IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 2, 2020, accepted September 26, 2020, date of publication September 29, 2020, date of current version October 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027794

An Approach to Detecting Diabetic Retinopathy Based on Integrated Shallow Convolutional Neural Networks

WANGHU CHEN[1], **BO YANG**[1], **JING LI**[1], **AND JIANWU WANG**[2]
[1]Institute of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China
[2]Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Corresponding author: Wanghu Chen (chenwh@nwnu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61967013 and Grant 61462076.

ABSTRACT The early detection of Diabetic Retinopathy (DR) is critical for diabetics to lower the blindness risks. Many studies represent that Deep Convolutional Neural Network (CNN) based approaches are effective to enable automatic DR detection through classifying retinal images of patients. Such approaches usually depend on a very large dataset composed of retinal images with predefined classification labels to support their CNN training. However, in some occasions, it is not so easy to get enough well-labelled images to act as model training samples. At the same time, when a CNN becomes deeper, its training will not only take much longer time, but also be more likely to lead to overfitting, especially on a large training dataset. Therefore, it is meaningful to explore a simpler CNN based approach that is still effective on small datasets to classify retinal images. In this paper, an approach to retinal image classification is proposed based on the integration of multi-scale shallow CNNs. Experiments on public datasets show that, on small datasets, the proposed approach can improve the classification accuracy by 3% compared with current representative integrated CNN learning approaches. On the bigger dataset, the proposed approach can improve the classification accuracy by 3% to 9% compared with other representative approaches such as traditional CNN, LCNN and VGG16noFC. The evaluation also represents that, though the classification accuracy of the proposed approach declines by 6% on the smallest dataset containing only 10% samples of the original dataset, its time cost declines to about 30% of that on the original dataset.

INDEX TERMS Convolutional neural network, diabetic retinopathy, image classification, integrated learning, performance integration.

I. INTRODUCTION

Diabetic Retinopathy (DR) is an eye condition that can cause vision loss and blindness in people who are suffering from diabetes [1]. The survey of International Diabetes Federation shows that the incidence of DR is increasing year by year, and there will be more than 6 hundred million patients suffering from DR in 2045 [2]. It is considered that DR has ranked in the first place in diseases threatening eye health of human beings. So, early diagnosis and treatment are very important for people suffering from DR to prevent blindness happening. It is generally believed that early diagnosis and treatment can lower the blindness rate of DR patients to 5% [3].

Traditional measures to diagnosing DR mainly depend on doctors to differentiate 2D color images of eye fundus manually. In such conditions, the diagnosis precision greatly relies on the abilities and experiences of doctors. Considering the patients suffering from DR are a very huge population, it is meaningful to explore the automatic classification of retinopathy images for the diagnosis of DR.

In the past decades, many studies have paid attention to the classification of retinal or other medical images. Conventional machine learning based approaches usually need to extract features offline beforehand, such as those based on SVM, KNN, Regression and so on [4]. In recent years, due to the explosive growth of image data on the Internet, deep learning has been widely used in image classification [5], [6]. For example, deep CNN based approaches to detecting DR are proposed in [7], [8] and [9]. To get

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan-Li Sun.

high precision, such approaches usually use very deep networks and very large scale of datasets. However, in the real world, it is often not so easy to get high-quality labelled datasets for model training. It is also well known that manual labelling of a dataset is a very hard and error-prone job. If mistakes happen in dataset labelling, the accuracy of the model learning from the labelled dataset will be affected greatly. At the same time, the training of a very deep neural network often takes a long time or needs special computing environments. Moreover, it is considered that when a CNN becomes deeper and there are not enough data samples, it is easier to cause the model overfitting [10], [11].

Therefore, it is interesting to explore whether shallow neural networks can also get good performance on the classification of medical images, even on a labelled dataset not so big. In this paper, focusing on the scenario of lacking high-quality labelled training samples, we intend to explore the classification of retinal images based on the integrated multi-scale shallow CNNs.

The expected contributions include: (1) Multi-scale shallow CNNs is introduced to the detection of Diabetic Retinopathy based on the classification of retinal images. Because these shallow CNNs can sense image features under various vision-related receptive fields, the approach can improve the classification of images when there are not enough high-quality labelled samples. (2) A performance integration model of base learners, each of which is a shallow CNN, is proposed. It really shows advantages in classification accuracy compared with similar approaches according to the experiments. (3) A policy of sample distribution for multi-scale base learners based on 2-D entropy of images are also explored in the paper.

The remaining sections are arranged as follows. In section II, the related work is analysed. Our proposed shallow CNNs based approach to integrated classification of Diabetic Retinopathy images will be explored in III. In section IV, the processing of the images used in the paper is introduced. Then, experiments and analyses are discussed in section V, and conclusions and future work are given in section VI.

II. RELATED WORK

Lots of studies have explored the detection of diseases by the classification of medical images in the past decade, and many interesting approaches based on various image processing algorithms have been proposed. As to the early detection of DR, most proposed approaches are based on conventional machine learning [4]. Generally, these methods use some feature extraction algorithms to extract the pathological features in the retinal images at first, and then put the pathological features into various classification algorithms to do retinal image classification, such as Support Vector Machines (SVMs), k-Nearest Neighbour (k-NN), regression and so on. Because these methods need to define or extract features of medical images beforehand, their performances of classification highly depend on the policies to the extraction of image features [4].

Feature extraction is essential for the classification of images. The gray-level histogram was used to extract features of images in early studies and was widely applied into image retrieval for its simplicity and high efficiency [12]. In [13], texture feature extraction approach based on local binary patterns is proposed, which is very simple and can effectively extract the features in the image. It is considered that the operator SIFT proposed by Lowe marks the beginning of a new era of local feature extraction [14]. Its extension to Principal Factor Analysis, PCA-SIFT [15], can reduce the complexity and time consumption of feature extraction because pathological position information is added to the feature space to improve the classification performance [16]. These approaches are really effective for the feature extraction of gray-level medical images. However, because gray-level medical images are usually too smooth and lack of obvious edges, some valuable features might be lost. Therefore, convolutional network based end-to-end deep learning systems for feature extraction of medical records are proposed [17]. On a very large scale of image dataset, these approaches show big advantages in feature extraction. However, in some occasions, it is difficult to provide enough well-labelled training samples. In such condition, the effect of the feature extraction of medical images will be affected greatly.

Deep Learning has been widely used in the classification of medical images in recent years. A survey on deep learning based algorithms for the detection of DR is given in [4]. In order to accurately utilise and organise a large amount of Internet image data, a large image library based on the WordNet structure is introduced in [5], which provides a good research condition for researchers in the field of computer vision and other related fields. A deep convolutional neural network is proposed in [18] to classify 1.2 million images and shows good effect. In [19], Human ECG signals are converted into binary images and delivered to a CNN to do recognitions. The approach provides a new perspective for the application of CNNs and shows good performance in signal recognition. Moreover, compared with fully connected architectures, a CNN computationally has fewer connections [20]. So, its training is easier and faster. To prevent the training of a CNN from overfitting, dropout [21] was introduced to address the internal covariate shift, and batch normalisation was added to the network designing as well. This can improve the performance of a CNN to some extent [22]. On the other way, some studies intend to reduce the training time of CNNs by using GPUs (Graphics Processor Units) [23]. In [24], a novel CNN model is proposed, which is composed of the convolutional, subsampling, max pooling, and batch normalisation layers. In practical applications, it is really not so easy to provide enough well-labelled data for model learning. In such situations, the training of a deep neural network often has to face some problems, especially overfitting [10], [11]. According to some studies to the topological conjugation of hidden layers of multilayer neural networks, it is proved that a deep neural network can be replaced by a shallow one with fewer layers [25]. A shallow neural network model for

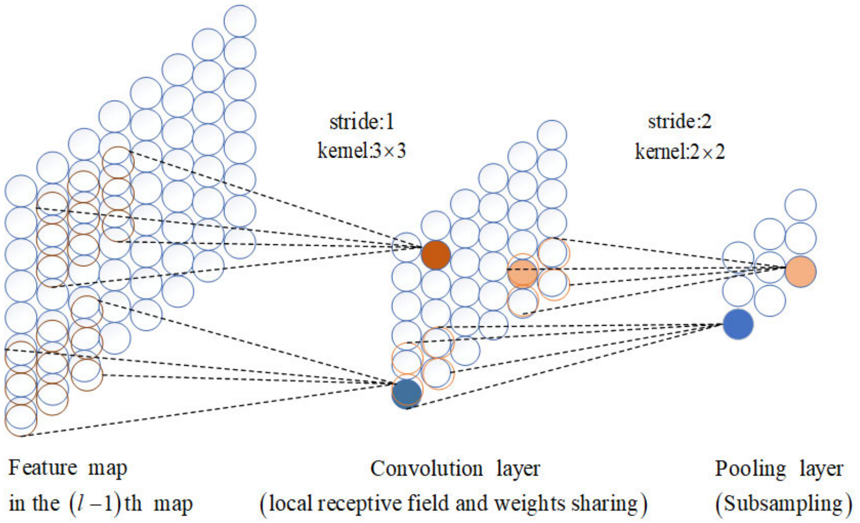

FIGURE 1. A graphical illustration of three key mechanisms (i.e., local receptive field, weights sharing, and subsampling) in convolutional neural networks.

graphical symbol recognition usually needs fewer parameters and is easier to finish model learning. More importantly, it may work well on unbalanced small datasets as well [26]. In [27], focusing on network intrusion detection, a shallow neural network is proposed and compared with a deep one. It is found that the shallow one shows some advantages in both classification effect and time consumption [27]. A novel speech emotion recognition approach is proposed in [28] based on the combination of deep and shallow neural networks, which takes a Deep Belief Network (DBN) to extract and recognise the speech emotion features automatically, and then uses a shallow neural network to obtain the final recognition results [28]. So, it seems to be an feasible measure to use shallow neural networks to solve problems mentioned above. In [29], it is statistically confirmed that the shallow models can achieve better performance than the deep model that does not use a regularisation technique. Some different polices have been explored to improve shallow feed forward networks. For example, a random initialisation gradient descent algorithm is proposed for a two-layer neural network with the quadratic activation function, which can make the network convergent to the global optimal at linear speed [30].

According to the discussion above, CNNs really show big advantages in image classification. However, current CNN based approaches usually use very deep networks and are trained on a vast of samples in computing environments like GPUs. Considering that there are not enough well labelled image samples in some applications, these approaches may have to face some problems such as overfitting, high time consumption and so on. It may be a feasible way to address these concerns to repeatably distribute image samples to different based learners, each of which is a shallow CNN with various size of convolutional layers, and then integrate their performance. So, in this paper, we intend to explore a DR early detection learning approach through the classification of retinal images based on shallow CNNs combined with performance integration. With such a measure, we also intend to make the model simpler and to reduce the computation cost.

III. MULTI-SCALE SHALLOW CNNS BASED INTEGRATED LEARNING

A. CONVOLUTIONAL NEURAL NETWORK

Similar to ordinary Neural Networks, a Convolutional Neural Network (CNN) is also composed of neurons that have learnable weights and biases [24]. Each neuron receives some inputs, performs a dot product and connects to the next layer of neurons, and finally applies the result to a non-linear activation function. CNN has become one of the most popular model in deep learning, for its good performance in the classification of images, video, text and so on. Especially, CNN shows big advantages in object, face, and scene recognitions. It can learn directly from image data to perform image classification, and does not need to extract image features manually beforehand.

A typical CNN consists of convolutional layers interspersed with pooling layers, followed by fully connected layers as in a standard multilayer neural network [24], [31]. Such a structure enables CNN to better utilise spatial and configuration information by taking 2D or 3D images as input. Unlike other deep neural networks, CNN exploits three mechanisms of local receptive field, weights sharing, and subsampling as illustrated in Fig. 1 [24]. Usually, a stride of the size of the receptive field in pooling layers is set equal to the size of the receptive field for sub-sampling, which thus helps a CNN to be translation invariant. Such measures can greatly help to reduce the degrees of freedom of a model. Moreover, compared with other deep or feedforward neural networks, a CNN usually needs less parameters and has become an attractive

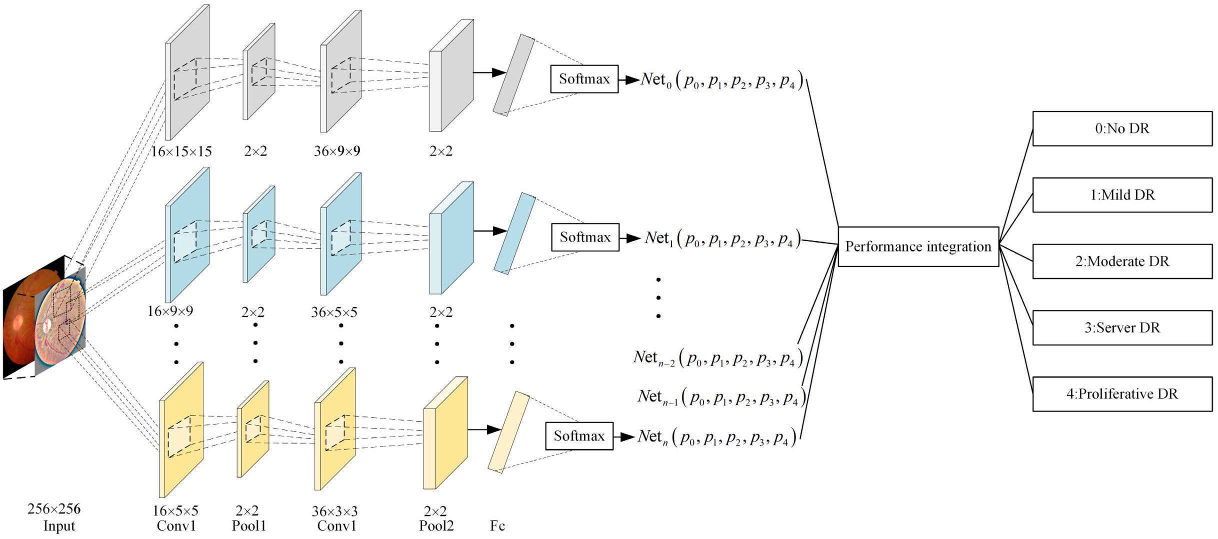

FIGURE 2. **The proposed integrated learning model, which is composed of multi-scale shallow CNNs as base learners. Each base learner has two convolutional layers and two pooling layers followed with one full connected layer.**

deep learning model [18], [20]. Therefore, it has been widely applied in medical image analyses [32], such as breast cancer image classification, classification of leukocytes in peripheral blood images and so on [33].

Deep convolutional neural networks have been popularly used in medical image processing and show big advantages. Considering that deep networks often work on very big datasets, their computation cost are often very high. Exp., GoogLeNet is a deep convolutional neural network architecture, which has gained popularity due to its network architecture and usage in several data science challenges. It introduces a new layer called Inception and the overall number of layers even reaches 100 [23]. At the same time, deep networks also have to face some common challenges like overfitting and poor robustness. Moreover, in many occasions, there are not enough well labeled samples in medical field. In reality, data labelling is always a challenging work in machine learning tasks. These mentioned are very challenging problems to be addressed. Therefore, in this paper, we intend to explore a multi-scale shallow CNNs based integrated learning model for the detection of DR.

B. MULTI-SCALE SHALLOW CNNs BASED INTEGRATED MODEL

According to the analyses to the dataset of retinal images provided, we proposed a shallow CNNs based integrated learning model as shown in Fig. 2. The model proposed is composed of L base learners, each of which acts as a shallow CNN to learn the image features under a specific vision-related receptive field (Fig. 2). To be more concise, in Fig. 2, we only represent the details of 3 out of these base learners, such as their convolutional layers, pooling layers and full connected layers. The output of all learners will be integrated according to a policy proposed to enhance the sensing to multi-scale features and improving the classification finally (see in III-C).

The preprocessed images are delivered to L base learners with different scales, each of which is a shallow CNN that can extract features under various vision-related receptive fields. All these features extracted will be transferred to the full connected layer (FL) finally. To maintain the effectivity of the data and avoid overfitting, a Dropout Layer (DL) is added succeeding to the full-connected layer. Dropout usually randomly deactivates a fraction of the units or connections (e.g., 50%), in a network on each training iteration. A ReLu Layer(RL) is also provided after each of convolutional layers and the full-connected layer to maintain the nonlinear properties of the whole model. Each base learner has an output layer with a Softmax function for classification, and the output of all base learners will be integrated according to our proposed performance integration policy (Eq. 2 and 3).

Each base learner contains 2 Convolutional Layers (CLs) interspersed with 2 Pooling Layers (PLs). CLs are used for local feature detection at different positions in the input feature maps with learnable kernels, and PLs down-sample the feature maps of the preceding convolution layer. The neurons of CL are partially connected to those in a local area of its pre-layer, so as to extract local features. PLs are sensitive to the local area and can enable the secondary feature extraction. The alternative functions of these two kinds of neurons can help improve the learning of the network because this can hold the invariance of displacement and rotation. The feature mapping can be considered as a kind of conventional computation as follows [24],

$$X_j^l = f\left(\sum_{i \in M_j} X_i^{l-1} * K_{ij}^l + b_j^l\right) \tag{1}$$

where l is the index of a layer of the CNN, K is a convolution kernel, M_j is a combination of input feature images, b_j is a bias for output features of each layer, $f(\cdot)$ is a nonlinear activation function and the notation '$*$' denotes a convolution operator.

It is mentioned that, in this paper, ReLU [34] is chosen as the activation function. It performs a threshold operation to each element of the input and any value less than zero is set to zero. PL acts on different areas images, and merges the similar features from the front layer and extracts features further. This measure can largely reduce the parameters produced during the training and is effective to avoid overfitting, and improve the training efficiency finally.

C. PERFORMANCE INTEGRATION

Ensemble Learning is a kind of approaches to combine some simple learners together to get better effects [35]. As it is discussed above, we provide multi-scale shallow CNNs, each of which is known as a base learner, to support feature extraction under various vision-related perceptive fields. If an ideal sample distribution policy with repeatable data sampling is provided as well, it may get better accuracy even on a small dataset.

Obviously, the output of the base learners should be integrated finally. Currently, such integration is usually realised based on Mean or Voting of the prediction results of all base learners. Though these integration approaches are applied to image classification applications, there seems little theoretical arguments to support their effectivity or can be improved. In this paper, we introduce a simple principle, which enhances the performance of strong base learners and reduce that of the weak base learners, to explore a more effective integration approach. Therefore, we call such an approach **performance integration** in this paper.

The proposed performance integration can be realised according to Eq. 2 and 3 as follows.

$$A = \frac{a_i}{\sum_{i=0}^{L-1} a_i} \tag{2}$$

$$C = \frac{\sum_{i=0}^{L-1} (A_i \times Net_i\,(p_0, p_1, \cdots, p_{n-1}))}{L} \tag{3}$$

In Eq. 2, L is the account of base learners, a_i is the classification accuracy of the i^{th} base learner, and $\sum_{i=0}^{L-1} a_i$ is the sum of the accuracies of all base learners. So, when to determine the class of a given input sample, A_i will reflect the impact factor of the i^{th} base learner in the whole integrated model according to its performance (see in Eq. 3). Obviously, it satisfies the condition, $0 \leq i \leq L-1$. In Eq. 3, $Net_i\,(p_0, p_1, \cdots p_{n-1})$ represents the output of the i^{th} base learner when a specific sample is input to the integrated model, where p_j $(0 \leq j \leq n-1)$ represents the probability that the current sample falls into class j with base learner i. Because Diabetic Retinopathy is divided into 5 categories, such as NO DR, Mild DR, Moderate DR and so on, the length of the vector $(p_0, p_1, \cdots p_{n-1})$ is 5. That is say that $n = 5$ in Eq. 3. Thus, when a specific sample is given, the class to which it belongs to will be determined by C based on the outputs of all base learners.

According to Eq. 2 and 3, if the accuracy of the i^{th} base learner is higher than that of others, its impact factor A_i will be great than that of others. It means that this base learner will have a higher impact on the classification result than other ones do. To be contrary, if its factor is smaller, it will be considered as a weak base learner and should have more slight impact on the final classification result (see in Eq. 3). Therefore, based on our performance integration policy proposed, the effect of the strong base classifiers will be enhanced in the classification, and that of the weak ones will be weakened.

IV. DATA REPROCESSING

The dataset, Diabetic Retinopathy Detection, used in this paper comes from the platform Kaggle, which contains 35,126 labeled images. A clinician has rated the presence of diabetic retinopathy in each image with a level from 0 to 4. So, the label of an image will reflect the level of DR evaluated through its features, such as 0-No DR, 1-Mild, 2-Moderate, 3-Severe and 4-Proliferative DR. Fig. 3 illustrates 5 retinal images falling into different categories.

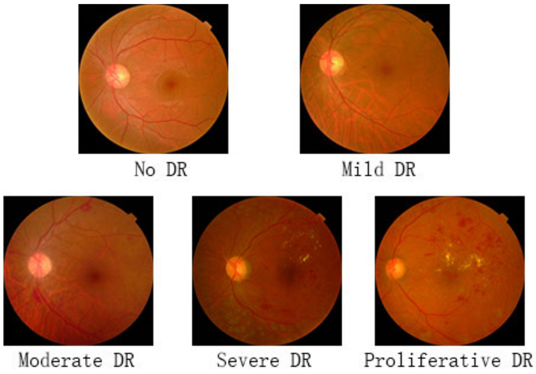

FIGURE 3. A demo of retinal images falling into 5 classes, 0-No DR, 1-Mild, 2-Moderate, 3-Severe, 4-Proliferative DR.

Because there usually exist noises in a dataset, it is necessary to clean the data before further processing. The images used in the experiments are cleaned by removing the black background at first (Fig. 4). At the same time, because the original images are very big and not in the same size, to avoid the deviation might be caused by the inconsistence of data, all images are transformed into a size of 256 × 256 pixels. Moreover, to reduce the noises existing in dim images, histogram equalization is performed on them. At last, to speed up the convergency and improve the learning effect, all images are normalized using the policy represented in Eq. 4. In Eq. 4, x is the original value of a pixel, y is the normalized value, $P_{\max}$ and $P_{\min}$ are the maximum and minimum values of the pixels in the image respectively.

$$y = \frac{x - P_{\min}}{P_{\max} - P_{\min}} \tag{4}$$

The whole preprocessing to a primary retina image is illustrated in Fig. 4, 5 and 6.

V. EXPERIMENTS AND ANALYSES

A. EXPERIMENT DESIGN

A learning system using our proposed approach is implemented with Python and the famous deep learning framework, TensorFlow, on Google platform with GPU of

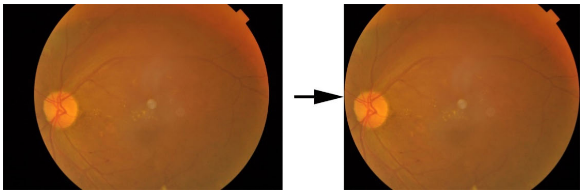

FIGURE 4. A retina image preprocessed by removing the dark background.

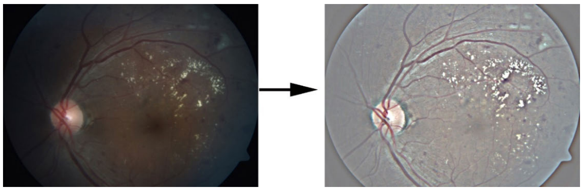

FIGURE 5. A retina image preprocessed with Histogram Equalisation.

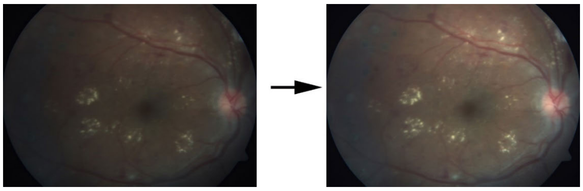

FIGURE 6. A retina image preprocessed with normalization.

Tesla P100. The images used in the experiments are reprocessed beforehand to remove noise, and are normalised as well. Moreover, in order to analyse the performance of our approach on small datasets, 10 sub-datasets with different sizes are derived from the primary dataset. The smallest derived dataset contains 200 images and the largest one is composed of 3500 such images. For each dataset, 70% of its samples selected randomly are used for model training and the remained 30% of samples are used to test the precision and the generalisation capability of the model. Because the distribution of images among these sub-datasets has impact on the model learning, it is critical to reasonably extract image samples for these sub-datasets from the original dataset. Given an image, its entropy can reflect its information amount. More importantly, the 2-D entropy of an image can further reflect the spatial characteristics of its pixel distribution. Therefore, a 2-D entropy based controlled random sample extraction policy is proposed in this paper to derive sub-datasets from the original dataset.

For a pixel in an image, let i represents its gray value and j represents the mean of gray values of its neighbouring pixels. Then, the pair (i, j) will reflect the comprehensive features at the pixel, including both its gray value and the surrounding gray value distribution, where $(0 \leq i, j \leq 255)$. Supposing the width of the image is M pixels, and its height is N, the probability that the feature (i, j) happens in the image can be evaluated according to $p_{i,j} = \frac{f(i,j)}{M \times N}$, where $f(i, j)$ means the frequency of the happening of feature (i, j). Then, the 2-D entropy of the image can be evaluated according to the following equation (Eq. 5).

$$H = -\sum_{i=0,j=0}^{255} p_{ij} \ln p_{ij} \tag{5}$$

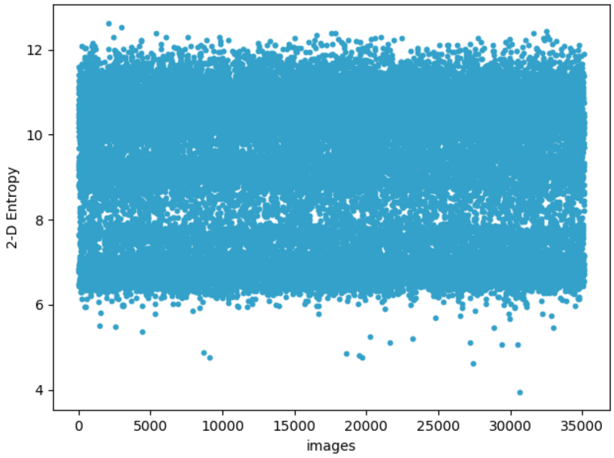

FIGURE 7. 2-D entropy distribution of dataset images.

Based on Eq. 5, we get the 2-D entropy of all images in the original dataset in the experiments. As it is shown in Fig. 7, almost all 2-D entropies of images in the original dataset distribute in the interval [6, 12]. To derive 10 sub-datasets from the original dataset, we partition the interval into 10 segments with equal widths at first. Thus, all original images will be partitioned into 10 groups according to their 2-D entropy. Then, images in different groups will be randomly selected into a sub-dataset with equal probabilities. This can ensure each sub-dataset to cover images with various features as far as possible.

According to our explorative experiments, when the number of base learners of our proposed model increase from 2 to 5, the accuracy of its classification will rise from 82% to 86% correspondingly. However, when it goes beyond 5, the accuracy has no outstanding increase. So, considering of the integration efficiency at the same time, we use 5 base learners in our proposed model. Each of them is a shallow CNN identified by Net-i, where $0 \leq i \leq 4$. Table 1 gives the details of each base learner, including the size of each convolutional layer, and the input and output of the full connected layer. Taking Net-0 as an example, its first convolutional layer has 16 convolutional kernels with a size of 15 × 15. Its second convolutional layer has 36 convolutional kernels, each of which has a size of 9 × 9. The full connected layer of Net-0 contains 128 features.

TABLE 1. The structural details of each base learner.

	Size of 1^{st} CL	Size of 2^{nd} CL	Output of FL	Output of FL
Net-0	$16 \times 15 \times 15$	$36 \times 9 \times 9$	1×128	1×5
Net-1	$16 \times 9 \times 9$	$36 \times 5 \times 5$	1×128	1×5
Net-2	$16 \times 11 \times 11$	$36 \times 9 \times 9$	1×128	1×5
Net-3	$16 \times 5 \times 5$	$36 \times 3 \times 3$	1×128	1×5
Net-4	$16 \times 5 \times 5$	$36 \times 3 \times 3$	1×128	1×5

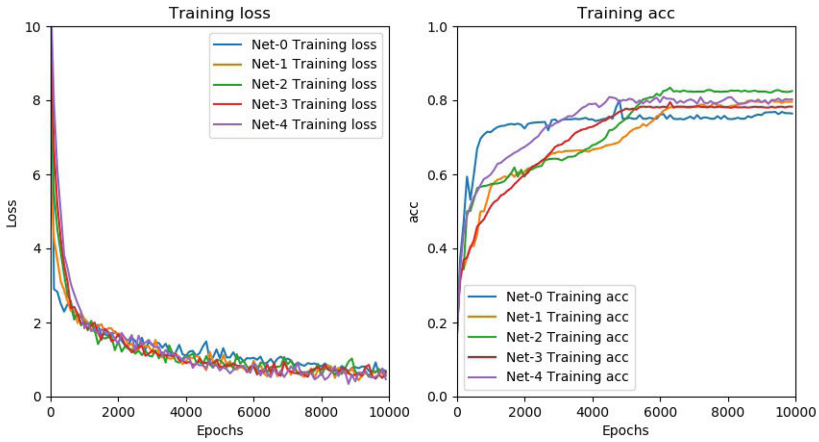

FIGURE 8. The loss and accuracy (Acc) curves of each base leaner.

In the next section, to evaluate the performance of our proposed model in retinal image classification, it will be compared experimentally with some related CNN based approaches.

B. RESULTS AND ANALYSES

To observe the training process of five base learners of our proposed integrated shallow CNN model, Net-0, Net-1, Net-2, Net-3 and Net-4, on the same dataset, the loss and accuracy of each base learner are illustrated in Fig. 8. For example, as to base learner Net-0, its accuracy curve shows an ascending trend as the loss descends during the training process. When the iterations reach 7000 times, the network becomes convergent and the accuracy gets up to about 0.77 as the loss declines to about 0.75. The other base learners represent the similar trends as the loss and the accuracy varying. The classification accuracy and the iterations needed for the convergency of each base learner are represented in table 2. From Fig. 8 and table 2, it is found that different sizes of CLs really have impacts on the convergency speed and the accuracy of the model. It seems that Net-2 have advantages in the accuracy and convergency speed compared with other base learners. That is to say that the integration of multi-scale shallow CNNs is a feasible way to get a better integrated learner for the classification of medical images.

TABLE 2. The loss and iterations of each base learner.

	Accuracy of Classification	Iterations Needed for Convergency
Net-0	0.77	7000
Net-1	0.79	8000
Net-2	0.82	6000
Net-3	0.78	7000
Net-4	0.80	9000

In Fig. 9, the performance of our proposed integrated CNN model is compared with its best base learner, which is a traditional CNN, on both the original dataset and the 10 sub-datasets mentioned above. It shows that as the size of the training dataset increases from 200 to 2000, the classification accuracy of our integrated model is improved correspondingly. It means that the increase of training samples is really helpful to improve the classification effect. When the amount of the training samples goes beyond 2000, it is found that the accuracy of our proposed integrated model on the three remained bigger sub-datasets, which has a size of 2500, 3000 and 3500 respectively, will no longer be improved as the training sample increasing. Especially, on the biggest sub-dataset, which has 3500 images, the accuracy of our proposed model is higher than those on other sub-datasets. Fig. 9 also shows it is similar when using the traditional CNN model. It found that, on the biggest sub-dataset, the classification accuracy of the best base learner reaches 0.82, and that of our integrated model is improved to 0.86 owing to the proposed performance integration policy. This details can also be seen in table 3. So, it shows that the integration of multi-scale shallow CNN can really improve the classification accuracy.

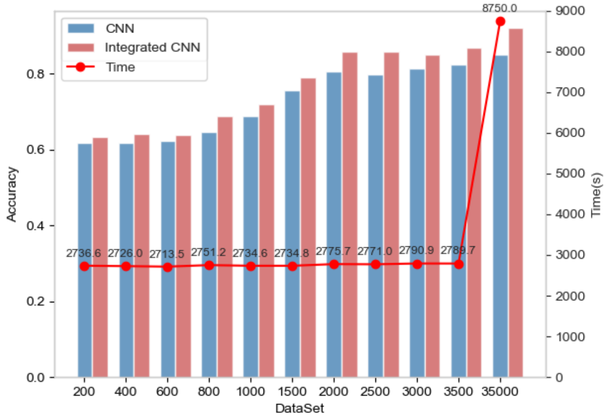

FIGURE 9. The comparison of accuracy and training time between the original dataset and 10 sub-datasets.

Though the classification accuracy of our model on the sub-dataset with 3500 images is 6% lower than its classification accuracy on the original dataset, the size of the

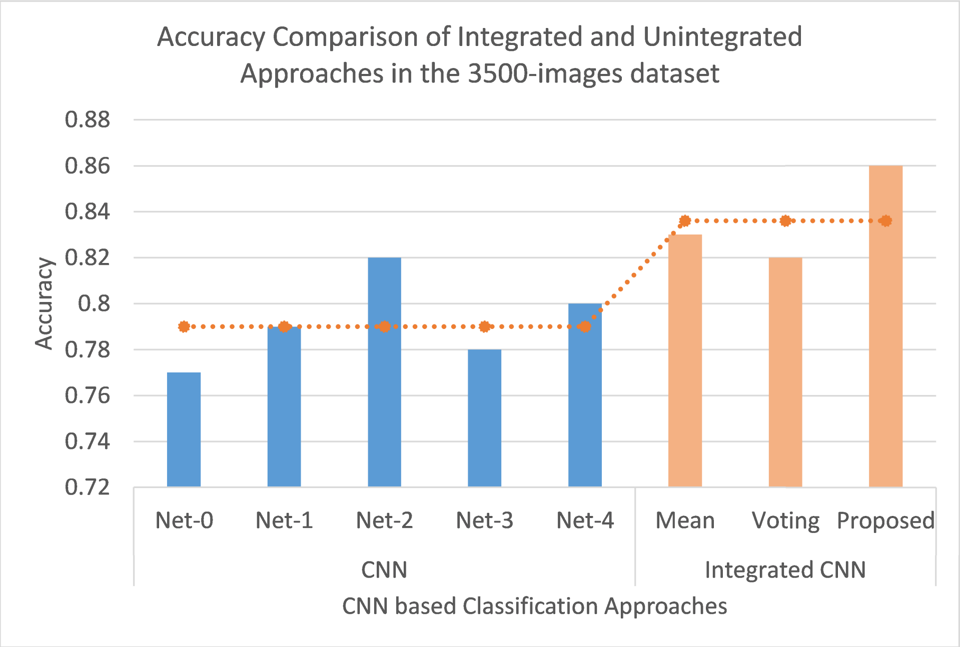

FIGURE 10. The accuracy comparison of shallow CNN based integrated model with the traditional CNNs.

TABLE 3. The classification accuracy and time cost on different datasets.

Dataset	Accuracy of the Best Base Learner	Accuracy of the Proposed Model	Training Time(s)
200	0.62	0.63	2736.6
400	0.62	0.64	2726.0
600	0.62	0.64	2713.5
800	0.65	0.69	2751.2
1000	0.69	0.72	2734.6
1500	0.75	0.79	2734.8
2000	0.80	0.84	2775.7
2500	0.80	0.85	2771.0
3000	0.81	0.85	2790.9
3500	0.82	0.86	2789.7
35000	0.85	0.92	8750.0

sub-dataset is only 10% of that of the original dataset and the training time cost declined greatly (Fig. 9). Table 3 gives the details of the time cost when the model is trained on sub-datasets with different size. For example, it takes only 2736.6s to finish the model training on the sub-dataset with 200 images, and 2789.7s on that with 3500 images. Compared with the original dataset with 35,000 images, the training time cost on the sub-datasets declines by a rate up to 70%. That is to say that when enough labeled samples can not be provided and time consumption is complained on some occasions, the proposed approach can be an acceptable measure to do image classification tasks.

Based on the analyses above and experimental results, considering both the classification effect and time cost, the sub-dataset with 3500 images is selected as the optimal training samples to evaluate our model further through the comparison with current similar models.

To verify the effectivity of our Performance Integration policy, we compared our proposed model two representative integrated shallow CNN models using Mean and Voting as policies to integrate the classification results of their base learners respectively. Fig. 10 represents that each base learner really shows disadvantages independently in accuracy compared with the integrated learners. Though Net-2 performs best on both the accuracy and the convergency speed in all base learners, its accuracy is still lower than those of all integrated models from 1% to 4%. The dash line in Fig. 10 shows the average accuracy of all base learners and that of all integrated models respectively. It shows the average accuracy of 3 different integration approaches is higher than that of the 5 base learners by 4.6%. This further verifies that the integration of shallow CNNs is a feasible and effective way to improve the classification effects of medical images. On the other hand, from the figure, we can also find that our proposed performance integration model has advantages in accuracy compared with other two integrated CNN models. Table 4 gives their accuracies on the sub-dataset containing 3500 DR images. It is mentioned that our proposed model use the policy called Performance Integration. Just as it is shown in table 4, the integration approach proposed in this paper can improve the classification accuracy by 3% compare with that based on Mean, and 4% compared with that based on Voting. Fig. 10 illustrated the comparison of all 5 base leaners and integrated shallowing CNNs based on Mean, Voting and Performance Integration. It can be found that our proposed model can improve the accuracy from 2% to 9% compared with all these models.

To further verify the advantages of our proposed approach, it is compared with other approaches both on all small sub-datasets and the original dataset. It is necessary to mentioned that VGG16noFC is an approach to detecting retinopathy by

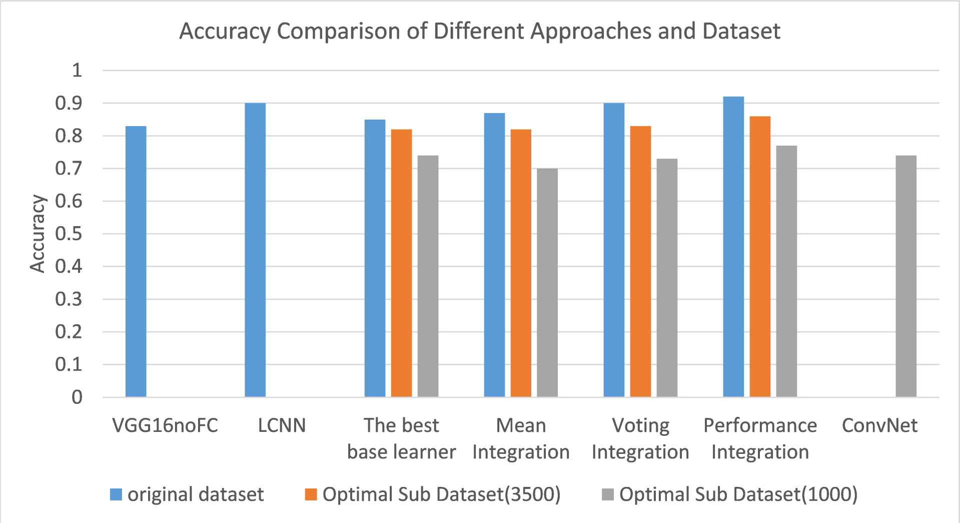

FIGURE 11. The accuracy comparison of the proposed shallow multi-scale CNNs based approach, Performance Integration, with current approaches on different datasets.

TABLE 4. The accuracy comparison of different integration approaches.

Integration Approach	Classification Accuracy
Mean based Integration	0.83
Voting based Integration	0.82
Performance based Integration	0.86

extracting pathological features in the retinal image based on CNN [36]. Another approach having good effect to classify retinal images is called LCNN, which uses a contrast enhancement filter to preprocess retinal images beforehand, and then the preprocessed images are put into a shallow neural network with three convolutional layers and two pooling layers for classification [37]. According to [38], ConvNet also shows good performance on 1000 retina images selected from the original dataset.

TABLE 5. The accuracy comparison of different approaches and datasets.

	Classification Approach	Classification Accuracy
Original dataset	VGG16noFC	0.83
	LCNN	0.89
	ShallowNet+Mean integration	0.87
	ShallowNet+Voting integration	0.9
	ShallowNet+Performance integration	0.92
Optimal Sub-dataset(3500)	ShallowNet+Mean integration	0.82
	ShallowNet+Voting integration	0.83
	ShallowNet+Performance integration	0.86
Optimal Sub-dataset(1000)	ConvNet	0.74
	ShallowNet+Mean integration	0.70
	ShallowNet+Voting integration	0.73
	ShallowNet+Performance integration	0.77

As it is shown in table 5 and Fig. 11, the three integrated approaches all have higher accuracy compared with VGG16noFC and LCNN on the original dataset. When comparing with VGG16noFC and LCNN, our proposed performance integration approaches has a 9% and 3% higher accuracy respectively on original dataset. In addition, our approaches is 3% higher than ConvNet on the dataset composed of 1000 images, which once again verifies that our approaches is effective on small datasets. Even our integrated model is trained on the much smaller optimal sub-dataset, whose size is only 10% of that of the original dataset, its accuracy is only 3% lower than that of LCNN on the original dataset. Nevertheless, table 5 and Fig. 11 together show that our proposed approach based on shallow CNNs and performance integration has advantages in accuracy compared with other approaches on the same dataset. Although the performance of our approach decreases when the datasets become much smaller, the decrease degree is acceptable considering of the high efficiency. Importantly, when there are no enough labeled samples, our proposed approach can act well on the small datasets to do classification of medical image.

VI. CONCLUSION AND FUTURE WORK

In this paper, multi-scale shallow CNNs combined with performance integration is introduced to the early detection of Diabetic Retinopathy through the classification of retinal images. Owing to the feature sensing under various vision-related receptive fields by different base learners and the repeatable dataset sampling, it can do image classification well when there are not enough high-quality labelled samples. According to the experiments, the performance integration model shows advantage in accuracy compared with other integration models like that based on Mean and Voting. Moreover, the proposed approach also performs well on small datasets when considering of both classification effect and efficiency compared with other approaches.

In future work, we will enhance our proposed approach as follows. (1) More effective approach to integrating shallow CNNs will be explored to further improve the classification accuracy. (2) The transformation of image samples and the repeatable sampling of the dataset will be combined to improve the performance of integrated shallow CNN model.

REFERENCES

[1] D. S. Fong, L. Aiello, T. W. Gardner, G. L. King, and R. Klein, "Retinopathy in diabetes," *Diabetes Care*, vol. 27, no. 1, p. 7, 2004.

[2] *International Diabetes Federation*. [Online]. Available: https://diabetesatlas.org/

[3] A. V. Chappelow, K. Tan, P. K. Kaiser, and N. K. Waheed, "Reply," *Amer. J. Ophthalmol.*, vol. 153, no. 4, p. 781, Apr. 2012, doi: 10.1016/j.ajo.2012.01.002.

[4] H. Thanati, R. J. Chalakkal, and W. H. Abdulla, "On deep learning based algorithms for detection of diabetic retinopathy," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Auckland, New Zealand, Jan. 2019, pp. 1–7.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[6] J.-E. Liu and F.-P. An, "Image classification algorithm based on deep learning-kernel function," *Sci. Program.*, vol. 2020, pp. 1–14, Jan. 2020, doi: 10.1155/2020/7607612.

[7] H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma, and W. Qian, "An interpretable ensemble deep learning model for diabetic retinopathy disease classification," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 2045–2048.

[8] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Classification of diabetic retinopathy and normal retinal images using CNN and SVM," in *Proc. 12th Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, Surabaya, IN, USA, Jul. 2019, pp. 152–157.

[9] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.

[10] S. Zhou, C. Chen, G. Han, and X. Hou, "Deep convolutional neural network with dilated convolution using small size dataset," in *Proc. Chin. Control Conf. (CCC)*, Guangzhou, China, Jul. 2019, pp. 8568–8572.

[11] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning Earth observation classification using ImageNet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016, doi: 10.1109/LGRS.2015.2499239.

[12] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 7, pp. 729–736, Jul. 1995, doi: 10.1109/34.391417.

[13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

[15] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, USA, 2004, pp. 1–5.

[16] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1019–1032, Jun. 2013, doi: 10.1109/TMI.2013.2247770.

[17] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, "Deepr: A convolutional net for medical records," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 22–30, Jan. 2017, doi: 10.1109/JBHI.2016.2633963.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Harrahs Harveys, Lake Tahoe, 2012, pp. 1097–1105.

[19] J. Xu, T. Li, Y. Chen, and W. Chen, "The impact of bathtub water temperature on personal identification with ECG signal based on convolutional neural network," in *Proc. IEEE 4th Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2018, pp. 1341–1345.

[20] Q. Zhou, J. Xing, W. Chen, X. Zhang, and Q. Yang, "From signal to image: Enabling fine-grained gesture recognition with commercial Wi-Fi devices," *Sensors*, vol. 18, no. 9, p. 3142, Sep. 2018, doi: 10.3390/s18093142.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, 2015, pp. 448–456.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[24] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[25] C.-H. Chang, "Deep and shallow architecture of multilayer neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2477–2486, Oct. 2015, doi: 10.1109/TNNLS.2014.2387439.

[26] S. Dey, A. Dutta, J. Llados, A. Fornes, and U. Pal, "Shallow neural network model for hand-drawn symbol recognition in multi-writer scenario," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Kyoto, Japan, Nov. 2017, pp. 31–32.

[27] D. E. Kim and M. Gofman, "Comparison of shallow and deep neural networks for network intrusion detection," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Las Vegas, NV, USA, Jan. 2018, pp. 204–208.

[28] J. Wang and Z. Han, "Research on speech emotion recognition technology based on deep and shallow neural network," in *Proc. Chin. Control Conf. (CCC)*, Guangzhou, China, Jul. 2019, pp. 3555–3558.

[29] K. Pasupa and W. Sunhem, "A comparison between shallow and deep architecture classifiers on small dataset," in *2016 8th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE),* Yogyakarta, IN, USA, 2016, pp. 1–6.

[30] S. Xia and Y. Shi, "Learning shallow neural networks via provable gradient descent with random initialization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 5616–5620.

[31] S. Lawrence, C. L. Giles, A. Chung Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Feb. 1997, doi: 10.1109/72.554195.

[32] G. Chen, Y. Chen, Z. Yuan, X. Lu, X. Zhu, and W. Li, "Breast cancer image classification based on CNN and bit-plane slicing," in *Proc. Int. Conf. Med. Imag. Phys. Eng. (ICMIPE)*, Shenzhen, China Nov. 2019, pp. 1–4.

[33] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," *Med. Biol. Eng. Comput.*, vol. 55, no. 8, pp. 1287–1301, Aug. 2017, doi: 10.1007/s11517-016-1590-x.

[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines vinod nair," in *Proc. 27th Int. Conf. Mach. Learn.* Haifa, Israel, 2010, pp. 21–24.

[35] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning," *Decis. Support Syst.*, vol. 57, pp. 77–93, Jan. 2014, doi: 10.1016/j.dss.2013.08.002.

[36] G. Garcia, J. Gallardo, A. Mauricio, J. Ló́pez, and C. Del Carpio, "Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images," in *Proc. Int. Conf. Artif. Neural Netw.*, Alghero, Italy, 2017, pp. 635–642.

[37] M. S. Chowdhury, F. R. Taimy, N. Sikder, and A.-A. Nahid, "Diabetic retinopathy classification with a light convolutional neural network," in *Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng.*, Rajshahi, Bangladesh, Jul. 2019, pp. 1–4.

[38] M. Arora and M. Pandey, "Deep neural network for diabetic retinopathy detection," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, faridabad, Haryana, Feb. 2019, pp. 189–193.

WANGHU CHEN received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, in 2009. He was a Visiting Scholar with the University of California, San Diego, from 2013 to 2014. He is currently a Professor with the Institute of Computer Science and Engineering, Northwest Normal University, China. He is a PI of five research projects, including the National Science Funds. He is also a PC member of international workshops. He has published more than 40 articles. His research interests include big data and cloud computing.

BO YANG received the bachelor's degree in mathematics and applied mathematics from Shenyang University, in 2017, where he is currently pursuing the master's degree. His research interests include image processing and computer vision.

JING LI received the master's degree from Lanzhou University, China, in 2013. She is currently an Associate Professor with the Institute of Computer Science and Engineering, Northwest Normal University, China. Her research interests include big data, scientific workflow, and service-oriented computing.

JIANWU WANG received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, in 2007. He is currently an Assistant Professor with the Department of Information Systems, University of Maryland, Baltimore County (UMBC). He has published more than 100 articles with more than 1500 citations. His research interests include big data, scientific workflow, distributed computing, and service-oriented computing. He is a program committee member of over 40 conferences/workshops. He serves as an associate editor or editorial board member for four international journals and the Co-Chair for three related workshops. He serves as a reviewer for over 15 journals or books.