multiple neurons to fit the data distribution. However, a single fully connected layer sometimes makes it difficult to solve the nonlinear problem well. By increasing the number of fully connected layers, we can significantly improve the nonlinear fitting ability of the network [43]. Additionally, dropout layers are typically introduced before adding the fully connected layers. This operation randomly excludes some neurons and their associated connections during the training process, providing the semi-supervised medical image classification model with partially incomplete signals, thereby enhancing the model's generalization performance.

Specifically, whenever a new input element enters the model, the dropout layer randomly selects a subset of neurons and connects these subsets with the shared weights. Specific neurons or connections do not restrict this operation, so it can effectively deal with the problem of over-fitting [44], [45]. Introducing the dropout layer into the model helps improve its generalization performance on various classification problems.

At the end of this paper's semi-supervised medical image classification model, two dropouts and two fully connected layers are introduced. Specifically, the size of the first fully connected layer is 100, and the size of the second fully connected layer is 8. Before these two fully connected layers, the model performs a dropout operation to ensure that some data information is incomplete during training, thus improving the model's generalization ability.

IV. EXPERIMENT

This paper mainly uses the Python programming language and PyTorch deep learning framework to build and train the network model. The CPU is AMD Ryzen 7 5800X, the memory is 64G, and the GPU is an NVIDIA RTX 3090 with 24G. To effectively use the GPU, the environment of this experiment also includes CUDA11.1 and CUDA Deep Neural Network Library (CUDNN).

Additionally, we conducted experiments on two large-scale medical datasets, ISIC 2019 and OCT. We compared our method with traditional classical algorithms [46], [47], [48] and state-of-the-art frontier algorithms [23], [33], [49] in several experimental metrics to accurately evaluate the performance of our approach in medical image classification tasks.

A. EXPERIMENTAL DETAILS AND EVALUATION INDICATORS

Because the sizes of the images in the two medical data sets are not the same, this paper adjusts all the pictures in the data set to 224 × 224 to meet the training requirements of the network. The parameters in model training are set as follows: the total number of training rounds is 200, the batch number is 64, and the learning rate is set from 0.01 and gradually decays to 0.001 in the later stage of training. In addition, the Adam algorithm is used as the choice optimizer to optimize the semi-supervised model.

This study used five widely recognized indicators, including accuracy, sensitivity, specificity, F1-Score, and AUC (area under the curve), to comprehensively evaluate the effectiveness of the semi-supervised medical image classification model. This multi-dimensional evaluation system deeply helps us understand the model's performance in different aspects and provides a more comprehensive evaluation and result analysis for our research.

Accuracy is the percentage of correctly classified samples to the total number of samples, and its formula is as follows.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{8}$$

TABLE 1. Comparison of AUC indicators of eight categories on ISIC 2019 test set: Different semi-supervised models.

Method	Temporal ensembling [42]	Virtual adversarial training [43]	Mean teachers [44]	Dual Students [45]	Boostmis [23]	Automatic diagnosis [29]	Ours
Labeled	20%	20%	20%	20%	20%	20%	20%
Unlabeled	80%	80%	80%	80%	80%	80%	80%
MEL	84.34	83.27	84.57	85.22	85.64	84.69	**86.79**
NV	85.26	85.49	86.08	85.38	85.15	86.42	**86.57**
BCC	84.89	84.79	85.13	84.67	86.37	86.54	**86.88**
AKIEC	82.54	83.26	82.49	83.12	82.51	83.79	83.37
BKL	83.51	84.23	83.57	83.79	84.17	83.21	**85.52**
DF	81.23	82.71	82.89	82.68	82.72	83.26	**83.77**
VASC	80.98	81.26	81.08	82.56	81.25	81.63	**82.96**
SCC	81.34	82.09	82.85	83.23	83.07	82.96	**83.79**
Average AUC	83.01	83.38	83.58	83.83	83.80	84.09	**84.96**

TABLE 2. Accuracy of different semi-supervised models in ISIC 2019 test set.

Method	backbone	epoch	Labeled	Unlabeled	Accuracy
Upper Bound	DenseNet121	200	100%	0%	90.28
Temporal ensembling[42]	DenseNet121	200	20%	80%	84.79
Virtual adversarial training[43]	DenseNet121	200	20%	80%	85.23
Mean teachers[44]	DenseNet121	200	20%	80%	85.02
Dual Students[45]	DenseNet121	200	20%	80%	85.74
Boostmis[23]	DenseNet121	200	20%	80%	85.63
Automatic diagnosis[29]	DenseNet121	200	20%	80%	86.36
Ours	DenseNet121	200	20%	80%	87.07

Sensitivity, also known as recall or actual positive rate, refers to the ratio of correctly classified positive samples to the total number of positive samples in the dataset. In medical image classification studies, it signifies the model's ability to detect positive cases accurately. The formula is as follows:

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

Specificity refers to the ratio of correctly classified negative samples to the total number of negative examples in the data set, which shows the ability of the model to identify negative cases accurately. Generally speaking, the higher the specificity, the lower the misdiagnosis rate of adverse results. The formula is as follows:

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

F1-score is a metric that combines recall and accuracy to evaluate the merit of a model and maximizes both by achieving a balance between the two, as follows:

$$\text{F1-score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{11}$$

The definition of AUC refers to the area enclosed by the receiver operating characteristic curve (ROC) and the coordinate axis. The range of AUC is [0,1], and the higher the value, the better the classification effect of the model. The

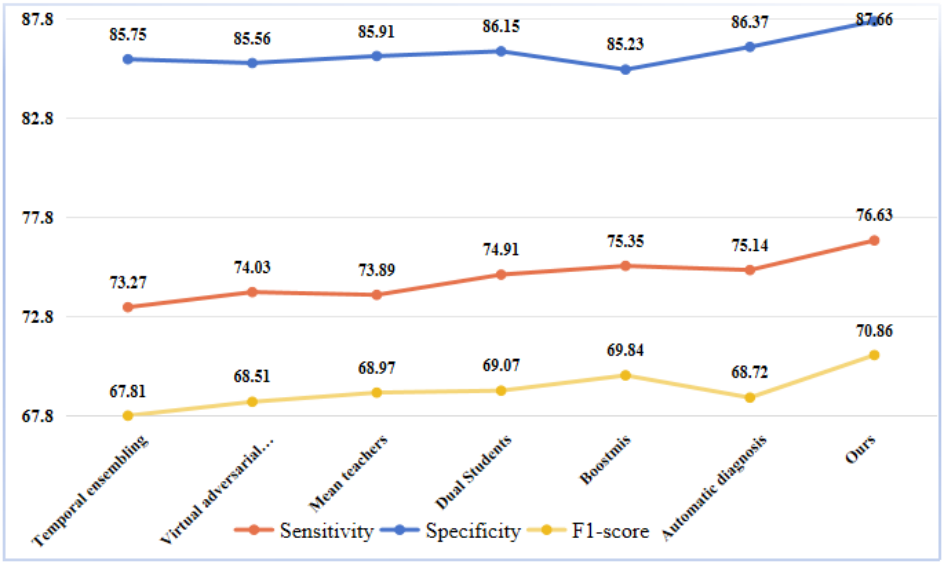


FIGURE 5. Performance comparison of different semi-supervised models on ISIC 2019 test set: Sensitivity, specificity, and F1 value.

TABLE 3. Influence of different percentage of ISIC 2019 labelling data training on semi-supervised medical image classification model.

Method	epoch	Labeled	Unlabeled	AUC	Accuracy
UpperBound	200	100%	0	87.37	90.28
Ours	200	10%	90%	83.85	85.27
Ours	200	20%	80%	84.96	87.07
Ours	200	30%	70%	85.73	87.81
Ours	200	40%	60%	86.16	88.64

formula is as follows:

$$\mathrm{AUC} = \frac{\sum_{ins_i} \mathrm{Rank}_{ins_i} - \frac{M\times(M+1)}{2}}{M \times N} \tag{12}$$

B. ISIC 2019 DATASET

Skin cancer image ISIC2019 data set [50] is divided into eight categories, comprising 25,331 images. Specifically, it includes 4,522 melanoma (MEL) samples, 12,875 melanocyte nevus (NV) samples, 3,323 basal cell carcinoma (BCC) samples, 867 actinic keratosis/Bowen's disease (AKIEC) samples, 2,624 benign keratosis (BKL) samples, and 200 samples. Fig 4 shows each category image in the ISIC2019 dataset.

Table 1 shows the AUC values of each category of different semi-supervised image classification methods on the ISIC 2019 test set when 20% labelled data and 80% unlabeled data are used for training. The results show that our model performs well on this test set, and the average AUC value is 84.96%, which is superior to other semi-supervised image classification methods. More specifically, our model has

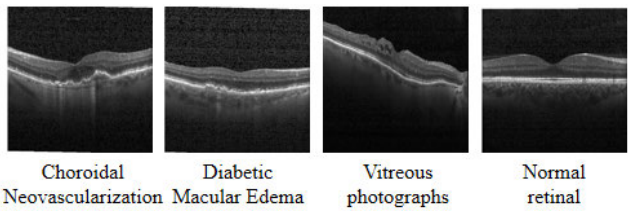


FIGURE 6. Actual samples from each category in the retinopathy OCT dataset.

achieved higher results than other semi-supervised models in seven categories when comparing the AUC values of various types. This result fully proves the excellent performance of our model in the semi-supervised skin cancer image classification task.

According to the results in Table 2, when training with 20% marked data and 80% unlabeled data, our method achieves excellent accuracy, reaching 87.07%. This performance exceeds the existing semi-supervised image classification model. Specifically, our approach improves the accuracy by 2% to 3% compared with some classical semi-supervised classification models, and it is also 0.7% to 1.4% higher than some cutting-edge semi-supervised classification methods. These results strongly prove the outstanding performance of our proposed method in the field of semi-supervised image classification.

To verify the performance of the semi-supervised medical image classification model proposed in this paper more comprehensively, we compare different semi-supervised medical image classification models with three key indicators: specificity, sensitivity, and F1 value, and the specific results are shown in Fig 5. Encouragingly, our method offers a noticeable improvement in specificity, sensitivity, and F1

TABLE 4. Comparison of AUC indicators of four categories on the retinopathy OCT test set: Different semi-supervised models.

Method	Temporal ensembling [42]	Virtual adversarial training [43]	Mean teachers [44]	Dual Students [45]	Boostmis [23]	Automatic diagnosis [29]	Ours
Labeled	20%	20%	20%	20%	20%	20%	20%
Unlabeled	80%	80%	80%	80%	80%	80%	80%
CNV	91.73	91.93	93.26	91.57	92.43	93.17	**93.74**
DME	91.08	91.24	90.37	91.99	92.14	91.86	**92.57**
DRUSEN	89.72	91.19	89.89	90.83	91.65	91.87	**92.35**
NORMAL	92.34	92.08	91.39	91.93	92.94	92.75	92.42
Average AUC	91.21	91.53	91.23	91.58	92.29	92.41	**93.27**

TABLE 5. Accuracy of different semi-supervised models in the retinopathy OCT test set.

Method	backbone	epoch	Labeled	Unlabeled	Accuracy
Upper Bound	DenseNet121	200	100%	0%	95.21
Temporal ensembling[42]	DenseNet121	200	20%	80%	91.12
Virtual adversarial training[43]	DenseNet121	200	20%	80%	91.35
Mean teachers[44]	DenseNet121	200	20%	80%	91.27
Dual Students[45]	DenseNet121	200	20%	80%	91.24
Boostmis[23]	DenseNet121	200	20%	80%	92.43
Automatic diagnosis[29]	DenseNet121	200	20%	80%	91.87
Ours	DenseNet121	200	20%	80%	92.85

value. Compared with other methods, the improvement range is 1% to 3%. These data fully prove the superiority of our approach.

It is worth noting that the performance of the semi-supervised image classification model is closely related to the proportion of labeled samples used. Therefore, we have also profoundly studied the influence of different labeled data proportions on the model performance, and the specific results are shown in Table 3. Obviously, with the increase in the proportion of tag data, we observed that the accuracy and AUC value of the network on the ISIC 2019 test set gradually improved. In particular, our network can achieve an accuracy of 85.27% even if trained with only 10% labeled data.

TABLE 6. Influence of different percentage of retinopathy OCT labelling data training on semi-supervised medical image classification model.

Method	epoch	Labeled	Unlabeled	AUC	Accuracy
UpperBound	200	100%	0%	95.25	95.21
Ours	200	10%	90%	91.79	92.05
Ours	200	20%	80%	93.27	92.85
Ours	200	30%	70%	93.56	93.18
Ours	200	40%	60%	93.82	93.57

C. RETINOPATHY OCT DATASET

The OCT dataset [51] of retinopathy contains a total of 108,312 images, which are categorized into four groups: 37,206 Choroidal Neovascularization (CNV) images, 8,617 vitreous photographs (Drusen, DRUSEN), 11,349 Diabetic Macular Edema (DME) images, and 51,140 Normal retinal images (NORMAL). Fig 6 displays specific examples of pathological pictures for each category.

Firstly, this paper compares the AUC values of different semi-supervised models in each category on the Retinopathy OCT Dataset, and the results are shown in Table 4. Specifically, this paper uses 20% labeled samples and 80% unlabeled samples to train different semi-supervised models.

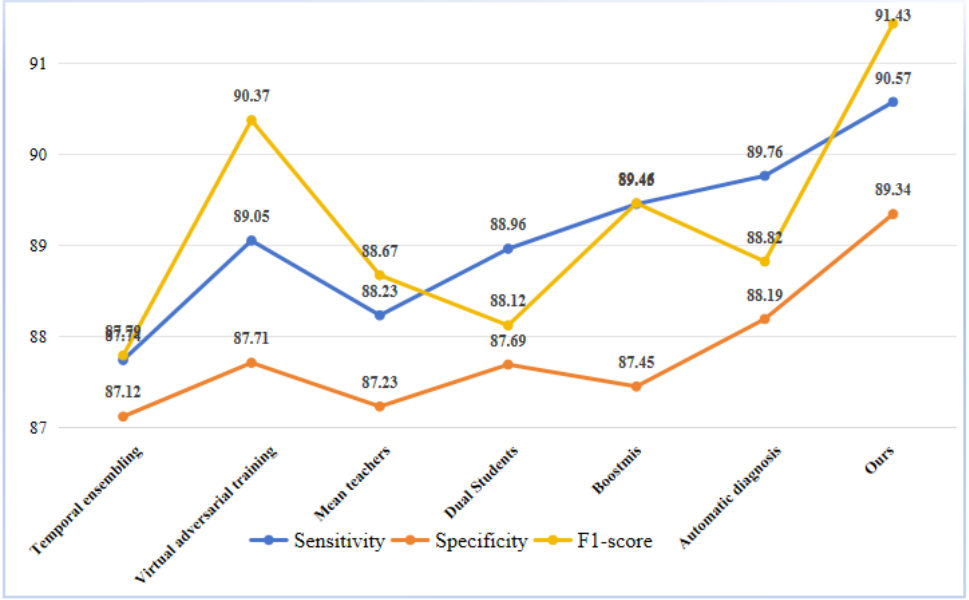


FIGURE 7. Performance comparison of different semi-supervised models on the retinopathy OCT test set: Sensitivity, specificity, and F1 value.

The AUC values of the model in this paper are higher than those of other semi-supervised models in three categories, and the best average AUC is obtained. This result proves that our network has good robustness and generalization ability.

Secondly, this paper uses 20% labeled samples and 80% unlabeled samples in the Retinopathy OCT Dataset to train different semi-supervised image classification models and compares their accuracy scores, as shown in Table 5. It is not difficult to see that the method in this paper has achieved the highest accuracy rate of 92.85%, which is improved compared with other semi-supervised models, which emphasizes the advantages of this model in accurately and effectively classifying various retinal OCT images.

Then, to further demonstrate the advantages of our model, this paper also uses 20% labeled samples and 80% unlabeled samples in the Retinopathy OCT Dataset to train different semi-supervised image classification models and makes a comprehensive comparison on three indicators, namely specificity, sensitivity and F1 score, as shown in Fig 7. This model's specificity, sensitivity, and F1 scores are 90.57%, 89.34%, and 91.43%, respectively, significantly improved compared to the previous models. These results again confirm our model's effectiveness and superiority in all aspects of medical image classification.

Finally, as shown in Table 6, we investigated the performance impact of using different ratios of labeled samples to train the semi-supervised medical image classification model. Its accuracy and AUC values increase with the percentage of labeled samples in the training samples. Notably, the model in this paper achieves an AUC value of 91.79% and an accuracy of 92.05% under the combination of using 10% labeled samples and 90% unlabeled image data. With limited labeled samples, the model in this paper can still achieve significantly better accuracy and AUC values.

V. DISCUSSION

Nowadays, semi-supervised deep learning is increasingly used in clinical medical imaging diagnosis. However, in the medical field, existing models have poor robustness, generalization ability, and classification performance due to problems such as lack of image data, large image size, and low image quality. Therefore, this paper proposes a new semi-supervised medical image classification model and conducts many experiments on two major public medical data sets, ISIC 2019 and retinopathy OCT. By comparing with some classic and cutting-edge semi-supervised algorithms, our method has improved to varying degrees in five experimental indicators, including accuracy, sensitivity, specificity, F1-Score, and AUC, proving that our method can significantly improve the model's generalization performance and image classification accuracy. However, our method still has some limitations:

1) While introducing quadratic neurons and additional layers enhances the model's performance, it also creates a more complex structure, potentially resulting in higher computational costs and longer training times.
2) Although this paper provides a detailed explanation of the model's architecture and principles, there is still a need for a deeper understanding of its internal workings and mechanisms. This understanding is crucial for fine-tuning and optimizing the model and explaining its prediction process.
3) This paper primarily validates the model using publicly available international datasets. However, testing the model on more diverse and specific datasets in

real-world medical applications will contribute to a more comprehensive understanding of its performance and adaptability.

4) Rare disease datasets are often smaller and exhibit noticeable class imbalance. However, this paper does not delve into the model's performance on datasets related to rare diseases.

VI. CONCLUSION

With the continuous improvement of computer performance and the continuous improvement of deep learning technology, semi-supervised image classification technology is increasingly used in the field of clinical medical image classification. However, the existing semi-supervised image classification model has some shortcomings in robustness, generalization ability, and accuracy due to excessive medical image size, uneven distribution of various types of images, and low image quality. This paper proposes a brand-new semi-supervised learning model to solve the above problems. In short, this paper improves the DenseNet network by introducing secondary neurons to enhance the feature extraction ability of the model. In addition, at the end of the network structure, we added two Dropout layers and two fully connected layers to enhance the nonlinear fitting ability of the network. In this paper, experiments are carried out on two large medical public data sets, and the results are compared with some classic and cutting-edge semi-supervised image classification models. The results show that the semi-supervised model proposed in this paper has improved robustness, generalization ability, and classification accuracy.

Our future research work will focus on the following aspects:

- Optimizing Model Complexity: The model can be optimized by employing techniques such as model pruning to reduce complexity without compromising performance.
- Enhancing Model Interpretability: Delve deeper into the internal workings of semi-supervised learning models, striving to improve their interpretability for a more thorough optimization and understanding of their prediction mechanisms.
- Expanding Testing to Different Datasets: Test and validate the model on a more diverse dataset, particularly those related to specific medical conditions or representing more comprehensive demographic profiles. This is crucial for genuinely assessing its generalizability and applicability in real-world scenarios.
- Investigating Model Performance on Rare Disease Data: Conduct in-depth research on the model's performance on rare disease datasets and explore transfer learning strategies to leverage limited data effectively.
- Continuous Parameter Optimization: Refine model parameters, including learning rates and initialization strategies, to adapt to different medical imaging tasks and improve overall performance.

ACKNOWLEDGMENT

(Sai Li and Peng Kou are co-first authors.)

REFERENCES

[1] F. Radenović, G. Tolias, and O. Chum, "CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples," in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Springer, 2016, pp. 3–20.

[2] C. Affonso, A. L. D. Rossi, F. H. A. Vieira, and A. C. P. de Leon Ferreira, "Deep learning for biological image classification," *Expert Syst. Appl.*, vol. 85, pp. 114–122, Nov. 2017.

[3] K. B. Obaid, S. Zeebaree, and O. M. Ahmed, "Deep learning models based on image classification: A review," *Int. J. Sci. Bus.*, vol. 4, no. 11, pp. 75–81, 2020.

[4] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit. Lett.*, vol. 141, pp. 61–67, Jan. 2021.

[5] A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavaf, and E. A. Fox, "Natural language processing advancements by deep learning: A survey," 2020, *arXiv:2003.01200*.

[6] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[7] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102805.

[8] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales, "When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 384–392.

[9] Md. T. H. Fuad, A. A. Fime, D. Sikder, M. A. R. Iftee, J. Rabbi, M. S. Al-Rakhami, A. Gumaei, O. Sen, M. Fuad, and M. N. Islam, "Recent advances in deep learning techniques for face recognition," *IEEE Access*, vol. 9, pp. 99112–99142, 2021.

[10] Y. Liu, H. Li, J. Yan, F. Wei, X. Wang, and X. Tang, "Recurrent scale approximation for object detection in CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 571–579.

[11] M. Zhang, H. Li, G. Xia, W. Zhao, S. Ren, and C. Wang, "Research on the application of deep learning target detection of engineering vehicles in the patrol and inspection for military optical cable lines by UAV," in *Proc. 11th Int. Symp. Comput. Intell. Design (ISCID)*, vol. 1, Dec. 2018, pp. 97–101.

[12] J. Ryu and S. Kim, "Small infrared target detection by data-driven proposal and deep learning-based classification," *Proc. SPIE*, vol. 10624, pp. 134–143, May 2018.

[13] J. J. Jeong, A. Tariq, T. Adejumo, H. Trivedi, J. W. Gichoya, and I. Banerjee, "Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation," *J. Digit. Imag.*, vol. 35, no. 2, pp. 137–152, Apr. 2022.

[14] V. Narayan, P. K. Mall, S. Awasthi, S. Srivastava, and A. Gupta, "FuzzyNet: Medical image classification based on GLCM texture feature," in *Proc. Int. Conf. Artif. Intell. Smart Commun. (AISC)*, Jan. 2023, pp. 769–773.

[15] T. Huynh, A. Nibali, and Z. He, "Semi-supervised learning for medical image classification using imbalanced training data," *Comput. Methods Programs Biomed.*, vol. 216, Apr. 2022, Art. no. 106628.

[16] P. Aggarwal, N. K. Mishra, B. Fatimah, P. Singh, A. Gupta, and S. D. Joshi, "COVID-19 image classification using deep learning: Advances, challenges and opportunities," *Comput. Biol. Med.*, vol. 144, May 2022, Art. no. 105350.

[17] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: A literature review," *BMC Med. Imag.*, vol. 22, no. 1, p. 69, Dec. 2022.

[18] N. Hasan, A. Nadaf, M. Imran, U. Jiba, A. Sheikh, W. H. Almalki, S. S. Almujri, Y. H. Mohammed, P. Kesharwani, and F. J. Ahmad, "Skin cancer: Understanding the journey of transformation from conventional to advanced treatment approaches," *Mol. Cancer*, vol. 22, no. 1, p. 168, Oct. 2023.

[19] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J. Biomed. Informat.*, vol. 86, pp. 25–32, Oct. 2018.

[20] C. Xue, Q. Dou, X. Shi, H. Chen, and P.-A. Heng, "Robust learning at noisy labeled medical images: Applied to skin lesion classification," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1280–1283.

[21] E. Santos, R. Veras, H. Miguel, K. Aires, M. L. Claro, and G. B. Junior, "A skin lesion semi-supervised segmentation method," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jul. 2020, pp. 33–38.

[22] V. Narayan, P. K. Mall, A. Alkhayyat, K. Abhishek, S. Kumar, and P. Pandey, "Enhance-Net: An approach to boost the performance of deep learning model based on real-time medical images," *J. Sensors*, vol. 2023, pp. 1–15, May 2023.

[23] W. Zhang, L. Zhu, J. Hallinan, S. Zhang, A. Makmur, Q. Cai, and B. C. Ooi, "BoostMIS: Boosting medical image semi-supervised learning with adaptive pseudo labeling and informative active annotation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 20634–20644.

[24] S. Zhou, S. Tian, L. Yu, W. Wu, D. Zhang, Z. Peng, and Z. Zhou, "ReFixMatch-LS: Reusing pseudo-labels for semi-supervised skin lesion classification," *Med. Biol. Eng. Comput.*, vol. 61, no. 5, pp. 1033–1045, May 2023.

[25] S. Zhou, S. Tian, L. Yu, W. Wu, D. Zhang, Z. Peng, and Z. Zhou, "Growth threshold for pseudo labeling and pseudo label dropout for semi-supervised medical image classification," *Eng. Appl. Artif. Intell.*, vol. 130, Apr. 2024, Art. no. 107777.

[26] B. Sheng, X. Chen, T. Li, T. Ma, Y. Yang, L. Bi, and X. Zhang, "An overview of artificial intelligence in diabetic retinopathy and other ocular diseases," *Frontiers Public Health*, vol. 10, Oct. 2022, Art. no. 971943.

[27] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT classification using a multi-scale convolutional neural network ensemble," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 1024–1034, Apr. 2018.

[28] T. Schlegl, S. M. Waldstein, H. Bogunovic, F. Endstraßer, A. Sadeghipour, A.-M. Philip, D. Podkowinski, B. S. Gerendas, G. Langs, and U. Schmidt-Erfurth, "Fully automated detection and quantification of macular fluid in OCT using deep learning," *Ophthalmology*, vol. 125, no. 4, pp. 549–558, Apr. 2018.

[29] Y. Rong, D. Xiang, W. Zhu, K. Yu, F. Shi, Z. Fan, and X. Chen, "Surrogate-assisted retinal OCT image classification based on convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 253–263, Jan. 2019.

[30] L. Huang, X. He, L. Fang, H. Rabbani, and X. Chen, "Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1026–1030, Jul. 2019.

[31] Y. Sun, H. Zhang, and X. Yao, "Automatic diagnosis of macular diseases from OCT volume based on its two-dimensional feature map and convolutional neural network with attention mechanism," *J. Biomed. Opt.*, vol. 25, no. 9, Sep. 2020, Art. no. 096004.

[32] J. M. A. Ramos, "Semi-supervised deep learning for ocular image classification," Ph.D. dissertation, Universidad Nacional de Colombia, Bogotá, Colombia, 2022.

[33] X. Wang, F. Tang, H. Chen, C. Y. Cheung, and P.-A. Heng, "Deep semi-supervised multiple instance learning with self-correction for DME classification from OCT images," *Med. Image Anal.*, vol. 83, Jan. 2023, Art. no. 102673.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[35] F. Fan, W. Cong, and G. Wang, "A new type of neurons for machine learning," *Int. J. Numer. Methods Biomed. Eng.*, vol. 34, no. 2, p. e2920, Feb. 2018.

[36] J.-X. Liao, H.-C. Dong, Z.-Q. Sun, J. Sun, S. Zhang, and F.-L. Fan, "Attention-embedded quadratic network (qttention) for effective and interpretable bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[37] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[38] S. H. Haji and A. M. Abdulazeez, "Comparison of optimization techniques based on gradient descent algorithm: A review," *PalArch's J. Archaeol. Egypt/Egyptol.*, vol. 18, no. 4, pp. 2715–2743, 2021.

[39] A. Jamin and A. Humeau-Heurtier, "(Multiscale) cross-entropy methods: A review," *Entropy*, vol. 22, no. 1, p. 45, Dec. 2019.

[40] S. Sukparungsee, Y. Areepong, and R. Taboran, "Exponentially weighted moving average—Moving average charts for monitoring the process mean," *PLoS One*, vol. 15, no. 2, Feb. 2020, Art. no. e0228208.

[41] Y. Cui, D. Wu, and J. Huang, "Optimize TSK fuzzy systems for classification problems: Minibatch gradient descent with uniform regularization and batch normalization," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 12, pp. 3065–3075, Dec. 2020.

[42] M. Ćalasan, S. H. E. A. Aleem, and A. F. Zobaa, "On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function," *Energy Convers. Manage.*, vol. 210, Apr. 2020, Art. no. 112716.

[43] J. Zhang, Y. Sun, L. Guo, H. Gao, X. Hong, and H. Song, "A new bearing fault diagnosis method based on modified convolutional neural networks," *Chin. J. Aeronaut.*, vol. 33, no. 2, pp. 439–447, Feb. 2020.

[44] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, Apr. 2017.

[45] L. Tian, G. Sihai, and L. Hongjie, "Research on fault diagnosis method of aluminum electrolytic cell based on feature extraction," in *Proc. 32nd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, May 2017, pp. 127–131.

[46] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," 2016, *arXiv:1610.02242*.

[47] T. Miyato, S.-I. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: A regularization method for supervised and semi-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1979–1993, Aug. 2019.

[48] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[49] Z. Ke, D. Wang, Q. Yan, J. Ren, and R. Lau, "Dual student: Breaking the limits of the teacher in semi-supervised learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6728–6736.

[50] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020.

[51] L. Carrera-Escalé, A. Benali, A.-C. Rathert, R. Martín-Pinardel, C. Bernal-Morales, A. Alé-Chilet, M. Barraso, S. Marín-Martinez, S. Feu-Basilio, J. Rosinés-Fonoll, T. Hernandez, I. Vilá, R. Castro-Dominguez, C. Oliva, I. Vinagre, E. Ortega, M. Gimenez, A. Vellido, E. Romero, and J. Zarranz-Ventura, "Radiomics-based assessment of OCT angiography images for diabetic retinopathy diagnosis," *Ophthalmol. Sci.*, vol. 3, no. 2, Jun. 2023, Art. no. 100259.

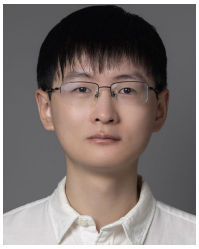
SAI LI (Member, IEEE) received the Ph.D. degree in physics and electronics from Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, in 2020. He is currently with Zaozhuang University. His research interests include computer vision, machine learning, pattern recognition, hyperspectral imagery, and image processing.

PENG KOU received the B.E. degree from the College of Surveying and Mapping Engineering, Shandong University of Science and Technology, Shandong, China, in 2014. He is currently with Liupanshui Planning and Design Institute of Surveying and Mapping, Liupanshui, China. His research interests include computer vision, machine learning, pattern recognition, and image processing.

MIAO MA is currently pursuing the M.Phil. degree in computer science with the University of New South Wales (UNSW). Prior to this, she gained experience as a Database System Developer in multiple IT companies. Her research interests include database systems and data mining.

HAOYU YANG received the B.S. degree from the School of Gifted Young, University of Science and Technology of China, in 2023. He is currently pursuing the M.S. degree with the College of Computing, Georgia Institute of Technology. He has cooperated with the Researcher Feilong Ma and an Associate Professor Weiwei Zhuang. His research interests include digital image processing, computer vision, and autonomous driving.

SHUO HUANG received the Ph.D. degree in physics and electronics from Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, in 2020. He is currently conducting postdoctoral research with Shanghai Institute of Technical Physics, Chinese Academy of Sciences. His research interests include computer vision, machine learning, pattern recognition, and image processing.



ZHENGYI YANG received the Ph.D. degree in computer science from the School of Computer Science and Engineering, University of New South Wales (UNSW). He is currently an Associate Lecturer and the Ph.D. Supervisor with the School of Computer Science and Engineering, UNSW. His expertise lies in developing efficient and scalable algorithms and systems for managing and processing large-scale data, including graph data, relational data, spatial–temporal data, and high-dimensional data.