Image Anal Stereol 2023;42:77-91 doi: 10.5566/ias.2857
Original Research Paper

DEEP NEURAL NETWORK-BASED ENSEMBLE MODEL FOR EYE DISEASES DETECTION AND CLASSIFICATION

AFSANA AHSAN JENY[1,2], MASUM SHAH JUNAYED[1,2] AND MD BAHARUL ISLAM✉[,1,3,4]

[1]Department of Computer Engineering, Bahcesehir University, Yildiz, Ciragan Cd, Besiktas, 34349, Istanbul, Turkey, [2]Department of CSE, University of Connecticut, Storrs, CT, United States, [3]Department of Computer Science & Engineering, Daffodil International University, DIU Road, Dhaka 1341, Bangladesh, [4]College of Data Science and Engineering, American University of Malta, Triq Dom Mintoff, BML 1013, Malta
e-mail: afsanaahsan1996@gmail.com, masumshahjunayed@gmail.com, bislam.eng@gmail.com
(Received December 25, 2022; accepted June 11, 2023)

ABSTRACT

Fundus images are the principal tool for observing and recognizing a wide range of ophthalmological abnormalities. The automatic and robust methods based on color fundus images are urgently needed since few symptoms are observable in the early stages of the disease. Experts must manually evaluate images to detect diseases for screening procedures to be effective. Due to the complexity of the screening procedure and the shortage of experienced personnel, developing successful screening-based treatments is costly. Although existing automated approaches strive to address these issues, they cannot handle a wide range of diseases and real-world circumstances. We design an automated deep learning-based ensemble method to detect and classify eye diseases from fundus images to address the abovementioned problems. A deep CNN-based model is proposed in the ensemble method that incorporates a mix of 20 layers, including the activation, optimization, and loss functions. The contrast-limited adaptive histogram equalization (CLAHE) and Gaussian filter are utilized in the pre-processing step to get more explicit images and eliminate noise. To avoid overfitting in the training phase, augmentation techniques are applied. Three pre-trained CNN models, including VGG16, DenseNet201, and ResNet50, are employed to compare and assess the efficiency of the proposed CNN model. Experimental results demonstrate that the ensemble approach outperforms recent approaches, which is comparatively state-of-art in the ODIR publicly available dataset.

Keywords: Convolutional neural network, Deep learning, Ensemble model, Medical imaging, Ocular Disease.

INTRODUCTION

Infections of the fundus are the significant and primary cause of impairment in humans on a global scale (Jonas *et al.*, 2014). According to Costagliola *et al.* (2009), the number of people living with diabetic retinopathy (DR) will exceed 400 million by 2030, while the prevalence of severe glaucoma could exceed 80 million by 2020. Fig. 1 illustrates the efficacy of the most prevalent ocular conditions, including DR, glaucoma, cataracts, age-related macular degeneration (AMD), uncorrected refractive error (URE), corneal opacity, and other disorders. In the case of ophthalmic diseases, they are becoming a big issue for public health worldwide. It is possible that the ophthalmic condition might induce lifelong blindness since it has irreversible symptoms.

To improve patients' quality of life, early detection of eye problems is vital. A WHO research from 2021 estimates that there are 2.2 billion people with visual impairments globally, with around 1 billion preventable cases (WHO, 2022). On the other hand, there are fewer ophthalmologists compared to the number of patients. In addition, manually scanning the fundus is a very time-consuming process that places a large amount of reliance on the expertise of ophthalmologists. Consequently, it is essential to have an automated computer-aided diagnostic tool for scanning ocular diseases.

It is challenging to develop a computer-aided diagnosis system, for instance, Microaneurysm is a crucial guideline for DR screening. Furthermore, it is difficult to accurately detect ocular disorders due to inadequate contrast between the lesion and background pixels, an irregular lesion morphology, and significant variations between the same lesion spots produced by different cameras. Then fluorescein angiography and optical coherence tomography (OCT) are two techniques to explore the retina in greater depth. Optical coherence tomography (OCT) image acquisition methods have recently been developed. However, these devices are more expensive than fundus cameras. The advent of low-cost, high-capacity computers in recent years has made artificial intelligence (AI) potentially crucial to the analysis of medical images. To boost disease diagnosis, AI

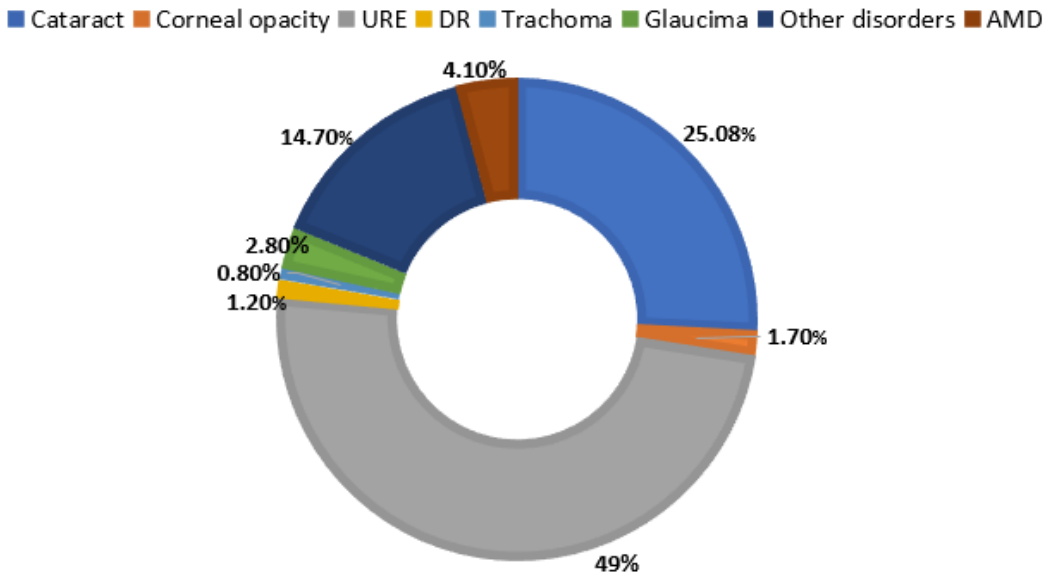

Fig. 1: *The prevalence of different visual disorders.*

research aims to develop models that can work with image-capturing tools.

Deep learning models for eye disease detection have been constructed, and their performance has been considered noteworthy. However, there are still some drawbacks. Firstly, some prior research has been focused on identifying and classifying only one to three disorders: DR (Wang *et al.*, 2020; Jordi *et al.*, 2019; Sarki *et al.*, 2020), glaucoma (Orfao and Haar, 2021; Raghavendra *et al.*, 2018; Chai *et al.*, 2018; Gupta *et al.*, 2022), and cataracts (Junayed *et al.*, 2021b; Khan *et al.*, 2021). However, considering the real needs of patients with fundus illness daily, we believe that developing a more effective and reliable fundus testing approach to detect various diseases is crucial. Secondly, it's difficult to train a single model to attain high disease detection accuracy with limited fundus imaging data and inevitable picture noise. Thirdly, several deep learning-based methods (Li *et al.*, 2019a; 2020; Gour and Khanna, 2021; Yang and Yi, 2022) utilized the pre-trained models or proposed model with a lot of blocks. These models had heavy architectures due to the increased number of layers, weights, and parameters. Because of this, classifying and detecting eye disorders using these deep learning-based algorithms require a significant amount of computing cost.

Contributions. We suggest an end-to-end ensemble model based on deep learning to diagnose eight distinct fundus disorders to overcome the aforementioned issues. The main contributions of our study are as follows:

- We design an automated deep learning-based ensemble method to detect and classify eye diseases from fundus images. Due to smartphones' widespread availability, this method might be utilized by the general public as a remote screening tool, particularly in impoverished countries where few ophthalmologists are accessible.
- In this article, we propose a deep CNN-based model as an attachment for the ensemble approach. To decrease the computational cost, the CNN model combines 20 layers, along with the activation function, optimization technique, loss function, and batch size.
- To get clearer images and eliminate noise, contrast-limited adaptive histogram equalization (CLAHE) and Gaussian filter are employed in the pre-processing step, followed by augmentation approaches.
- Three pre-trained CNN models, including VGG16, DenseNet201, and ResNet50, are compared to demonstrate and assess the efficiency of the proposed CNN model. In addition, the ensemble technique is compared to contemporary state-of-the-art techniques, showing that our method outperforms others.

RELATED WORKS

Several machine learning and deep learning techniques have been used to study the detection of ocular disorders. Some of the approaches to eye disorders are extensively described in this section.

For the ODIR database, Islam *et al.* (2019) proposed a shallow CNN approach that requires initial fundus image training. Images of the left and right eyes accompany every label in the ODIR database. The labels associated with eye disorders were removed after they carefully examined each picture on their own. This streamlined the task but did not provide a model that can concurrently recognize and categorize many diseases for a set of photos. Then a complete integrated learning model was then given by Qummar *et al.* (2019) by combining five CNN models: Resnet 50, InceptionV3, Xception, Dense121, and Dense169. Five categories were used in this paradigm to categorize DR: normal, mild, moderate, severe, and PDR (Infectious Eye Disease). The model's accuracy percentage is 80.8%. Jordi *et al.* (2019) proposed an approach in which the ODIR database would be changed to solve the challenge of multi-class categorization. VGG16 and InceptionV3 are the pre-trained architectures that categorize fundus pictures taken from the ODIR database. Following this, the same database is used to make use of the substantial spatial correlations between the two-color fundus pictures. A transfer learning technique based on ResNet architecture and a Dense Correlation Network was proposed by Li *et al.* (2019b). This approach, however, is computationally costly (for example, the trainable parameters are 74.2M), and the accuracy is 82.7%. Another study, (Wang *et al.*, 2020), created a multi-label classification clustering algorithm using the pre-trained EfficientNet architecture. In the first part, the EfficientNet network is utilized for feature extraction, and in the second, a custom neural network architecture is employed for multi-label classification. This method of experimentation also uses the ODIR 2019 dataset, which has an accuracy of 89%. However, they didn't use preprocessing techniques to reduce visual picture noise and boost accuracy.

Several studies attempted identifying two or three eye diseases using pre-trained deep learning models. For instance, Li *et al.* (2019a) successfully attained an accuracy of 98.6% when using the VGG16 model to categorize AMD and DME in a dataset of 207,130 pictures gathered through OCT imaging. An 18 convolutional neural network was built by Raghavendra *et al.* (2018) to diagnose just Glaucoma using fundus photographs. The model used in this paper had an accuracy of 98.13%. An R-CNN, an FCN, and a customized CNN model were provided to identify Glaucoma as part of the research conducted by Chai *et al.* (2018). Through the use of the dataset in the process of model evaluation, they were able to achieve a success rate of 91.51%. Khan *et al.* (2020) used the transfer learning pipelines to detect three forms of eye illness, including normal, Cataract, and Hypertensive Retinopathy were improved by using the average of estimates from pre-trained CNN models, consisting of ResNet50, InceptionResNetV2 (Siciarz and McCurdy, 2022), EfficientNetB0 (Gaur *et al.*, 2021), and EfficientNetB2 (Ayana *et al.*, 2022). An enhanced and adaptive histogram equalization technique based on morphological processes was used instead of raw photos. In binary classification, the proposed ensemble-based method beat the pre-trained CNN models. Each of the five models had an accuracy rate of 82.57%, 80.63%, 80.67%, 84.22%, and an ensemble model accuracy rating of 85.08%, respectively. Orfao and Haar (Orfao and Haar, 2021) employed a variety of pre-trained models, including InceptionV3, Alexnet (Chen *et al.*, 2021), VGGNet, and ResNet, to detect Glaucoma, Diabetic Retinopathy, and Cataracts from fundus photos. The InceptionV3 model, which was 225 MB in size and had an accuracy of 99.30% and an F1-Score of 99.39%, produced the best results. After that, when Histogram of Oriented Gradients (HOG) features are employed for feature extraction, the accuracy of the Support Vector Machine (SVM) is 76.67%, and its F1-score is 76.48%, respectively. The researchers Junayed *et al.* (2021b) next used a suggested CNN model to categorize only cataract photos using datasets that were freely accessible. This article uses specific pre-processing processes, such as normalization and data augmentation. These techniques were employed to improve the accuracy of the data. Their precision was exceptional; nonetheless, the test was designed only to find cataracts, not to score where they were in the eye. Another study by Khan *et al.* (2021) selected an architecture based on the VGG19 (Han *et al.*, 2021) model to diagnose cataracts autonomously from color fundus pictures. This model achieved an accuracy and prediction rate of 97.47%, respectively.

In addition, several recent works (Li *et al.*, 2020; Gour and Khanna, 2021; He *et al.*, 2021; Yang and Yi, 2022) utilized the ODIR database to identify and categorize eight types of eye illnesses. For instance, (Li *et al.*, 2020) tried to detect multiple eye diseases using state-of-the-art deep neural networks. When it comes to multi-disease categorization, expanding the number of nodes in a network isn't enough; a well-structured neural network-based approach is required, they claimed. After that, Gour and Khanna (2021) employed a transfer learning-based convolutional neural network (CNN) to detect eight categories of eye disease. On the ODIR database, VGG16 pre-trained architecture with SGD optimizer appears to be better for multi-class multi-label fundus picture classifications than four other pre-trained CNN architectures with two other optimization techniques. 84.93%, 85.57% achieved

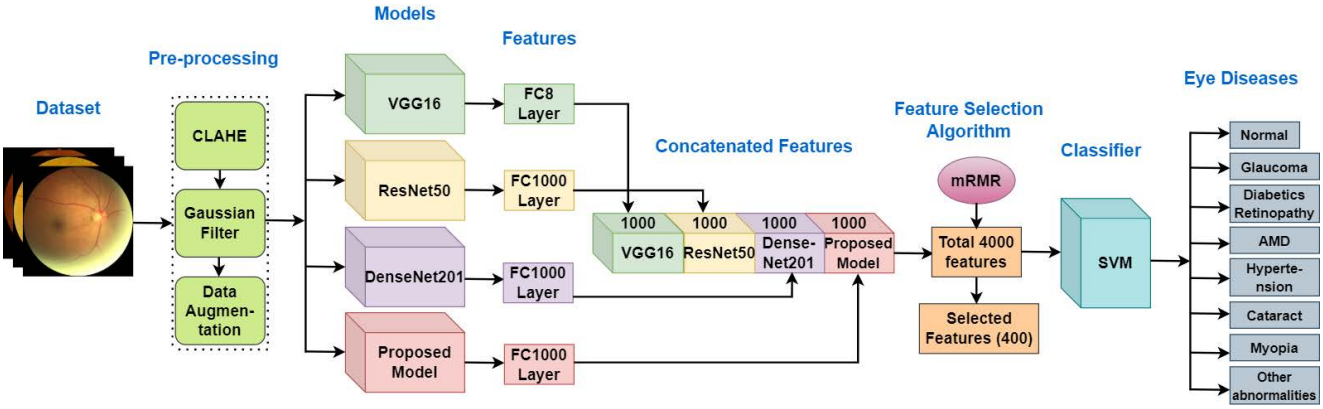

Fig. 2: *The overall flowchart of the proposed-ensemble method includes a pre-processing step, a step of ensembling by three pre-trained models, our proposed CNN model with a feature selection algorithm, and finally, a classification step.*

the AUC, F1-score, and accuracy with VGG16 and SGD optimizer, and around 85%, respectively. After that, He *et al.* (2021) proposed another work to detect eye disease using the ODIR database. In He *et al.* (2021), the pre-trained models are used as a feature extractor, a module for spatial correlation, and a classifier for classification scores. 93% and 91.3% obtain the AUC and F1-Score; however, their proposed network is too expensive (i.e., 74.2M parameters) because of their proposed feature correlation module. Recently, Yang and Yi (2022) suggested another deep learning-based network including too many modules such as DS block, DSR block, and SE block to decrease the computation required and improve the calculation of the feature of the images and filter the features. The accuracy, precision, F1 value, and kappa score of the DSRA-CNN network suggested in this research are respectively 87.90%, 88.50%, 88.16%, and 86.17% when tested on the ODIR dataset. However, employing many modules, their architecture showed computational cost issues (i.e., around 26M), and it may face time complexity problems.

The literature review results show that several experiments have been conducted employing pre-trained deep learning algorithms on a few diseases, such as DR, cataracts, and glaucoma. In contrast, few studies have been published employing the recommended deep-learning techniques to diagnose eye diseases. Consequently, several barriers still exist, including raising model accuracy while decreasing computational burden, reducing the number of training parameters, and including more common eye diseases.

MATERIALS AND METHOD

This article suggests a novel method for detecting and classifying eight different ophthalmological diseases utilizing fundus images. Fig. 2 depicts the proposed framework, which is divided into three phases: First, a pre-processing step is done with three methods, including Contrast Limited Adaptive Histogram Equalization (CLAHE), Gaussian filter, and data augmentation. Secondly, the pre-processed images are placed into our proposed-ensemble model for feature extraction up to the fully connected (FC) layer; then, the features are concatenated and selected most efficient features by the mRMR algorithm. And finally, the classification process is done by the SVM classifier in 8 distinct eye disease classes. The following subsections provide more in-depth information on each stage of the suggested structure.

DATASET

For this research, we leverage color fundus images from the Ocular Disease Intelligent Recognition (ODIR) collection (ODIR-2019, 2022). There are a total of 8 types of illnesses or disorders represented among the 3098 healthy, 1406 diabetic, 224 glaucoma, 265 cataracts, 293 AMD, 107 hypertension, 242 pathological myopia, and 791 other samples (total of 6426 images). Among others, Canon, Zeiss, and Kowa cameras are often employed to record fundus photos, which are subsequently saved in JPG format with different sizes and resolutions. Professional ophthalmologists perform all classification techniques in the dataset. Several examples from the ODIR dataset for each class are shown in Fig. 3.

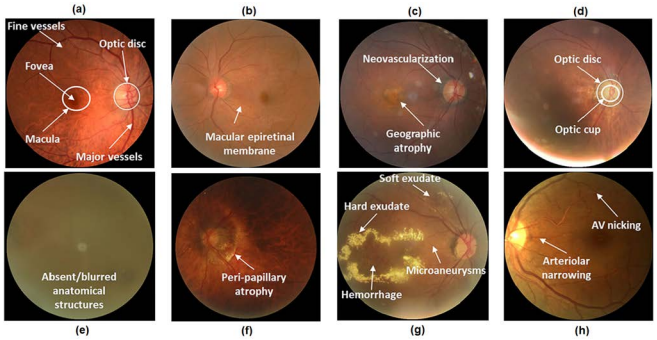

Fig. 3: *The sample images from the ODIR dataset. (a), (b), (c), (d), (e), (f), (g), and (h) show Normal, Other Abnormalities, AMD, Glaucoma, Cataract, Myopia, Diabetic Retinopathy, and Hypertension images, respectively.*

DATA PRE-PROCESSING

The overall efficiency of the model is highly dependent on the quality of the input pictures. In order to improve the efficacy of the image classifications, image pre-processing is essential. In the past, the system's efficiency suffers from the pictures' blurry borders. This means that all of the photos need to be regulated and their characteristics improved. Our suggested method divides the pre-processing phase into three distinct phases: the Contrast Limited Adaptive Histogram Equalization (CLAHE) (Sahu *et al.*, 2019) phase, the Gaussian filter (Wang *et al.*, 2018) phase and the data augmentation phase.

At first, the images are cropped to eliminate the extra black pixels surrounding the retina. Then CLAHE method is employed to enhance the images. CLAHE effectively improves the poor contrast of medical pictures, brightens fundus pictures (Sahu *et al.*, 2019), and improves border and regional information. CLAHE is performed to the L channel of the eye pictures with more excellent contrast, with a tile size of 8 by 8 and a Clipping Limit of 5.0. The CLAHE approach could generate some noise in the images. Thus we employ the Gaussian filter, as shown in eq (1), to eliminate it.

$$G_f(a,b) = Ae^{\frac{-(a-\lambda_a)^2}{2\beta_a^2}+\frac{-(b-\lambda_b)^2}{2\beta_b^2}} \quad (1)$$

where λ and A represent the mean and amplitude, as well as β indicates the standard deviation for both of the variables a and b, respectively. The results of applying the CLAHE and Gaussian filters to the photos are shown in Fig. 4.

With deep learning-based models, having a significant amount of training data is crucial. Overfitting is possible during neural network training due to the small size of the initial eight-class dataset. Five additional data augmentation procedures, such as random rotation by 30 degrees to the left or right, horizontal flipping, shearing, scaling, and translation, are implemented to enhance the size of the training dataset and improve the classification performance and overall efficiency of the CNN. After that, the pictures are resized to a specific ratio (256 $\times$ 256) for optimal viewing.

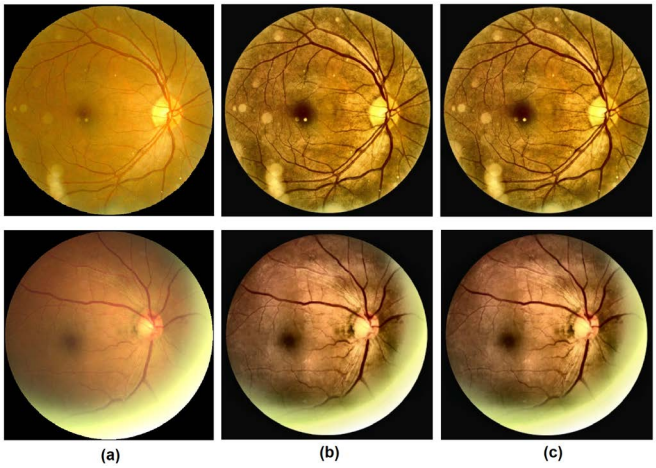

Fig. 4: *The sample images of the pre-processing (a) normal images, (b) CLAHE images, and (c) Gaussian filter images from the ODIR dataset.*

PROPOSED-ENSEMBLE MODEL

The term "ensembles of models" refers to merging the results of several different statistical models into a single comprehensive prediction. This paves the way for the model's capabilities to reflect the improved estimations more properly and achieve greater performance than any single contributing model. In this study, the proposed framework, shown in Figure 2, is carried out in three-steps processes for the automatic diagnosis of ophthalmological disease from fundus images. In the first step, our proposed CNN model with three well-known pre-trained models, including VGG-16 (Simonyan and Zisserman, 2014), ResNet50 (He *et al.*, 2016), and DenseNet201 (Jaiswal *et al.*, 2021), respectively, are employed as feature extractors. In the second step, the eye disease images are extracted by the proposed CNN, VGG-16, ResNet50, and DenseNet201. These features are acquired before the softmax layer of each model. Obtaining these features via traditional techniques is challenging and requires great expertise, increasing the overall cost.f

On the other hand, feature extraction is performed automatically using deep learning architectures quickly. However, the final FC layer of each model, including our proposed model, yielded a total of 4000

features when concatenated. Since distinct low-level features (redundant features) are acquired separately utilizing three well-known pre-trained models with the proposed CNN, an excessive number of duplicate features might reduce classification accuracy and increase execution time. Therefore, to identify the most compelling features, the mRMR method (Yan and Jia, 2019) is utilized in the third step. After that, 400 efficient features are chosen randomly by the mRMR method and fed into the Support Vector Machines (SVM) (Soumaya *et al.*, 2021) classifier for the classification with 5-fold cross-validation. SVM is a mechanism for conceptually separating data from many classes in the most efficient manner. Decision boundaries, or hyperplanes, are established for this purpose. SVM can produce good outcomes in high-dimensional spaces while also making optimal use of memory. The proposed CNN model for this work and the mRMR feature selection algorithm are described in detail below.

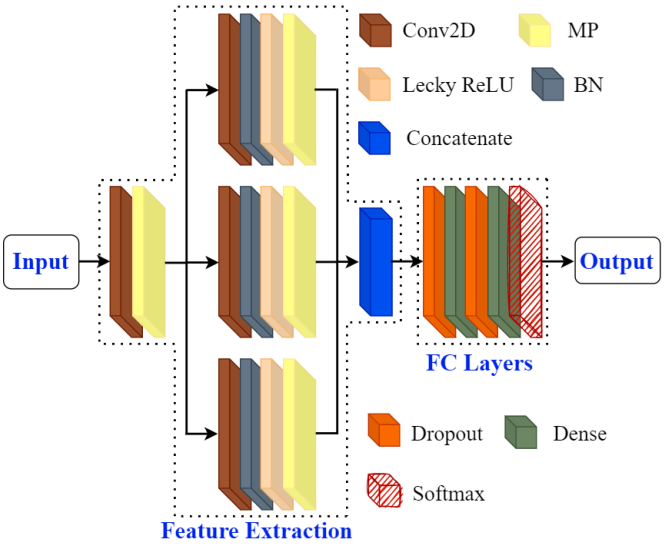

Fig. 5: *The design of our suggested CNN model.*

Proposed CNN Model

CNN's are the most widely used deep learning algorithms for training medical images to classify anomalies in medical images (Chan *et al.*, 2020). This is because CNN keeps distinguishing features when evaluating input images. Regarding eye disease images, spatial connections are critical in retinal pictures, such as where blood vessels begin to collapse and how yellow liquid accumulates around the retinal area. Sometimes, the pre-trained CNN models are not appropriately suited for medical images. In addition, their computational cost (i.e., parameters, execution time) is excessively high due to the employment of several layers. Because the pre-trained models were created for certain particular scenarios, such as object identification (Dhillon and Verma, 2020) and normal image classification (Zaidi *et al.*, 2022). Considering this, we introduce a deep CNN model to recognize and classify eight different types of eye diseases. Fig. 5 depicts the proposed architecture for the model. This model has 20 layers, some of which are used for feature extraction and classification. These layers include conv (convolution), MP (max-pooling), BN (batch normalization), dropout, and DL (dense layers). The feature extraction and classification modules include 15 and 5 layers, correspondingly. The phase of feature extraction and the processes for classification are merged in deep learning-based techniques, but in manual feature extraction methods, these two phases are kept distinct from one another.

As the first step in training a CNN, this network takes input images. To improve its effectiveness, this model starts with an input layer size of 256×256 and gradually adds more and more CNN layers. The hyperparameter may be adjusted using a variety of settings. For the first convolutional layer, we use 32 filters, a 3×3 kernel size, and a zero Padding layer. Then, we use a multi-protocol (MP) layer. The output from this feature is split into three separate convolutional layers. A convolutional layer, a BN layer, a Leaky ReLU (LR), and a Max Pooling (MP) layer make up each convolutional layer. It uses an LR activation function with MP and BN layers after each convolution layer. When using an MP layer with a stride of 2, the data representation size is significantly reduced. Since a higher pixel count is correlated to further parameters, which need massive amounts of data, this mostly works to shrink the picture size. Convolutional layer 2 is shrunk from 256×256 to 128×128, but filter size and stride remain the same at 16 filters each. We use the same stride and filter size for the third convolutional layer but increase the size to 64×64. Completion, adaptation, overfitting, and effectiveness on unexplored features are all enhanced when BN layers are utilized as regularization layers.

The results from these three units are combined and sent to fully connected (FC) layers for classification. The purpose of these stages (flattening, dense, dropout) is to detect visual problems. The filtered characteristics of eye diseases are collected using two sets of dense and dropout layers, with the dense layers specified by 256 and 512 flattened neurons. To prevent the model from overfitting, 70% and 50% of the neurons in the hidden layers are set to 0 at every iteration during the training step, respectively. The proposed CNN model's settings are shown in Table 1. Finally, the SVM is used in place of softmax to categorize eight distinct disorders.

Table 1: *The settings of the proposed CNN model*

Blocks	Operator	Filters	Configuration	Stride	Output Shape
1	Conv2D	32	KS: 3x3; LeckyReLU	-	256x256x32
	MP	-	KS: 2x2	2	128x128x16
2	Conv2D	64	KS: 1x1; LeckyReLU	-	128x128x64
	BN	-	KS: 2x2	-	128x128x64
	MP	-	KS: 2x2	2	56x56x64
3	Conv2D	32	KS: 3x3; LeckyReLU	-	128x128x32
	BN	-	KS: 2x2	-	128x128x32
	MP	-	KS: 2x2	2	56x56x32
4	Conv2D	96	KS: 7x7; LeckyReLU	-	128x128x96
	BN	-	KS: 2x2	-	128x128x96
	MP	-	KS: 2x2	2	56x56x48
5	Concate	-	-	-	-
FC	Flatten	-	-		-
	Dropout	-	0.7	-	-
	Dense	-	512; LeckyReLU	-	512
	Dropout	-	0.4	-	512
	Dense	-	256; LeckyReLU	-	256
	Softmax	-	8	-	8

Table 2: *The performance of the suggested technique for the extraction of features in comparison to the efficacy of four other methods for the extraction of features (VGG-16, Inception-v3, ResNet-50, and EfficientNet-B0) in terms of accuracy (Acc), sensitivity (Sen), specificity (Spec), precision (Pre), kappa (Kp) and F1-score (FS).*

Methods	Acc (%)	Sen (%)	Spec (%)	Pre(%)	Kp (%)	FS (%)
DenseNet201	89.49	82.49	81.56	88.79	86.02	85.11
VGG16	90.28	83.82	81.24	88.95	86.19	85.76
ResNet-50	91.20	84.16	83.62	89.05	87.26	86.98
Proposed CNN	93.81	85.77	85.09	91.63	88.01	**89.51**
Proposed Ensemble	**96.96**	**87.38**	**86.81**	**92.77**	**89.01**	89.35

mRMR feature selection

Regarding classification accuracy, feature selection is one of the most crucial aspects. The term ”feature selection” refers to the procedure of determining from among a large number of features that improve the effectiveness of a decision support system. The selection of features used in CNN designs helps to decrease workload, classification error, and noise by detecting better and more distinguishing features while also enhancing the efficiency of a CNN model. Some popular feature selection algorithms (Boutsidis *et al.*, 2008; Robnik-Šikonja and Kononenko, 2003; Raghu and Sriraam, 2018), for example, the LDA and PCA (Boutsidis *et al.*, 2008) feature selection algorithms work well only when applied to linear feature sets. On the other hand, ReliefF (Robnik-Šikonja and Kononenko, 2003) is a distance-based method for selecting features that provide both negative and positive weights. In addition, features with a negative weight from ReliefF are considered redundancy features.

On the other hand, the most frequently employed feature selection approach, NCA (Raghu and Sriraam, 2018), is a weight-based algorithm for generating positive weights. As a result, ReliefF and NCA are applied together to remove the redundant features. However, applying both models one after another is quite time-consuming. In contrast, the mRMR approach (Yan and Jia, 2019) has been presented to reduce the calculation costs while improving accuracy. Its goal is to find the best feature subset based on a trade-off between relevance and repetition among features and between feature and class labeling. This strategy is effective because it maximizes the relevance (L), also known as the mutual information (I), between the class and the features while simultaneously reducing the redundancy (D), also known as the average I between the features. The values of L and D can be determined by using eq. (5) and (6), respectively.

Table 3: *The performance of the suggested technique for the extraction of features in comparison to the efficacy of four other methods for the extraction of features (VGG-16, Inception-v3, ResNet-50, and EfficientNet-B0) in terms of accuracy (Acc), sensitivity (Sen), specificity (Spec), precision (Pre), kappa (Kp) and F1-score (FS).*

Methods	Acc (%)	Sen (%)	Spec (%)	Pre(%)	Kp (%)	FS (%)
DenseNet201	89.49	82.49	81.56	88.79	86.02	85.11
VGG16	90.28	83.82	81.24	88.95	86.19	85.76
ResNet-50	91.20	84.16	83.62	89.05	87.26	86.98
Proposed CNN	93.81	85.77	85.09	91.63	88.01	**89.51**
Proposed Ensemble	**96.96**	**87.38**	**86.81**	**92.77**	**89.01**	89.35

$$L(fs) = \frac{1}{|fs|} \sum_{a=1}^{|fs|} I(f_a, cl) \tag{2}$$

$$D(fs) = \frac{1}{|fs|^2} \sum_{a=1}^{|fs|} \sum_{b=1}^{|fs|} I(f_a, f_b) \tag{3}$$

where fs stands for the feature subset, cl for the class label (targeted value), and $|fs|$ is the size of the feature subset, and f_a and f_b stand for the ath and bth features, respectively. I(f_a,cl) represents the I value between the f_a feature and class cl. Similarly, I (f_a,f_b) is the I value between the f_a and f_b features. More thorough details about mRMR can be obtained from (Ding and Peng, 2005).

In this experiment, 6426 eye disease images are utilized across eight classes. The dimension of the feature vector derived from each model is 6426 $\times$ 1000. When these features are combined by three pre-trained models and our proposed model, a total of 6426 $\times$ 4000 feature vectors are produced. During this feature combination process, the number of features does not decrease. Each architecture's features are combined side by side, not over the other. This approach aims to integrate numerous feature subsets of different models to find an ideal subset of features to increase classification accuracy. After that, the mRMR feature selection approach is employed to optimize the features (randomly selects 400 features (100 features from each model) randomly). The pictures of eye diseases are classified into their respective classes in the last stage.

EVALUATION METRICS

Confusion Matrix (Jeny *et al.*, 2020; Yang *et al.*, 2021) provides six assessment criteria to evaluate our suggested method's success. The six evaluation metrics are accuracy ($\frac{TrPv+TrNv}{TrPv+TrNv+FlPv+FlNv}$), Recall/Sensitivity ($\frac{TrPv}{FlNv+TrPv}$), Precision ($\frac{TrPv}{TrPv+FlPv}$), F1-score ($2 \times \frac{Precision \times Recall}{Precision+Recall}$), Specificity ($\frac{TrNv}{FlPv+TrNv}$), Kappa ($k = \frac{2\times(TrPv\times TrNv - FlNv\times FlPv)}{(TrPv+FlPv)\times(FlPv+TrNv)+(TrPv+FlNv)\times(FlNv+TrNv)}$), and the AUC for the Receiver Operating Characteristic (ROC) curves [20], respectively. In this case, a True Positive (TrPv) comprises instances in which the forecast and actual result match. When our forecast of "no" matches the actual result ("no"), we have a true negative (TrNv). We have a false positive if we expect a YES and get a NO (FlPv). Finally, a false negative occurs when our prediction is NO, but the actual result is YES (FlNv).

RESULTS AND DISCUSSIONS

EXPERIMENTAL DETAILS

All of the experiments are conducted with the help of PyTorch (Kochgaven *et al.*, 2021) on a computer with an Intel Core i9-10850K CPU operating at 3.60 GHz, 64 GB of RAM, and an Nvidia GeForce RTX 2070 Super GPU that has 8 GB of memory. During the testing, the ADAM optimizer (Chauhan *et al.*, 2021) is used. A momentum of 0.8 with a learning rate of six times 10-4 and a mini-batch size of 32 with 80 epochs are employed. The RGB photos included in the input have a size of 256 $\times$ 256 pixels on each side.

Table 4: *According to levels of accuracy (Acc), sensitivity (Sen), and specificity (Spec), for every eye disease class by our proposed approach*

Disease Classes	Acc (%)	Sen (%)	Spec (%)
Normal	97.24	87.08	93.75
Glaucoma	96.43	82.96	89.98
Diabetics Retinopathy	99.76	87.85	83.59
AMD	95.64	89.17	81.57
Hypertension	96.24	88.99	85.28
Cataract	97.42	90.24	84.79
Myopia	95.88	87.79	91.03
Other abnormalities	97.13	85.36	84.41

RESULTS AND COMPARISON

Fig. 6 depicts a visual representation of our proposed method to find eight different types of eye diseases. From the examples, it can be observed

Table 5: *The comparison between our proposed approach with other existing approaches in terms of AUC, F1-Score (FS), and Kappa score (KS).*

Author	Method	AUC	F1-Score (FS)	Kappa (KS)
Islam *et al.* (2019)	CNN	80.50	85.00	31.00
Jordi *et al.* (2019)	Inception V3	88.38	89.84	51.86
Li *et al.* (2020)	VGG16	86.81	87.30	44.94
Wang *et al.* (2020)	EffifiNetB3	67.00	85.00	43.00
Gour and Khanna (2021)	Two I/PVGG16	84.93	85.57	-
He *et al.* (2021)	ResNet-101	93.0	91.30	63.70
Yang and Yi (2022)	DSRA-CNN	93.0	88.16	88.67
Proposed CNN (Ours)	SVM	92.00	**89.51**	88.01
Proposed-Ensemble	**SVM**	**95.90**	89.35	**89.01**

that our proposed model can effectively classify eye disease into eight distinct groups, each of which has a significant difference in score.

After analyzing the proposed-ensemble approach on the test set, the results were recorded separately in Table 4 for each of the eight eye disease classes regarding the accuracy, sensitivity/recall, and specificity value combinations. The class for diabetic retinopathy, cataracts, and myopia have the best accuracy, sensitivity, and specificity values, which are 99.76%, 90.24%, and 91.03%, respectively. Myopia also achieves the maximum value of 91.03%. The lowest order, on the other hand, is AMD, Glaucoma, and Diabetic Retinopathy. However, we achieve excellent values in all evaluation matrices, indicating that our proposed technique can detect eye illness images more accurately.

Table 3 compares the effectiveness of our method with three pre-trained model types, along with VGG-16, DenseNet201, and ResNet-50, with our proposed CNN model to show that it is capable of classifying the eight eye diseases in terms of accuracy, sensitivity, specificity, precision, F1-score, and Kappa score. We can see from this table that the proposed-ensemble approach outperforms VGG-16, DenseNet201, ResNet-50, and the proposed CNN with regards to accuracy (99.96%), sensitivity (87.38%), precision (92.77%), specificity (86.81%) and kappa (89.01%) except f1-score. Regarding the f1-score, our proposed CNN model acquires more scores than others, which is 89.51%. In addition, we can define that our suggested CNN model exhibits a comparable performance with the proposed-ensemble model.

The effectiveness of our ensemble classifier is assessed and compared to three pre-trained models (DenseNet201, VGG-16, and ResNet-50) and the proposed CNN model using the Receiver Operating Characteristic (ROC) Curve (Junayed *et al.*, 2021a) to ascertain how well it performs in Fig. 8. The false positive rate (FPR) and the true positive rate (TPR), respectively, at various thresholds, are shown along the x-axis and the y-axis of the ROC curves. The AUC (area under the curve) shows the identified data capacity by assessing whether the curve is located up or down the dashed line. According to the ROC curve, the highest AUC is achieved by the proposed-ensemble model which is around 22% (0.959 vs. 0.752), 11% (0.959 vs. 0.858), 5% (0.959 vs. 0.909), and 4% (0.959 vs. 0.920) more than DenseNet201, VGG-16, ResNet-50, and proposed CNN, respectively.

Fig. 7 (a) and (b) demonstrate the accuracy as well as loss during training and validation, correspondingly, for the optimum number of epochs (80 epochs) and batch size (32) for our proposed-ensemble technique. The blue line shows the training results (accuracy and loss) in Fig. 7 (a, b), while the orange line represents the results of the validation (accuracy and loss). Following a total of 80 epochs, the training accuracy is 97.16%, and the validation accuracy is 96.96%, while the training and validation loss are 0.31 and 0.33.

Table 6: *Comparison of three pre-trained models with the proposed CNN model in various variables (size, parameters, layers, epochs, run-time, and accuracy).*

Variables	DenseNet201	ResNet50	VGG16	Proposed CNN
Model Size	77MB	98MB	528MB	**27MB**
Parameters	20M	25.6M	138M	**1.749M**
Layers	201	177	41	**20**
Epochs	200	200	200	**80**
Accuracy	89.49%	91.20%	90.28%	**93.81%**

To show the effectiveness of our suggested CNN model, we display a comparison among three pre-trained models, i.e., DenseNet201, ResNet50, and VGG16, and our proposed CNN model in Table 6 regarding various factors (e.g., size, parameters, and so on). For every model, the feature is set to 100 (randomly). From this Table 6, it can be seen that our model shows 93.81% accuracy by utilizing

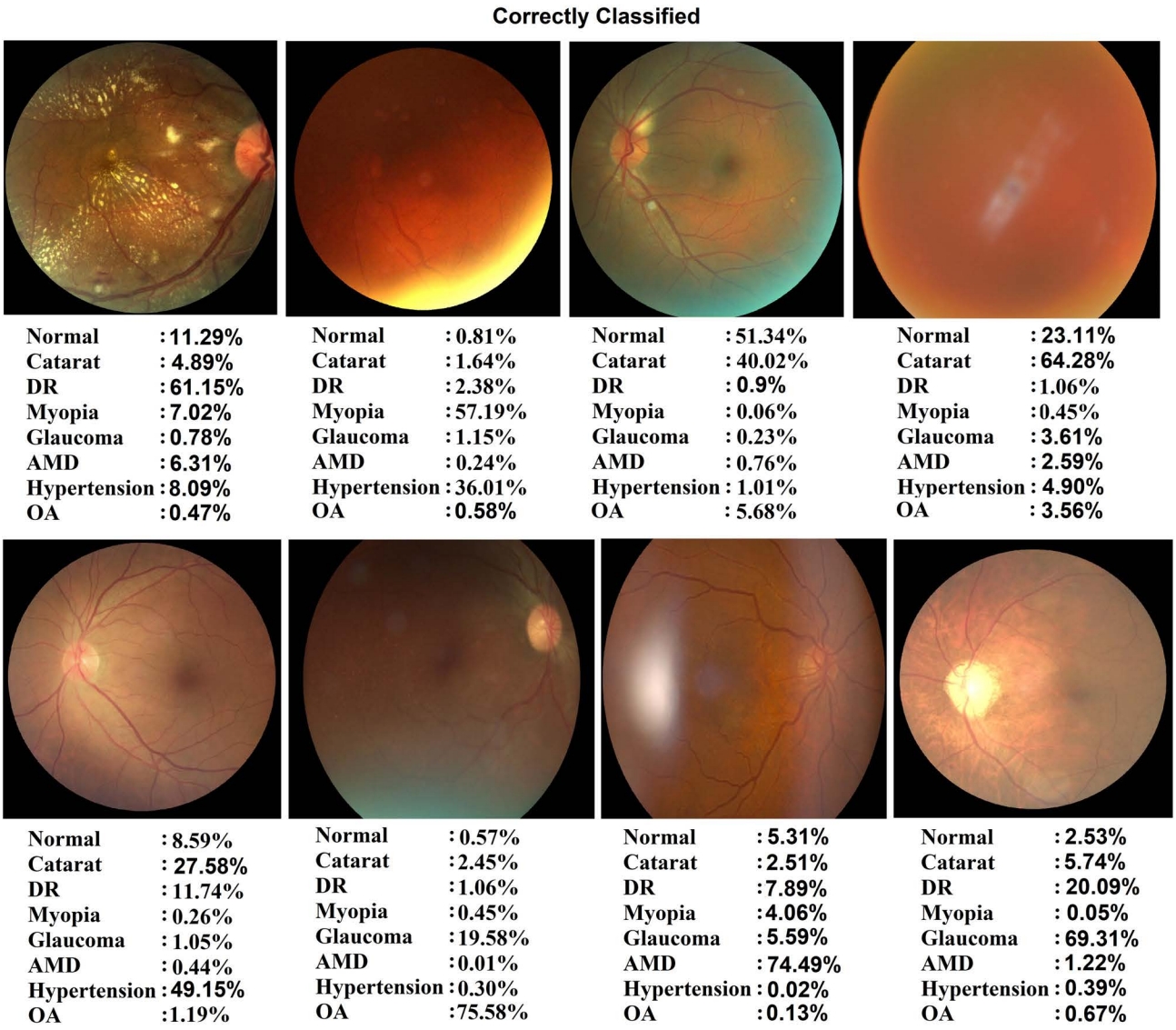

Fig. 6: *A graphical representation of the correctly classified cases using our proposed-ensemble model to detect eight different types of eye disease.*

only 80 epochs; on the other hand, all three pre-trained models need 2.5 times more (200 epochs). In contrast, our model size and parameters are only 27 MB and 1.749M, respectively, reflecting our method's superiority. Furthermore, even in terms of layers, our proposed CNN model needs only 20 layers which are approximately 10 (201 vs. 20), 9 (177 vs. 20), and 2 (41 vs. 20) times less than DenseNet201, ResNet50, and VGG16. Thus, it can be inferred that our suggested CNN model plays a critical and unrivaled role in detecting and classifying the eight various kinds of eye diseases.

COMPARATIVE STUDIES

Table 5 shows the AUC, F-score, and Kappa score findings of the ODIR dataset for the proposed technique and current approaches. According to Table 5, our proposed-ensemble model outperforms others (Islam *et al.*, 2019; Jordi *et al.*, 2019; Li *et al.*, 2020; Wang *et al.*, 2020; Gour and Khanna, 2021; He *et al.*, 2021; Yang and Yi, 2022) in terms of AUC, F1-Score, and Kappa score. However, all the articles utilized the pre-trained CNN models for classifying eye diseases except (Islam *et al.*, 2019; Yang and Yi, 2022). Islam *et al.* (2019) suggested a shallow CNN model trained from scratch for automated ocular illness detection from fundus pictures, showing quite minimal performance compared to other existing approaches. The CNN model we presented has an AUC of 92.00%, and a Kappa score of 88.01%, both of which are higher than (Islam *et al.*, 2019). Furthermore, when compared to the Gour and Khanna (2021), our

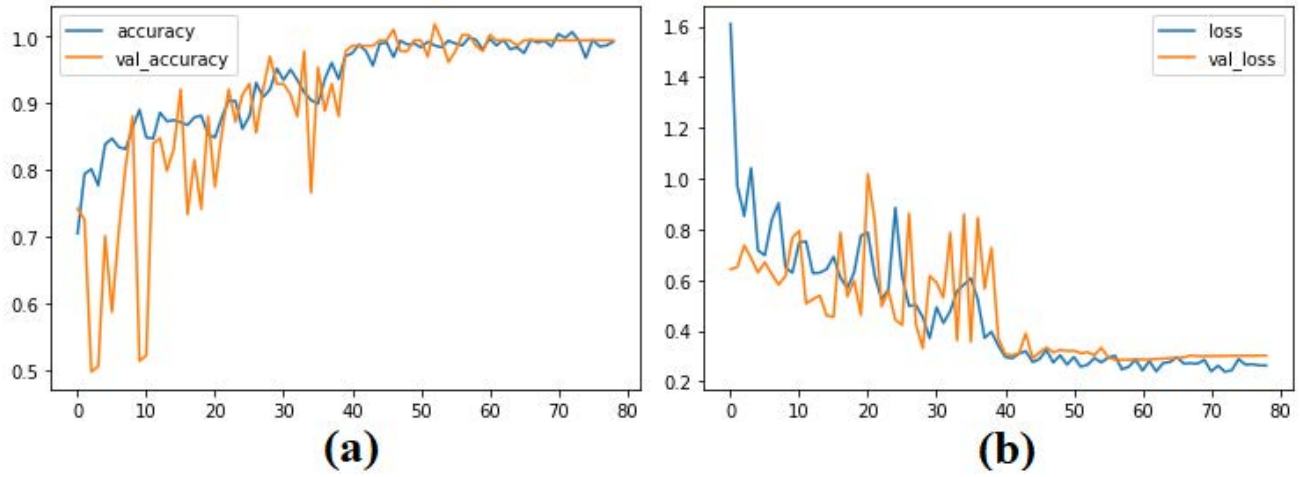

Fig. 7: *(a) The proposed-ensemble model's accuracy graph, and (b) its loss graph.*

proposed-ensemble model demonstrated around 11% (95.90 vs. 84.93) and 4% (89.35 vs. 85.57) more AUC, and F1-Score, respectively. Even our approach has a more AUC, F1-Score, and Kappa score than the current approach (Yang and Yi, 2022) ((95.90 vs. 93.00), (89.35 vs. 88.16), and (89.01 vs. 88.67)). Furthermore, it is essential to highlight that our proposed CNN model gets more F1-Score than our proposed-ensemble model (89.51 vs. 89.35).

Table 7: *The accuracy results for the comparative feature selection algorithms (NCA, ReliefF, NCAR, and mRMR) using Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM) classifiers.*

Classifiers	NCA	ReliefF	NCAR	mRMR
DT	80.15	81.28	81.95	**85.35**
KNN	78.76	79.87	80.35	**84.96**
NB	75.85	76.34	76.94	**87.90**
RF	88.56	88.94	89.34	**92.29**
SVM	89.28	89.35	89.54	**96.96**

ABLATION STUDIES

Some ablation studies are investigated to prove the feasibility of our proposed-ensemble model. In Table 7, the accuracy is represented for four feature selection algorithms (NCA, ReliefF, NCAR, and mRMR) employing five classifiers, including Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM) (Gayathri *et al.*, 2021; Lamba *et al.*, 2021; Ibrahim and Abdulazeez, 2021). In this experiment, the number of chosen features is set to 400 for all feature selection techniques. According to Table 7, the SVM classifier outperforms all other classifiers in terms of performance for all feature selection techniques. In contrast, the NCAR feature selection algorithm surpassed all different feature selection algorithms in terms of performance for all classifiers. Furthermore, the highest accuracy is achieved by the SVM classifier and the mRMR method, which is 96.96%. In contrast, the lowest accuracy is obtained by the NB classifier and the NCA algorithm, which is 75.85%, respectively.

Table 8: *The ablation study in terms of accuracy with or without the mRMR feature selection algorithm*

Approaches	Features	No feature Selection	Feature Selection
ResNet50-SVM	1000	0.899	0.912
DenseNet201-SVM	1000	0.871	0.894
VGG16-SVM	1000	0.896	0.902
Proposed model-SVM	1000	0.915	0.938
Proposed-Ensemble-SVM	4000	0.946	**0.969**

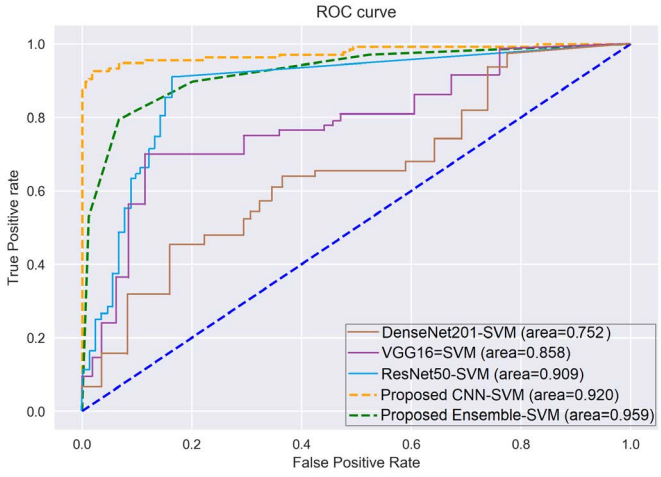

Fig. 8: *The efficiency of the ROC curves for the suggested-ensemble model using three pre-trained models (DenseNet201, VGG-16, and ResNet-50), in addition to the proposed CNN model.*

Another experiment is carried out with or without the mRMR feature selection algorithm in Table 8. In

this experiment, 100 features are assigned to each pre-trained model and the suggested model out of 1000 features. Similarly, 400 features are chosen randomly from 4000 features for the proposed ensemble model. From Table 8, it can be observed that without the mRMR algorithm, the accuracy is decreased for every model. For example, without feature selection, the accuracy of the proposed CNN model has a gap of 2.3% (93.8% vs. 91.5%). Similarly, we can achieve an accuracy of 96.9% with the suggested ensemble model by using the mRMR method (without 94.6%). Thus, it can be concluded that the mRMR method significantly influences the improvement of classification accuracy.

In Table 9, we display another experiment to show the effectiveness of our proposed CNN model, the compelling features, and the proposed-ensemble model. In this experiment, two, three, and four pre-trained models, including the proposed CNN model, are utilized successively. From this experiment, it can be said that the performance is not as best. Also, inference time (IT) is higher when we utilize two or three pre-trained models together (cases 1, 2, 4, and 7) as compared to when we include our suggested CNN model (cases 3, 5, 6, 8, and 9). For instance, when three pre-trained models are ensembled in cases 7, 8, and 9, the accuracy is still lower, and the IT is higher than when our suggested CNN model is included (91.74 vs. 95.02 vs. 94.77 and 3.126 vs. 2.089 vs. 2.131). Finally, in case 10, we integrated our suggested CNN model with the three pre-trained models and got the highest accuracy rate of 96.96%. Thus it can be summarized that our proposed CNN model has more prominent features; therefore, it demonstrates a more significant performance than the pre-trained models in terms of enhancing the accuracy of eye diseases.

FAILURE CASES

Fig. 9 shows a graphical illustration of the misclassification of images from different types of eye disease using our proposed-ensemble approach. Though our algorithm can correctly classify eye diseases into eight groups with a high score, there are a few misclassification situations, as shown in Fig. 9. However, when comparing classifying ratings, our method offers a few differences in scores (most of the time, the second-highest score). For example, the last row image was supposed to be the Diabetic Retinopathy (DR) class (score showed 34.88%), but it is classified in the Glaucoma class with a score of 50.67%. Drawing conclusions from a thorough examination of the misclassification images, the suggested approach has difficulty identifying images with shadows, intense light, poor clarity, or poor image quality. It will be considered with a large dataset with more classes in future work.

CONCLUSION

This article introduced an automated deep learning-based ensemble method to detect and classify eye diseases from fundus images. In this method, a deep CNN-based 20 layers model is proposed to lessen the burden on the computer's processing power, which adds the activation function, optimization algorithm, loss function, kernel sizes, and batch size. The pre-processing steps, such as CLAHE, Gaussian filter, and augmentation technique, are applied to remove the noise and enhance the number of images. Three pre-trained CNN models are used to compare and evaluate the proposed CNN model. To investigate the performance of the proposed model, the ablation study is performed with different cases. Experimental findings reveal that the ensemble strategy outperforms current state-of-the-art approaches in evaluation matrices. In the future, we want to assess our proposed CNN method using large-scale datasets to analyze and modify so that it can identify additional types of diseases.

DECLARATIONS

FUNDING

This work is partially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the 2232 Outstanding Researchers program, Project No. 118C301. Research and its contents are solely the authors' responsibility and do not necessarily represent the official view of the funding organizations. The funders had no role in study design, data analysis, algorithmic design, the decision to publish, or the preparation of the manuscript.

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies with human participants and/or animals performed by any of the authors.

CONFLICT OF INTEREST

We (authors) certify that there is no actual or potential conflict of interest related to this article.

AUTHOR'S CONTRIBUTIONS

Afsana Ahsan Jeny, Masum Shah Junayed: Methodology, Software, Validation, Writing-Original draft preparation. **Md Baharul Islam:** Conceptualization, Investigation, Supervision, Writing- Reviewing and Editing.

Table 9: *The classification accuracy and inference time acquired by our proposed CNN model, three pre-trained models, and mRMR method*

Case	Models	Features (mRMR)	IT (s)	Acc (%)
1	DenseNet201+VGG16	(100+100)=200	2.428	90.38
2	DenseNet201+ResNet50	(100+100)=200	2.879	91.15
3	DenseNet201+Proposed CNN	(100+100)=200	1.992	93.51
4	VGG16+ResNet50	(100+100)=200	2.779	91.58
5	VGG16+Proposed CNN	(100+100)=200	1.762	94.13
6	ResNet50+Proposed CNN	(100+100)=200	1.428	94.96
7	DenseNet201+VGG16+ResNet50	(100+100+100)=300	3.126	91.74
8	VGG16+ResNet50+Proposed Model	(100+100+100)=300	2.089	95.02
9	ResNet50+DenseNet201+Proposed CNN	(100+100+100)=300	2.131	94.77
10	DenseNet201+VGG16+ResNet50+Proposed CNN	(100+100+100+100)=400	3.451	**96.96**

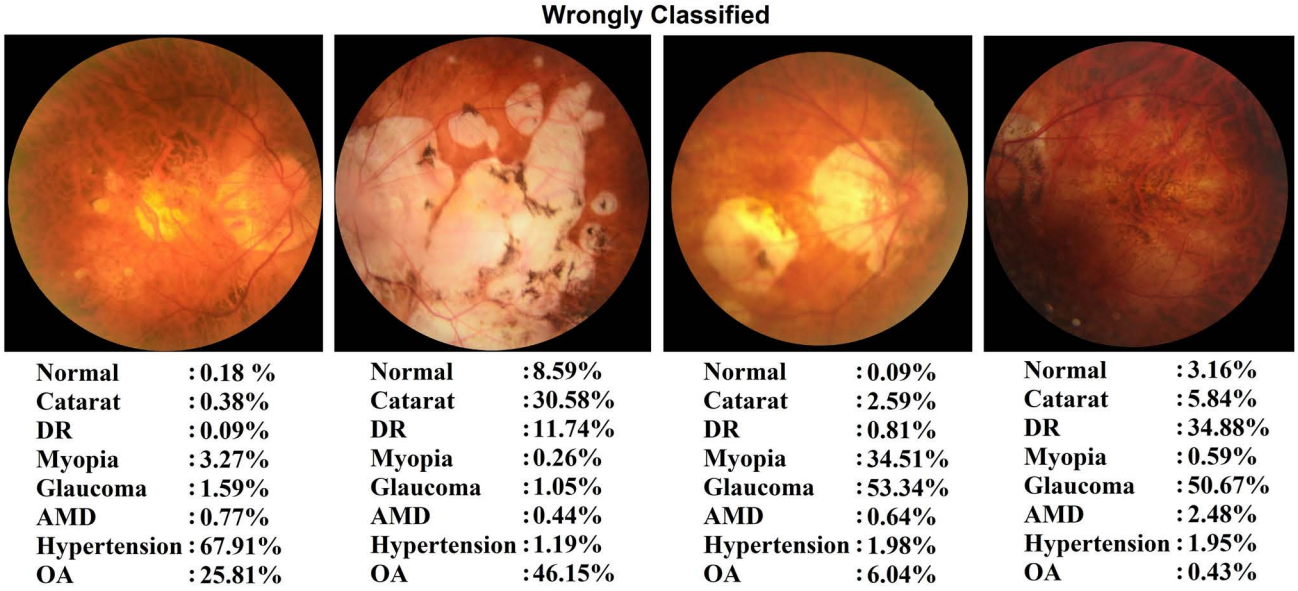

Fig. 9: *A graphical representation of the failure cases using our proposed-ensemble model to detect different types of eye disease.*

REFERENCES

Ayana G, Park J, Jeong JW, Choe Sw (2022). A novel multistage transfer learning for ultrasound breast cancer image classification. Diagnostics 12:135.

Boutsidis C, Mahoney MW, Drineas P (2008). Unsupervised feature selection for principal components analysis. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining.

Chai Y, Liu H, Xu J (2018). Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. Knowledge Based Systems 161:147–56.

Chan HP, Samala RK, Hadjiiski LM, Zhou C (2020). Deep learning in medical image analysis. Deep Learning in Medical Image Analysis :3–21.

Chauhan T, Palivela H, Tiwari S (2021). Optimization and fine-tuning of densenet model for classification of covid-19 cases in medical imaging. International Journal of Information Management Data Insights 1:100020.

Chen J, Wan Z, Zhang J, Li W, Chen Y, Li Y, Duan Y (2021). Medical image segmentation and reconstruction of prostate tumor based on 3d alexnet. Computer methods and programs in biomedicine 200:105878.

Costagliola C, Dell'Omo R, Romano MR, Rinaldi M, Zeppa L, Parmeggiani F (2009). Pharmacotherapy of intraocular pressure: part i. parasympathomimetic, sympathomimetic and sympatholytics. Expert opinion on Pharmacotherapy 10:2663–77.

Dhillon A, Verma GK (2020). Convolutional neural

network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence 9:85–112.

Ding C, Peng H (2005). Minimum redundancy feature selection from microarray gene expression data. Journal of bioinformatics and computational biology 3:185–205.

Gaur L, Bhatia U, Jhanjhi N, Muhammad G, Masud M (2021). Medical image-based detection of covid-19 using deep convolution neural networks. Multimedia systems :1–10.

Gayathri S, Gopi VP, Palanisamy P (2021). Diabetic retinopathy classification based on multipath cnn and machine learning classifiers. Physical and engineering sciences in medicine 44:639–53.

Gour N, Khanna P (2021). Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. Biomedical Signal Processing and Control 66:102329.

Gupta N, Hitendra G, Rohit A (2022). A robust framework for glaucoma detection using clahe and efficientnet. The Visual Computer 38:2315–28.

Han B, Du J, Jia Y, Zhu H (2021). Zero-watermarking algorithm for medical image based on vgg19 deep convolution neural network. Journal of Healthcare Engineering 2021.

He J, Li C, Ye J, Qiao Y, Gu L (2021). Multi-label ocular disease classification with a dense correlation deep neural network. Biomedical Signal Processing and Control 63:102167.

He K, Zhang X, Ren S, Sun J (2016). Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition.

Ibrahim I, Abdulazeez A (2021). The role of machine learning algorithms for diagnosing diseases. Journal of Applied Science and Technology Trends 2:10–9.

Islam MT, Imran SA, Arefeen A, Hasan M, Shahnaz C (2019). Source and camera independent ophthalmic disease recognition from fundus image using neural network. In: 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON). IEEE.

Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M (2021). Classification of the covid-19 infected patients using densenet201 based deep transfer learning. Journal of Biomolecular Structure and Dynamics 39:5682–9.

Jeny AA, Sakib ANM, Junayed MS, Lima KA, Ahmed I, Islam MB (2020). Sknet: A convolutional neural networks based classification approach for skin cancer classes. In: 2020 23rd International Conference on Computer and Information Technology (ICCIT). IEEE.

Jonas JB, Bourne RR, White RA, Flaxman SR, Keeffe J, Leasher J, Naidoo K, Pesudovs K, Price H, Wong TY, *et al.* (2014). Visual impairment and blindness due to macular diseases globally: a systematic review and meta-analysis. American journal of ophthalmology 158:808–15.

Jordi C, Joan Manuel N, Carles V (2019). Ocular disease intelligent recognition through deep learning architectures. Universitat Oberta de Catalunya Barcelona Spain .

Junayed MS, Islam MB, Jeny AA, Sadeghzadeh A, Biswas T, Shah AS (2021a). Scarnet: Development and validation of a novel deep cnn model for acne scar classification with a new dataset. IEEE Access .

Junayed MS, Islam MB, Sadeghzadeh A, Rahman S (2021b). Cataractnet: An automated cataract detection system using deep learning for fundus images. IEEE Access 9:128799–808.

Khan IA, Sajeeb A, Fattah SA (2020). An automatic ocular disease detection scheme from enhanced fundus images based on ensembling deep cnn networks. In: 2020 11th International Conference on Electrical and Computer Engineering (ICECE). IEEE.

Khan MSM, Ahmed M, Rasel RZ, Khan MM (2021). Cataract detection using convolutional neural network with vgg-19 model. In: 2021 IEEE World AI IoT Congress (AIIoT). IEEE.

Kochgaven C, Mishra P, Shitole S (2021). Detecting presence of covid-19 with resnet-18 using pytorch. In: 2021 International Conference on Communication information and Computing Technology (ICCICT). IEEE.

Lamba R, Gulati T, Alharbi HF, Jain A (2021). A hybrid system for parkinson's disease diagnosis using machine learning techniques. International Journal of Speech Technology :1–11.

Li F, Chen H, Liu Z, Zhang X, Wu Z (2019a). Fully automated detection of retinal disorders by image-based deep learning. Graefes Archive for Clinical and Experimental Ophthalmology 257:495–505.

Li N, Li T, Hu C, Wang K, Kang H (2020). A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In: International Symposium on Benchmarking, Measuring and Optimization. Springer.

Li T, Gao Y, Wang K, Guo S, Liu H, Kang H (2019b). Diagnostic assessment of deep learning algorithms

for diabetic retinopathy screening. Information Sciences 501:511–22.

ODIR-2019 (2022). International competition on ocular disease intelligent recognition. https://odir2019.grand-challenge.org/. [Online; accessed 05-January-2022].

Orfao J, Haar Dvd (2021). A comparison of computer vision methods for the combined detection of glaucoma, diabetic retinopathy and cataracts. In: Annual Conference on Medical Image Understanding and Analysis. Springer.

Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, Khan IA, Jadoon W (2019). A deep learning ensemble approach for diabetic retinopathy detection. IEEE Access 7:150530–9.

Raghavendra U, Fujita H, Bhandary SV, Gudigar A, Tan JH, Acharya UR (2018). Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Information Sciences 441:41–9.

Raghu S, Sriraam N (2018). Classification of focal and non-focal eeg signals using neighborhood component analysis and machine learning algorithms. Expert Systems with Applications 113:18–32.

Robnik-Šikonja M, Kononenko I (2003). Theoretical and empirical analysis of relieff and rrelieff. Machine learning 53:23–69.

Sahu S, Singh AK, Ghrera S, Elhoseny M, *et al.* (2019). An approach for de-noising and contrast enhancement of retinal fundus image using clahe. Optics Laser Technology 110:87–98.

Sarki R, Ahmed K, Wang H, Zhang Y (2020). Automated detection of mild and multi-class diabetic eye diseases using deep learning. Health Information Science and Systems 8:1–9.

Siciarz P, McCurdy B (2022). U-net architecture with embedded inception-resnet-v2 image encoding modules for automatic segmentation of organs-at-risk in head and neck cancer radiation therapy based on computed tomography scans. Physics in Medicine Biology .

Simonyan K, Zisserman A (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv14091556 .

Soumaya Z, Taoufiq BD, Benayad N, Yunus K, Abdelkrim A (2021). The detection of parkinson disease using the genetic algorithm and svm classifier. Applied Acoustics 171:107528.

Wang G, Gao Z, Zhang Y, Ma B (2018). Adaptive maximum correntropy gaussian filter based on variational bayes. Sensors 18:1960.

Wang J, Yang L, Huo Z, He W, Luo J (2020). Multi-label classification of fundus images with efficientnet. IEEE Access 8:212499–508.

WHO (2022). Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Online; accessed 15-January-2022].

Yan X, Jia M (2019). Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mrmr feature selection. Knowledge Based Systems 163:450–71.

Yang D, Martinez C, Visuña L, Khandhar H, Bhatt C, Carretero J (2021). Detection and analysis of covid-19 in medical images using deep learning techniques. Scientific Reports 11:1–13.

Yang Xl, Yi Sl (2022). Multi-classification of fundus diseases based on dsra-cnn. Biomedical Signal Processing and Control 77:103763.

Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B (2022). A survey of modern deep learning based object detection models. Digital Signal Processing :103514.