IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 September 2024, accepted 29 September 2024, date of publication 8 October 2024, date of current version 24 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3476412

RESEARCH ARTICLE

Eye Disease Detection Enhancement Using a Multi-Stage Deep Learning Approach

MD ZAHIN MUNTAQIM[1], TANGIN AMIR SMRITY[1], ABU SALEH MUSA MIAH[1], (Member, IEEE), HASAN MUHAMMAD KAFI[1], (Member, IEEE), TAOSIN TAMANNA[1], FAHMID AL FARID[2], (Member, IEEE), MD ABDUR RAHIM[3], HEZERUL ABDUL KARIM[2], (Senior Member, IEEE), AND SARINA MANSOR[2]

[1]Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology, Saidpur, Nilphamari 5311, Bangladesh
[2]Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia
[3]Department of Computer Science and Engineering, Pabna University of Science and Technology, Pabna 6600, Bangladesh

Corresponding authors: Sarina Mansor (sarina.mansor@mmu.edu.my), Abu Saleh Musa Miah (abusalehcse.ru@gmail.com), and Fahmid Al Farid (fahmid.farid@mmu.edu.my)

Multimedia University, Cyberjaya, Selangor, Malaysia (Grant Number: PostDoc(MMUI/240029)).

ABSTRACT Eye diseases, a significant global health concern, require timely detection to prevent vision loss. The alarming prevalence of eye diseases necessitates immediate action through early diagnosis, making it urgent to develop an automatic detection system. Many researchers have been working to develop such systems. Yet, existing solutions still face difficulties in achieving high-performance accuracy due to challenges like lacking feature effectiveness, high computational demands, and incomplete disease coverage. To overcome these challenges, we proposed a novel eye-disease detection system leveraging multi-stage deep learning technologies. In the study, we employed a preprocessing approach to ensure the system's robustness against rotation and translation, enhancing its effectiveness across varied conditions. Then, we employed a lightweight three-stage deep learning approach for extracting effective features and specific advantages. In the procedure, Stage 1 focuses on extracting fine-grained features using deep learning layers where the layers can automatically learn and identify complex patterns associated with various eye diseases, improving feature effectiveness and overall system accuracy. Then, we employed stage 2, which is constructed with two branches, each composed of convolutional blocks and identity blocks; this stage extracts hierarchical features by concatenating the outputs of the two branches. This hierarchical approach captures both low-level and high-level features, enhancing the extracted features' richness and robustness and leading to better classification performance. We concatenated the two branch features that fed into the classification module, producing a probabilistic eye disease presence map. By converting hierarchical features into precise disease predictions, this stage ensures accurate probabilistic outputs, aiding better decision-making and diagnosis. We evaluated the proposed model with OCT2017, Dataset-101, and Retinal OCT C8 datasets, demonstrating an accuracy improvement of up to 1% over existing state-of-the-art models in both multi-class and binary classification tasks. The lightweight design and reduced computational requirements of the model highlight its applicability for real-world deployment, particularly in resource-constrained environments. This computer-aided detection system offers a meaningful advancement in the field of automatic eye disease detection by providing a more accurate and efficient tool that can be deployed widely.

INDEX TERMS Eye disease classification, deep learning based classification, CNN-based classification.

I. INTRODUCTION

The eye is a sophisticated sensory organ that is crucial for our capacity to detect our surroundings. The global medical systems faces a significant problem due to the

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

increasing number of eye diseases affecting millions of individuals globally. The phrase ''eye diseases'' encompasses a broad spectrum of ocular issues, ranging from common refractive errors to severe conditions that result in visual impairment or complete loss of vision. The root causes and clinical symptoms of these diseases differ greatly. Choroidal neovascularization, Acrima, Cataract, Diabetic Macular Edema, Diabetic Retinopathy, Drusen, Glaucoma, Normal, Odir-5k, and Origa are a few common forms of eye diseases. Each disease has several different effects on the system of vision, as well as distinct signs and growth patterns. Cataracts are the most common cause of limited vision in most developing countries in South Asia; they account for most limited vision cases worldwide. In Bangladesh, approximately 750,000 people, including 40,000 children, suffer from limited vision, making it a significant health issue even though many of these cases are preventable with the right care and treatment [1], [2]. The World Health Organization has designated a region in Southeast Asia that includes Bangladesh that is home to 25% of the world's population as well as 33% of the 45 million limited vision people worldwide [3]. In this area, 80% of limited vision is caused by cataracts [2]. According to a recent prevalence survey [2], cataracts are responsible for 79.6% of cases of bilateral limited vision. Approximately 650,000 adults (95% Confidence Interval (CI) = 552 175 to 740 736) in Bangladesh are estimated to be limited vision as a result of cataracts among the adult population over 30 (roughly 44 million people). It is estimated that another 130,000 new cases arise every three years. Comparably, extrapolation indicates that 6.65 million (95% Confidence Interval (CI) 6.94 to 7.23) adults in one or both eyes have vision below 6/12 [2].

Optical coherence tomography (OCT) is a non-invasive, high-resolution, three-dimensional tomography imaging method that has been extensively utilized in the diagnosis of retinal diseases. During clinical analysis, ophthalmologists frequently use the morphology, thickness, and brightness of the retinal membrane structure in the OCT image to diagnose related diseases. This manual assessment procedure is laborious and frequently arbitrary. As such, methods for automatically identifying retinal OCT images are critical for the effective diagnosis and remote management of retinal disorders. Over the past 20 years, numerous algorithms have been developed for the classification of retinal OCT images. Additionally, and there aren't many publicly available benchmark datasets for assessing the current model and effectively training the deep learning model. Although dataset-101 [4] from kaggle is insufficient, The OCT dataset, which comprised 84,495 samples, was one of them. Kermany et al. [5] introduced it, and using deep learning, it produced a performance accuracy of 96%, which is low for implementing a classification system for eye diseases. The authors of papers [6], [7], [8], [9] aimed to deliver a favorable outcome by employing various techniques like iterative fusion CNN, multiscale multipath CNN, wide residual networks, and transfer learning. But they weren't successful in achieving high performance. The existing OCT image classification method only uses the features of the final convolutional layer for classification; it ignores the data from previous convolutional layers. To overcome the challenges, we proposed an eye disease classification system. Our key contributions are given below:

- **Novel Deep Learning Approach:** We introduce a novel, three-stage deep learning model designed specifically for eye disease detection. This model integrates carefully crafted convolutional and identity blocks, optimizing feature extraction and classification efficiency while maintaining a lightweight architecture. Our approach offers a significant departure from traditional methods, addressing key challenges such as inadequate feature effectiveness and high computational demands.
- **Robust Preprocessing and Augmentation:** To enhance the robustness of our system, we employ rigorous preprocessing and augmentation techniques. By implementing various image augmentation methods, we ensure that our system is rotation and translation-independent, thereby improving its adaptability across diverse datasets and real-world scenarios. This step is crucial for optimizing the performance and generalizability of our deep learning model.
- **Hierarchical Feature Extraction with Three-Stage Module:** We proposed a three-stage module. In Stage 1, we utilize convolutional layers, Batch Normalization, and activation layers to extract fine-grained features. Stage 2 enhances these features using two branches: Branch-1 consists of two convolutional blocks and four identity blocks, while Branch-2 comprises one convolutional block and two identity blocks. Both branches include global average pooling layers. The features from both branches are concatenated and fed into the classification module in Stage 3, which generates probabilistic maps for disease detection.
- **Superior Performance Evaluation:** Our proposed model undergoes rigorous evaluation on three benchmark datasets, OCT2017, Dataset-101, and Retinal OCT C8 for multi-class and binary classification tasks, respectively. The results demonstrate superior performance in terms of accuracy and efficiency compared to existing systems. Our model outperforms previous approaches, showcasing its potential to revolutionize automated eye disease detection and improve patient outcomes on a global scale.

II. LITERATURE REVIEW

Machine learning and deep learning have been used in numerous studies to predict eye diseases in the past [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]. We have thoroughly examined some related research papers and noted any shortcomings and mentioned them. Using deep learning and the local outlier factor (LOF) algorithm, the study [23] presented a novel automatic AMD detection method from

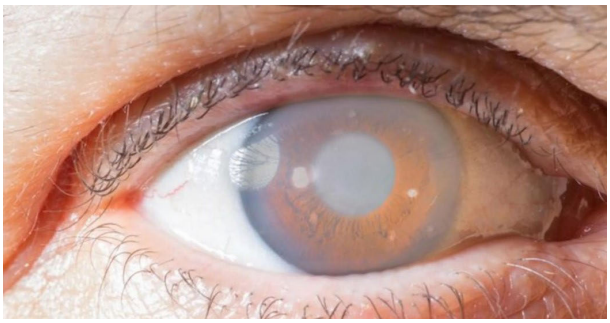

FIGURE 1. Glaucoma eye disorder adopted from [10].

optical coherence tomography (OCT) images. They used two publicly available datasets and got an accuracy of 99.87% on the UCSD dataset and 97.56% on the Duke dataset.

By determining the difference between the RPE layer and a second-order polynomial curve, Naz et al. [24] proposed an algorithm to identify OCT images affected by AMD. The RPE segmentation process was streamlined in time by employing an intensity-based threshold approach. Using a dataset comprising 25 AMD and 25 healthy images, the study detected AMD with an accuracy of 96.00%.

Arabi et al. [25] sampled the extracted layers and counted the number of white pixels in each sample after extracting the RPE layer using the binary threshold method. The number of pixels' mean values was determined and categorized. Using 16 images, they tested the method and were able to achieve a 75.00% accuracy rate.

For the purpose of detecting AMD from retinal OCT images, Thomas et al. [26] developed an algorithm based on RPE layer detection and baseline estimation using statistical techniques and randomization. A public dataset containing 2130 images was used to test the method, and the overall accuracy was 96.66%.

Sharif et al. [27] presented a method based on feature extraction and the support vector machine (SVM). First, the RPE layer was extracted by utilizing the graph theory dynamic programming technique, then a unique feature set consisting of features extracted from the difference signal of RPE and the inner segment outer segment layer of RPE was obtained. Finally, the SVM classifier was used to detect AMD-affected images from 950 OCT images,and an accuracy of 95.00% was obtained.

The authors of paper [7] proposes a transfer learning algorithm for analyzing medical images and achieving accurate diagnoses. Using a pre-trained deep learning model and optimizing it for medical image analysis is the main concept. Taking advantage of the model's pre-existing feature extraction capabilities, this technique modifies them to fit the unique domain of medical images. The study reports impressive results, achieving an accuracy of over 96.6% for OCT image classification tasks.

The study [8] presents a novel approach that uses Spectral-Domain Optical Coherence Tomography (SD-OCT) images to classify retinal diseases into multiple classes autometically. It uses the AOCT-NET convolutional neural network (CNN) architecture in conjunction with several image preprocessing techniques. The AOCT-NET model is trained on a sizable dataset of 109,309 SD-OCT images labeled for the five disease categories. With 95.30% overall accuracy, 95.20% sensitivity, and 98.40% specificity, the AOCT-NET model achieved remarkable classification accuracy.

The core architecture [28] of both SSL algorithms is a pre-trained ResNet-18 model. Originally trained on ImageNet, this pre-trained model extracts general feature representations useful for medical imaging domain transfer learning. The last layers of the pre-trained model are adjusted during training to suit the particular OCT image classification task better. Their method reached up to 99.69% accuracy.

III. DATASETS

To evaluate the proposed model, we used 3 benchmark publicly available datasets named OCT2017 [5], DATASET_101 [4], and Retinal OCT C8 [29] from Kaggle to conduct our experiments. Table 1 demonstrated the information of the datasets.

A. EYE DISEASE OCT2017 DATASET

Oct2017, illustrated in Figure 2, includes examples of three fundus diseases along with normal retina images. The OCT2017 dataset features images of three diseases: choroidal neovascularization (CNV), diabetic macular edema (DME), and Drusen, in addition to normal fundus images. This dataset contains 84,452 retinal OCT images across four classes: 83,484 images for training and 968 images for testing. Specifically, the training set consists of 36,205 CNV images, 10,348 DME images, 7,616 Drusen images, and 25,315 normal images, with an additional validation set of 1,000 images for each of the four classes [30]. Detailed information about the datasets is provided in Table 1.

B. EYE DISEASE 101 DATASET

There are two subsets in this dataset: train and test. Six classes are contained in each test and train directory. There are 9824 samples in the dataset. This dataset consisted of six classes, including Acrima, Origa, Odir-5K, Glaucoma, cataract, and retinal disease.

C. RETINAL C8 DATASET

The dataset consists of eight classes: AMD, CSR, DME, DR, DRUSEN, MH, and normal. The training dataset contains 15300 images in total; 6000 images are used for validation, and 2700 images are used for testing.

IV. PROPOSED METHODOLOGY

Our proposed methodology consists of four key steps, aligned with the deep learning standard code of practice, aiming to develop an efficient eye disease detection system. Figure 4 shows a diagram of our proposed methodology, which has been upgraded and modified from [31]. According to the figure, we first employed the image preprocessing approach,

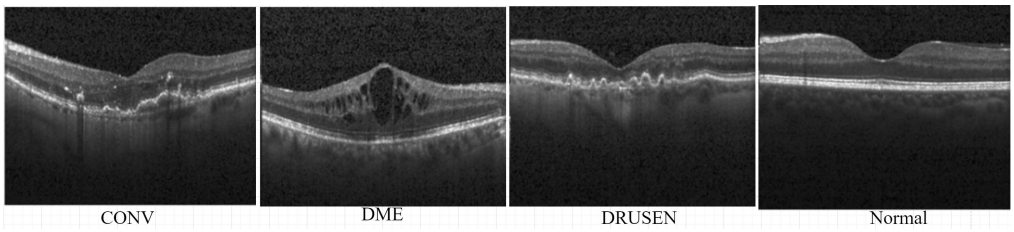

FIGURE 2. Sample Images of the optical coherence tomography (OCT) 2017 dataset.

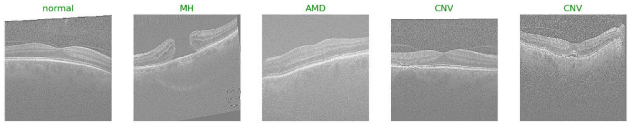

FIGURE 3. Sample Images of the Retinal C8 dataset [29].

TABLE 1. Overview of datasets used.

Dataset	Classes	Total Samples	Type
OCT2017 [5]	Multiclass classification: 4 Binary Classification: 2	84,495	Images
Dataset101 [4]	6	9825	Images
Retinal OCT-C8 [29]	8	24000	Images

including the augmentation technique. Then, we applied the multi-stage deep learning module, including three stages, which were first considered as the initial module to extract the grain feature of the dataset. We enhance the grain feature using a two-branch-based deep learning module where each branch is composed of a convolutional block and an identity block in the second stage. The third stage is composed of the LSTM module and classification module to generate the probabilistic maps. Figure 5 demonstrated the proposed multi-stage deep learning module. Figure 6 showed the convolutional module, and Figure 7 demonstrated the identity module [31].

- **Data Collection:** The initial stage of our methodology involves obtaining sufficient and accurate data, which is crucial for training and evaluating our deep learning model. We gather diverse datasets containing images of various eye diseases, ensuring comprehensive coverage and representation. This step aligns with the critical need for reliable data to facilitate robust model training.
- **Data Preprocessing and Augmentation:** Following data collection, we proceed to the preprocessing stage, where we implement robust preprocessing and augmentation techniques. This step ensures the quality and uniformity of the data while enhancing its adaptability across different scenarios. By incorporating various image augmentation methods, we create a rotation and translation-independent system, optimizing our model's performance and generalizability.
- **Training Deep Learning Models:** The core of our methodology involves training deep learning models tailored specifically for eye disease detection. We propose a novel three-stage deep learning approach:
  - **Stage 1:** Utilizes convolutional layers, Batch Normalization, and activation layers to extract fine-grained features.
  - **Stage 2:** Enhances these features using two branches. Branch-1 consists of two convolutional blocks and four identity blocks, while Branch-2 comprises one convolutional block and two identity blocks. Both branches include global average pooling layers. The concatenated features from both branches provide rich and robust hierarchical features.
  - **Stage 3:** The final classification module generates probabilistic maps for disease detection.

  This architecture enables efficient feature extraction and classification, addressing challenges such as inadequate feature effectiveness and high computational demands. The models are trained on the preprocessed data, leveraging the augmented dataset to enhance learning accuracy and robustness. Evaluation serves as a critical step in assessing the performance of the trained models. We rigorously evaluated the models on benchmark datasets, including OCT2017, Dataset-101, and Retinal OCT C8 for multi-class and binary classification tasks, respectively. The flowchart depicted in Figure 4 outlines the proposed methodologies, where each block is distinctly labelled, representing specific processing steps.

A. DATA PREPROCESSING

Pre-processing is the term used to describe the changes made to our data before we feed it into the algorithm. The process of transforming raw data into a clean dataset is known as data preprocessing. Following data collection, it was observed that the images exhibit varying resolutions and inconsistencies due to their origin from different datasets. So, data preprocessing techniques, including image resizing, were employed to standardize the images to valid resolutions, accommodating the specific input shapes required by different models. Additionally, data augmentation techniques, including horizontal flip, random flip, height shift, width shift, random zoom, random rotation, and random translation, were utilized to enhance the diversity of the model's training dataset. The suggested transformation and its potential range

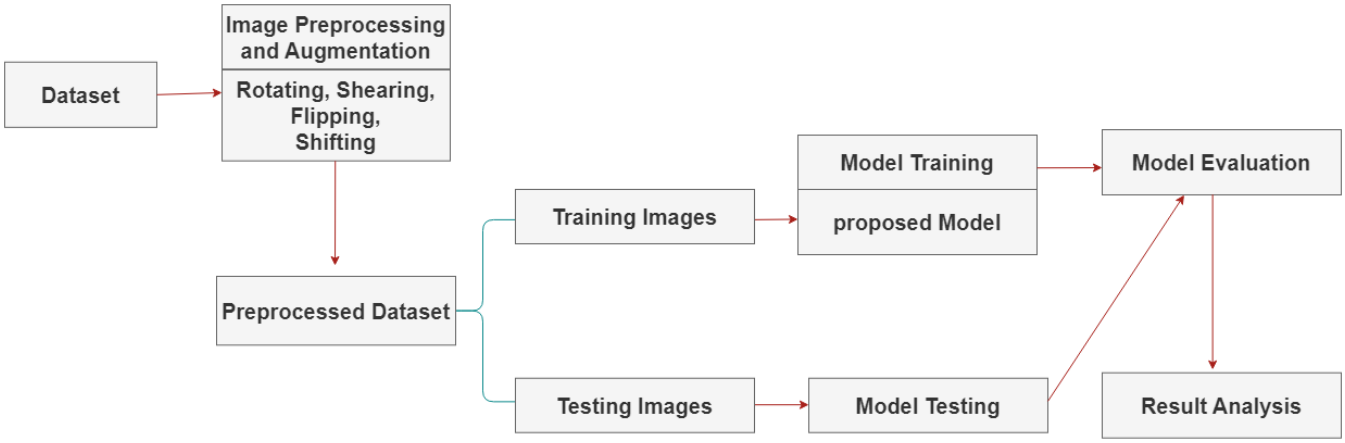

FIGURE 4. Proposed methodology.

for image augmentation are displayed in Table 1. The dataset's experimental test and image analysis are used to determine the range of each augmentation mentioned in the table.

TABLE 2. Augmentation techniques and possible ranges.

Augmentation Technique	Range
Rotation range	20
Width shift range	0.2
Height shift range	0.2
Shear range	0.2
Zoom range	0.2
Horizontal flip	True
Fill mode	nearest

B. STAGE-1: GRAIN FEATURE STAGE

The input to the network is an image with a size of 224×224 pixels and 3 color channels (RGB). The first layer is a convolutional layer with a filter size of 3 × 3 and a stride of 2. This layer will downsample the input image by a factor of 2. The Batch Normalization layer normalizes the output of the convolutional layer to have a mean of 0 and a standard deviation of 1. The activation layer applies an activation function ReLU to the output of the Btch Normalization layer. Subsequently, the output is directed towards branches 1 and 2.

C. STAGE-2: FEATURE ENHANCEMENT

The feature from Stage 1 is fed into the Stage 2 module to enhance the spatial feature extracted from Stage 1. In Stage 2, we extracted features using two branches to enhance the feature's range dependency and short-range dependency.

1) BRANCH-1 FEATURES

The Branch-1 of this model initially begins with a CONV-BLOCK, which we name CONV-BLOCK-1. This block is followed by two identity blocks, IDENTITY-BLOCK-1 and IDENTITY-BLOCK-2. We have another CONV-BLOCK, followed by two identity blocks, IDENITTY-BLOCK-3 and IDENITTY-BLOCK-4. Finally, the output passes through the Maxplooling layer and the Global Average Pooling layer and reaches the concatenation layer.

Conv Block: The input to the CONV-BLOCK is represented as x or x_skip, which could be the output of a previous block or a skip connection from an earlier layer in the network. The 3 × 3 CONV is the first layer in the CONV-BLOCK, which performs a convolution operation with a 3 × 3 kernel size. The stride is 2, which means the output feature map will have half the spatial dimensions of the input. Padding is set to ''same'', ensuring the spatial dimensions of the output are the same as the input. The 1 × 1 CONV layer reduces the number of channels in the feature map, preparing it for the next convolution layer. It also has a stride of 2, further decreasing the spatial dimensions of the output. The Batch Normalization layer normalizes the activations of the previous layer, improving the network's training stability and performance. The activation layer applies the ReLU activation function to the output of the Batch Normalization layer. 3 × 3 CONV is another second convolution layer in the CONV-BLOCK, which performs another convolution operation with a 3 × 3 kernel size. The stride is 2, and the padding is ''same'' similar to the first convolution layer. Add layer adds the output of the first 1 × 1 CONV-BLOCK (x_skip) to the output of the second 3 × 3 CONV-BLOCK (x). This is known as a skip connection or shortcut connection, allowing the network to learn an identity function and mitigate the vanishing gradient problem. The Batch Normalization layer normalizes the activations of the skip connection. The activation layer applies the ReLU activation function to the output of the skip connection.

Identity Block: This is the residual block, which takes two inputs: x and f. x is the input to the block, and f is the number of filters or feature maps. The block consists of two convolutional layers, followed by Batch Normalization

and an activation function. The CONV 3 × 3 is the first convolutional layer with a kernel size of 3 × 3 and uses "same" padding, meaning the output spatial dimensions are the same as the input. This layer is followed by Batch Normalization and an activation function, ReLU. The second convolutional layer also has a kernel size of 3 × 3 and uses "same" padding. This layer is followed by Batch Normalization. The input x is added to the output of the second convolutional layer, known as a residual or shortcut connection. This allows the network to learn the residual function

$$F(x) = H(x) - x$$

, where $H(x)$ is the output of the convolutional layers. The block's final output is the result of the addition, followed by an activation function ReLU.

2) BRANCH-2 FEATURES

This model's Branch-2 starts with a CONV-BLOCK. We refer to it as CONV-BLOCK-1. Two identity blocks come after this one. IDENTITY-BLOCK-1 and IDENTITY-BLOCK-2 are these blocks. The output ultimately travels through the Maxplooling, GlobalAveragePooling and finally concatenation layer.

The second CONV-BLOCK accepts an input of x or x_skip, which is a skip connection from an earlier layer of the network or the output of a previous block. The first layer in the CONV-BLOCK is called the 3 × 3 CONV, and it uses a 3 × 3 kernel size to conduct a convolution operation. The output feature map will have half the spatial dimensions of the input because the stride is 2. The "same" padding setting guarantees that the output's spatial dimensions match those of the input. In order to prepare the feature map for the subsequent convolution layer, the 1 × 1 CONV layer reduces the number of channels in the map. Its stride of two reduces the output's spatial dimensions even more. By normalizing the activations of the preceding layer, the Batch Normalization layer enhances the training stability and performance of the network. The ReLU activation function is applied by the activation layer to the Batch Normalization layer's output. A second convolution layer in the CONV-BLOCK, 3 × 3 CONV, carries out an additional convolution operation using a 3 × 3 kernel size. Padding is "same", and stride is 2, just like in the first convolution layer. The output of the second 3 × 3 CONV-BLOCK (x) is added to the output of the first 1 × 1 CONV-BLOCK (x_skip) using the add layer function. This kind of connection, also referred to as a shortcut or skip connection, helps the network learn an identity function and lessens the effects of the vanishing gradient issue. The skip connection activations are normalized by the Batch Normalization layer. ReLU activation function is applied by the activation layer to the skip connection's output.

The second residual block accepts the following two inputs: x and f. The block's input, denoted by x, and the number of filters, or feature maps, represented by f. The block has two convolutional layers, an activation function, and Batch Normalization after each. The first convolutional layer, called CONV 3 × 3, uses "same" padding, meaning that the input and output spatial dimensions are the same. Its kernel size is 3 × 3. Batch normalization and an activation function called ReLU come after this layer. The second convolutional layer uses "same" padding and has a kernel size of 3 × 3. The Batch Normalization layer comes after this one. The second convolutional layer's output also referred to as a residual connection or shortcut connection, is multiplied by the input x. The network is able to learn the residual function

$$F(x) = H(x) - x$$

as a result, where $H(x)$ is the convolutional layer's output. The addition result is the block's final output, and then an activation function called ReLU comes next.

D. STAGE-3: FEATURE FUSION AND CLASSIFICATION MODULE

The final features generated by concatenating the branch-1 and branch-2 of the 2nd stage feature according to the following Equation (1):

$$F_{Final} = F_{Branch-1} \bigoplus F_{Branch-2} \tag{1}$$

According to the concatenation, the branch-1 and branch-2 outputs are added to the concatenation layer, which produces F_{Final} that is fed into the classification module, which consists of the LSTM. In the classification module, LSTM produces the output that feeds into the Dense layer, which is a fully connected layer, also known as a dense layer. It takes the output of the LSTM layer as input and applies a weight matrix to it, followed by a bias term. The number of neurons in this layer is equal to the number of classes in the classification problem. The output of this layer is a vector of probabilities for each class. The classification layer represents the final output of the network, which is the classification result. It uses the Dense layer output to predict the input image's class label. The class label with the highest probability is chosen as the final prediction.

V. EXPERIMENTAL EVALUATION

The process of teaching a neural network to make precise predictions by providing it with enormous amounts of labeled data is known as deep learning model training. The network's internal parameters, like weights and biases, are adjusted based on the data until an accurate mapping of inputs to outputs is achieved. This process, which is frequently iterative, requires close attention to elements like training parameters, model architecture, and data quality. With the following three parameters, the models will be trained using the following dataset: the number of epochs, training dataset, and validation dataset. The first step of model training is to split the dataset into three partitions: the training dataset, the test dataset, and the validation dataset. In our case, approximately 80% of data are set as training dataset, 10% as testing, and 10% as validation. The number of times the model

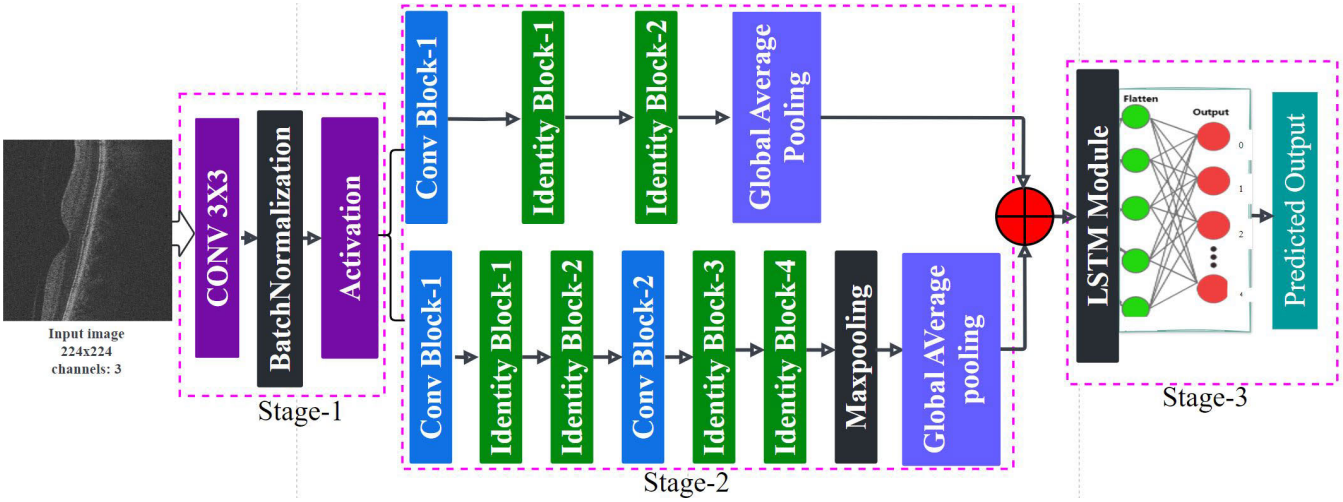

FIGURE 5. Architecture of our proposed CNN model.

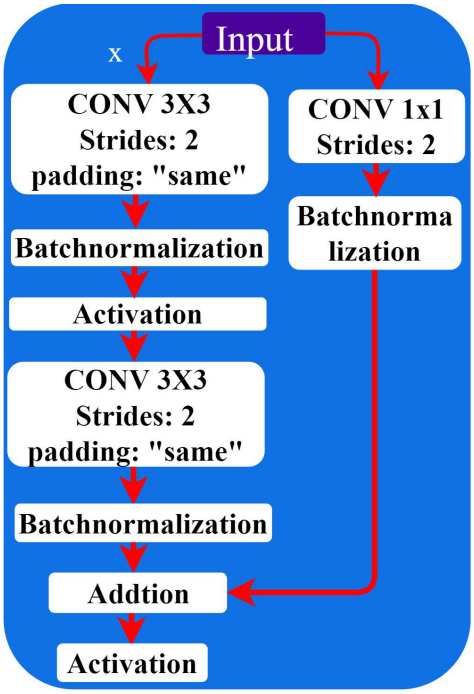

FIGURE 6. CONV-BLOCK architecture of the model [31].

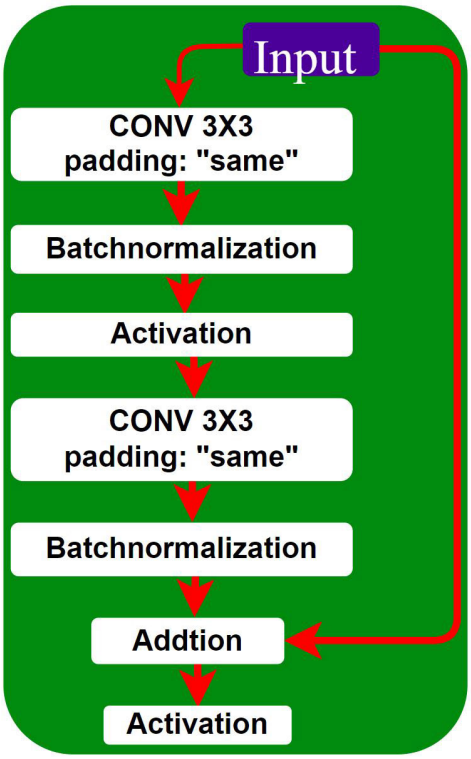

FIGURE 7. IDENTITY-BLOCK architecture of the model [31].

iterates through the data is called the number of epochs. A fixed number of epochs were employed to train all our models. Different activation functions were applied across various types of layers. Additionally, a specific learning rate was utilized to regulate the step size in the optimization process. A constant batch size was employed in each iteration during training. Furthermore, optimizers were used in the training process. an algorithm or method used to update the parameters of a neural network during the training process in order to minimize the error or loss function. Additionally, for monitoring the accuracy of your model during training and evaluation, a specific metric was utilized for evaluation purposes. Regularization techniques were implemented to mitigate overfitting. Overfitting, characterized by a model learning the training data excessively, capturing noise or irrelevant patterns, often results in poor performance on unseen or test data. Different filter sizes were applied across distinct convolutional layers. Additionally, a fixed stride size was utilized consistently across all models.

A. ENVIRONMENTAL SETTING

A stable, customized environment is necessary for deep learning model training. These are the basic arrangements we kept to throughout our research. On the hardware side, it used

12.7 GB of RAM, about 107.7 GB of primary storage, and an NVIDIA Tesla T4 GPU. In software side, Python 3 was used. Parameters that are set before the model is trained are called hyperparameters in a deep learning (DL) model. They are independent of the model's weights and aren't gained through training. The model's learning process and performance are greatly impacted by these hyperparameters. On the parameter side, we used 100 epochs for our model to train. kept image sizes 224×224, used loss function as 'sparse_categorical_cross entropy', 'Adam', 'RMSProp' as optimizer, learning rate of 0.001, 'Dropout', 'Early stopping' as Regularization, batch size 32, 'softmax', 'ReLU' as activation functions, padding size 3 × 3 and stride size 2 throughout the training. We used accuracy, precision, recall, and F1-score to show the performance matrix here.

B. PERFORMANCE ACCURACY WITH EYE DISEASE OCT2017 DATASET

Although the Oct2017 dataset contains four classes, some researchers used only two of these datasets to evaluate their model as a binary classification. In the study, we also showed performance accuracy with multi-class and binary classes.

1) MULTI CLASS CLASSIFICATION

The performance metrics of our proposed model for multi-class classification on the OCT2017 dataset are shown in Table 3. The model demonstrates high precision, recall, F1-score, and ROC-AUC scores across all classes. Specifically, for the CNV class, the model achieves a precision of 91.6%, a recall of 99.5%, an F1-score of 95%, and an ROC-AUC score of 99.9%. The DME class attains a perfect precision of 100%, a recall of 96.2%, an F1 score of 98.1%, and a ROC-AUC score of 99.9%. The DRUSEN class shows a precision of 99.1%, a recall of 95%, an F1 score of 98%, and an ROC-AUC score of 99%. The NORMAL class achieves perfect scores in precision and recall, with an F1 score of 99.5% and an ROC-AUC score of 99.9%. In multi-class classification, our model achieves 97.52% test accuracy on the Oct2017 dataset. In addition, Figure 8 and 9 demonstrated the accuracy and loss curve, and Figure 10 and Figure 11 demonstrated the confusion matrix and ROC curve for the OCT2017 multi-class dataset.

According to the confusion matrix, there were nine DME misclassified as CNV, twelve DRUSEN misclassified as CNV, one normal eye misclassified as CNV and one DME affected by CNV disease. The Correct classifications for CNV are 241, DME is 233, DRUSEN is 230, and normal eye is 240.

The state-of-the-art comparison for the proposed model with the OCT2017 dataset is demonstrated in Table 4. According to the comparison table, the ResNet50v2 model achieved the highest accuracy of 98.35%. With DenseNet-121 gaining 97.00%, our proposed model performs 97.52%. In comparison with the DenseNet-121 model, the difference is 0.52% larger. Interestingly, despite having fewer parameters, our proposed model performs almost as well as the State of the art model. Furthermore, our model completely outperforms DenseNet-121 in terms of F1 score, precision, and recall.

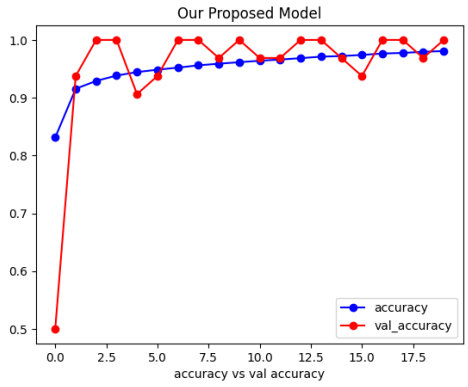

FIGURE 8. Learning curve of our proposed model on multi classification for Oct2017 dataset.

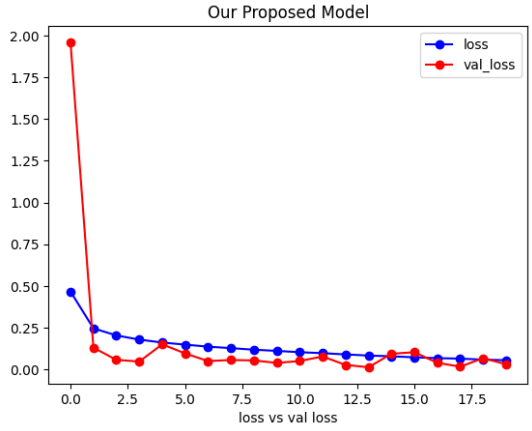

FIGURE 9. Loss curve of our proposed model on multi classification for Oct2017 dataset.

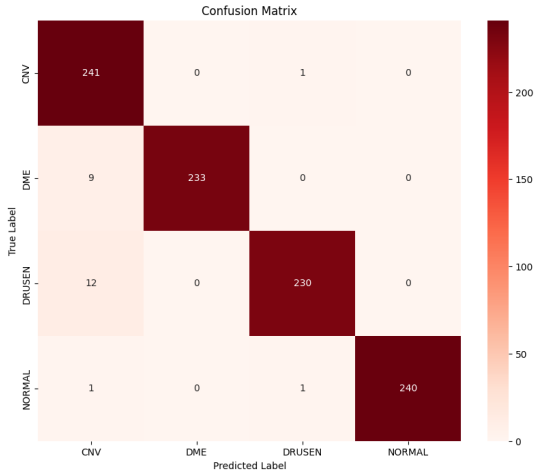

FIGURE 10. Confusion Matrix results of our proposed model on Oct2017 Dataset.

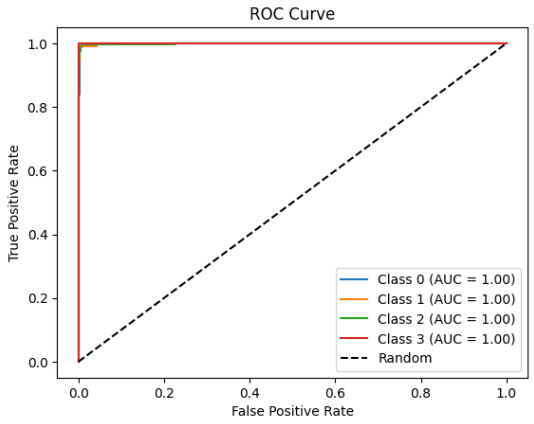

FIGURE 11. ROC curve results of our proposed model on Oct2017 Dataset.

TABLE 3. Performance metrics of our proposed model on multi class classification for Oct2017 Dataset.

Class	Precision	Recall	F1-score	ROC-AUC Score
CNV	91.6%	99.5%	95%	99.9%
DME	100%	96.2%	98.1%	99.9%
DRUSEN	99.1%	95%	98%	99%
NORMAL	100%	99.1%	99.5%	99.9%

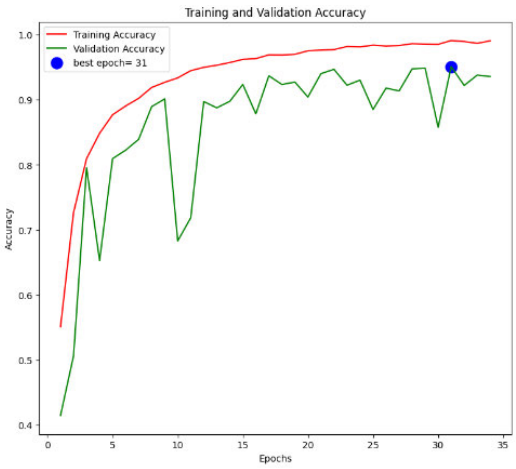

FIGURE 12. Learning curve of our proposed model on multiclass classification for Retinal C8 dataset.

C. PERFORMANCE ACCURACY OF RETINAL C8 DATASET

Our proposed model generated 94.81% of test accuracy on the Retinal C8 dataset after a sufficient number of epochs of training.

The confusion matrix in figure 14 reveals that 772 AMD cases were appropriately identified as AMD. 668 cases of CNV were appropriately identified as such. It was incorrect to classify 28 cases of CNV as DME. 43 CNV cases were mistakenly assigned to the DRUSEN category. 7 cases of CNV were mistakenly categorized as normal. 732 CSR occurrences were appropriately identified as such. 4 CSR cases were mistakenly assigned to DME. 2 CSR cases

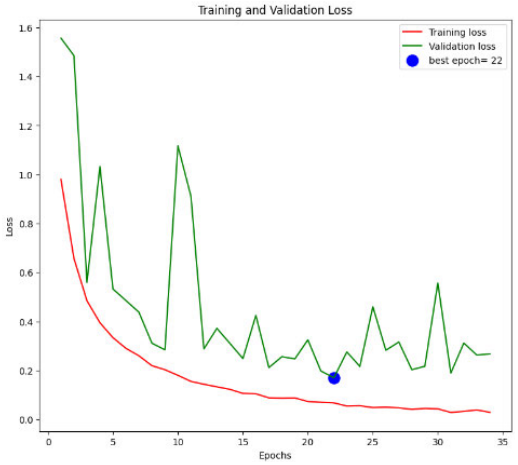

FIGURE 13. Loss curve of our proposed model on multiclass classification for Retinal C8 dataset.

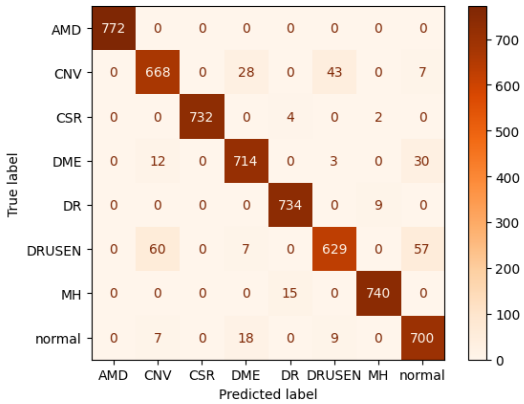

FIGURE 14. Confusion matrix results of our proposed model on Retinal C8 Dataset.

were mistakenly assigned to MH. 714 DME cases were appropriately identified as DME. 12 cases of DME were mistakenly assigned to CNV. 3 DME cases were mistakenly identified as DRUSEN. 30 DME cases were incorrectly identified as normal. 734 DR cases were appropriately categorized as such. 9 DR cases were incorrectly identified as normal. 629 DRUSEN cases were appropriately categorized as such. 50 DRUSEN cases were mistakenly identified as CNV. 7 DRUSEN cases were mistakenly identified as DME. 57 DRUSEN cases were incorrectly identified as normal. 740 MH cases were appropriately identified as MH. 15 cases of MH were incorrectly assigned to DME. 700 examples of normal were appropriately categorized as such. 7 cases of normal were mistakenly assigned to CNV. 18 cases of normal were mistakenly assigned to DME. 9 normal cases were mistakenly identified as DRUSEN.

D. PERFORMANCE ACCURACY WITH EYE DISEASE 101 DATASET

The performance metrics of our model for multi-class classification on the Eye Disease 101 dataset demonstrate in Table 7 which proves the effectiveness. For the acrima class,

TABLE 4. State of the Art models with multi-class Oct2017 dataset.

Model	Accuracy	F1-score	Precision	Recall	params
DenseNet-121	97.00%	97.01%	97.07%	97.00%	7.3M
ResNet50v2	98.35%	98.34%	98.37%	98.34%	23.5M
Ours	**97.52%**	**97.54%**	**97.69%**	**97.52%**	**0.372M**

TABLE 5. A comparison of our proposed model and previous researhces with the OCT2017 dataset.

Paper ID	Model	Accuracy
kermany et. al [7]	VGG16 Transfer Learning	96.6%
Alqudah [8]	AOCT-NET	97.1%
Melnychuk et. al [28]	WideResNet-50-2	99.69%
Chetoui et al. [32]	SVM and attention mechanisms	98.46%
Tsuji et al. [33]	Capsule Network Architecture	99.6%
Zhenwei Li et al. [30]	Swin Transformer	99.99%
Ours	**3 stages CNN model**	**97.52%**

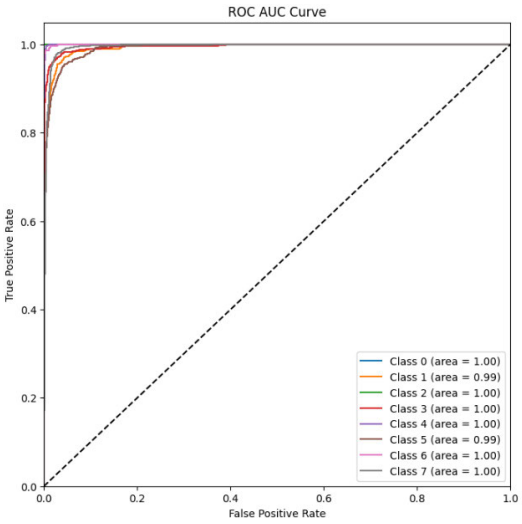

FIGURE 15. ROC curve results of our proposed model on Retinal C8 Dataset.

TABLE 6. Performance Metrices of our proposed model on Retinal c8 dataset.

	Precision (%)	Recall (%)	F1-score (%)
AMD	100	100	100
CNV	89	90	89
CSR	100	99	100
DME	93	94	94
DR	97	99	98
DRUSEN	92	84	88
MH	99	98	98
Normal	88	95	92

the model achieved perfect precision of 100.00%, recall of 99.3%, F1-score of 99.6%, and ROC-AUC score of 99.9%. The cataract class showed precision of 65.2%, recall of 50,

TABLE 7. Performance metrics of our model in multi-class classification on Eye Disease 101 dataset.

Class	Precision	Recall	F1-score	ROC-AUC Score
acrima	100.00%	99.3%	99.6%	99.9%
cataract	65.2%	50%	56.6%	97.22%
glaucoma	26.8%	31.5%	35.0%	91.8%
odir5k	97.92%	99%	98.45%	99.46%
origa	82.2%	100%	90.16%	98.61%
retina disease	27.27%	10%	14.63%	96.95%

TABLE 8. Comparison of State of the Art models with proposed model in multi class classification for Dataset-101.

Model	Accuracy	F1-score	Precision	Recall	params
DenseNet-121	94.84%	74.09%	76.98%	72.28%	7.3M
ResNet50v2	92.40%	65.51%	67.91%	63.78%	23.5M
Ours	**92.97%**	**68.7%**	**73.13%**	**66.10%**	**0.372M**

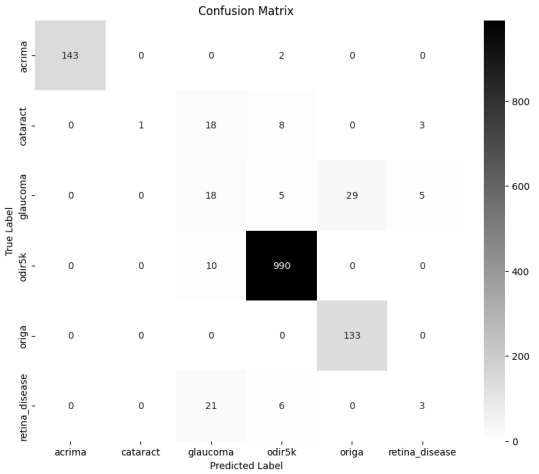

FIGURE 16. Confusion Matrix results of our proposed model on Dataset-101.

F1-score of 56.6%, and ROC-AUC score of 97.22%. For Glaucoma, the precision was 26.8%, recall 31.5%, F1-score 35.0, and ROC-AUC score 91.8. The odir5k class had high precision of 97.22%, recall of 99.46%, F1-score of 98.1%, and ROC-AUC score of 99.5%. The origa class demonstrated precision of 82.2%, recall of 100%, F1-score of 90.16%, and ROC-AUC score of 98.61%. For retina disease, the precision was 27.27%, recall 10%, F1-score 14.63%, and ROC-AUC score 96.5%. Table 8 demonstrated the state-of-the-art comparison of the proposed model with this dataset.

1) BINARY CLASSIFICATION

E. PERFORMANCE ACCURACY OF OCT-2017 DATASET

The proposed model's performance metrics for binary classification on the OCT2017 dataset are excellent, which is demonstrated in Table 10. For AMD, it achieves a precision 100%, a recall of 99.8%, an F1-score of 99%, and a ROC-AUC score 100%. For NORMAL, the precision is 98%, recall is 100%, F1-score is 99%, and ROC-AUC score is 100%. In the experiment, the Combination of classes CNV

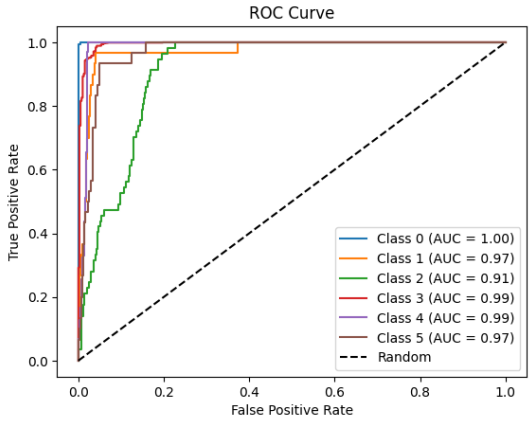

FIGURE 17. ROC curve results of our proposed model on Dataset-101.

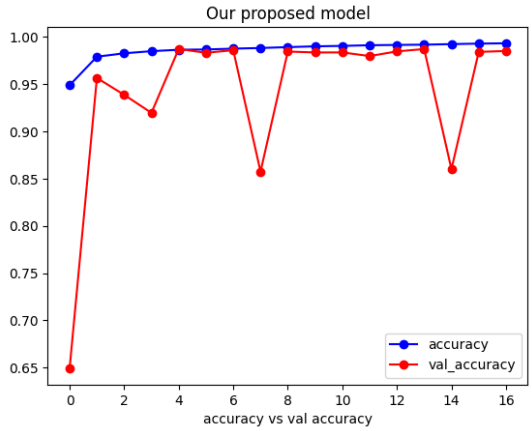

FIGURE 18. Learning curve of our proposed model on binary classification for Oct2017 dataset.

and DRUSEN refers to a new class called AMD (age-related macular degeneration). We trained our model on AMD vs. NORMAL classification and got a test accuracy of 99.33%. In addition, Figures 18 and 19 demonstrated the loss and accuracy curves of this dataset protocol. It is evident from the above table that our model achieves the highest ROC-AUC score 100% for both classes. Class AMD has a maximum precision 100%, while class NORMAL has a 25% lower precision but is still valid. Class NORMAL achieved the highest recall value of 100%, and both classes performed equally well, obtaining an F1 score of 99%. It is evident from the above table that our model achieves the highest ROC-AUC score 100% for both classes. Class AMD has a maximum precision of 100%, while class NORMAL has a 25% lower precision but is still valid. Class NORMAL achieved the highest recall value of 100%, and both classes performed equally well, obtaining an F1 score of 99%.

We acknowledge that the model's performance can be influenced by the characteristics of the datasets used for training. However, our approach demonstrates strong adaptability and robustness, as evidenced by its consistent performance across multiple datasets. To enhance generalizability, we employed data augmentation techniques and validated

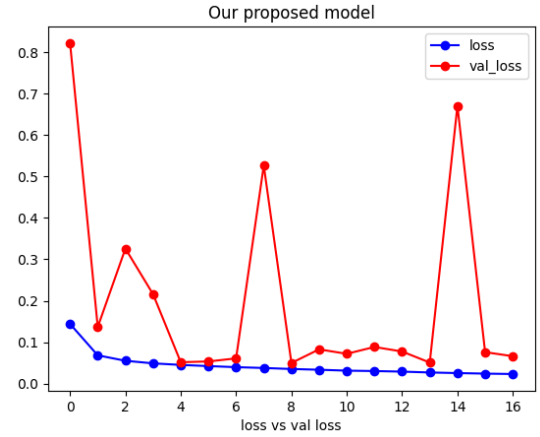

FIGURE 19. Loss curve of our proposed model on binary classification for Oct2017 dataset.

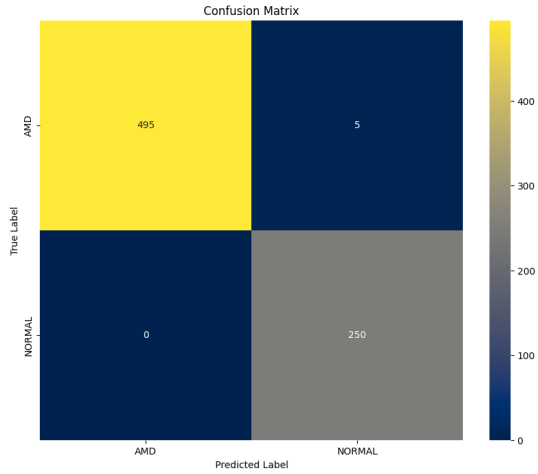

FIGURE 20. Confusion Matrix results of our proposed model on Oct2017 Dataset.

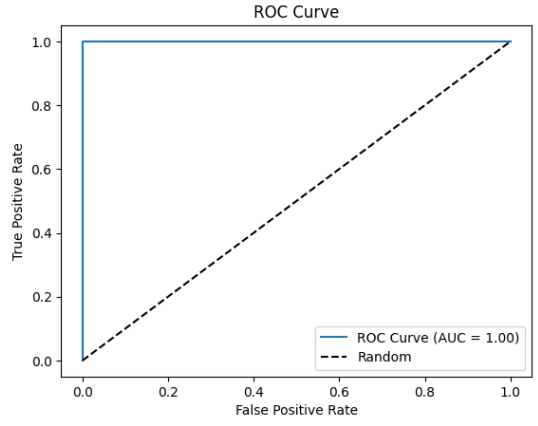

FIGURE 21. ROC curve results of our proposed model on Oct2017 Dataset.

the model on three diverse datasets, effectively minimizing the risk of overfitting. The model's use of dropout and early stopping further ensures it maintains high performance by preventing it from learning noise and overfitting to any particular dataset. In this study, we evaluated the proposed model on three different datasets: OCT2017, DATASET-101,

TABLE 9. A comparison between the suggested model and previous research on binary classification using the Oct2017 dataset.

Author Name	Model	Accuracy
Tingting et. al [23]	CNN & LOF algorithm	99.87%
Thomas et. al [26]	Multi scale Multi path CNN	99.78%
Thomas et. al [35]	Multi scale CNN	99.73%
kermany et al. [7]	InceptionV3 transfer learning	96.53%
Das et al. [9]	Multi-scale deep feature fusion	97.71%
Kaymak [34]	AlexNet transfer learning	98.26%
Fang et al. [6]	Iterative fusion CNN	93.40%
Huang et al [35]	DenseNet-121	99.20%
Ours	**3 stages CNN model**	**99.33%**

TABLE 10. Performance metrics of our proposed model for binary classification on Oct2017 Dataset.

Class	Precision	Recall	F1-score	ROC-AUC Score
AMD	100%	99.8%	99%	100%
NORMAL	98%	100%	99%	100%

and Retinal OCT-C8, achieving high test accuracies of 97.52%, 99.33%, and 94.81%, respectively. The consistent high performance across datasets with distinct characteristics demonstrates that our model effectively captures relevant features of eye diseases, regardless of variations in image quality or diversity. While slight variations in accuracy (such as the lower performance on the Retinal OCT-C8 dataset) suggest some sensitivity to dataset characteristics, they also highlight the model's robustness in maintaining high accuracy across diverse data sources. By testing the model on multiple datasets and implementing these advanced strategies, we have taken proactive steps to ensure that the model is highly generalizable and robust. This makes it a reliable tool for practical deployment in varied clinical settings. We are committed to further refining the model and exploring additional methods to enhance its applicability across even more diverse environments in future studies.

VI. DISCUSSION WITH POTENTIAL APPLICATION AND IMPACT ON OPHTHALMOLOGY

Our proposed lightweight deep-learning model for eye disease detection has significant potential to transform the field of ophthalmology by enabling real-time diagnosis and enhancing the efficiency and effectiveness of patient care. Below, we outline the key aspects of its application, integration into clinical practice, and potential impact:

A. REAL-TIME EYE DISEASE DETECTION

The system is designed to provide immediate assessments of eye conditions, such as Diabetic Retinopathy, Glaucoma, and Age-Related Macular Degeneration, using an optimized deep learning model for rapid image processing and classification. Its lightweight architecture allows deployment on edge devices, such as mobile phones or point-of-care diagnostic tools, enabling healthcare providers to perform instant, on-site evaluations. By combining imaging techniques with real-time analysis, the system delivers accurate diagnostic information quickly, supporting timely and informed clinical decisions. The detection system's user interface is designed to be intuitive and user-friendly, clearly displaying diagnostic results and risk assessments for easy interpretation. Key features include automated disease detection, real-time alerts for abnormal findings, comprehensive patient reports, and decision support for follow-up actions. The system integrates seamlessly with imaging methods, such as fundus photography and optical coherence tomography (OCT), and connects directly with electronic health record (EHR) systems. This integration streamlines the acquisition and organization of high-quality eye images, enhances data analysis, and improves diagnostic accuracy. It assists ophthalmologists in early detection of conditions, tailoring treatment plans to individual patients, and monitoring disease evolution over time, leading to improved patient outcomes and more effective clinical decision-making.

B. INTEGRATION INTO CLINICAL PRACTICE

For effective integration into clinical practice, the system is designed to operate within existing healthcare workflows with minimal disruption. Ensuring compatibility with EHR systems and imaging devices allows seamless data exchange and integration. Developing standardized protocols for data sharing, privacy, and security ensures compliance with regulatory standards. Comprehensive training for healthcare professionals facilitates smooth adoption and effective use of the system. Ongoing collaboration with ophthalmologists and other healthcare professionals through pilot studies, real-world validation, and iterative updates ensures the system is continuously refined to meet clinical needs.

C. OVERCOMING BARRIERS AND ENSURING EFFECTIVE DEPLOYMENT

Our system addresses potential adoption barriers, such as infrastructure compatibility, data privacy, and regulatory approvals, through compliance, clinical trials, and provider training. Collaboration with clinicians integrates clinical insights, refines model accuracy, and validates effectiveness in real-world scenarios. The system is designed for adaptability and scalability, overcoming challenges related to diverse healthcare environments, patient populations, and image quality, ensuring consistent performance across various settings.

VII. CONCLUSION

Our research represents a significant stride forward in addressing the pressing challenge of eye disease classification, a critical aspect of global healthcare requiring prompt detection to prevent vision loss. Recognizing the imperative for automated detection systems, we embarked

on developing an innovative approach to tackle existing weaknesses, including ineffective feature representation, high computational overheads, and incomplete disease coverage. Our pioneering eye-disease detection system harnesses the formidable capabilities of deep learning technologies. Through meticulous processing and the incorporation of diverse image augmentation techniques to ensure robustness against rotations and translations, we established a solid framework for our model's advancement. Our novel lightweight three-stage deep learning architecture, characterized by a judicious fusion of convolutional and identity blocks, facilitates efficient feature extraction and classification. Central to our methodology is the construction of a deep learning model comprising Stage 1, which extracts fine-grained features with convolutional layers, Batch Normalization, and activation layers; Stage 2, which enhances these features through two branches (Branch-1 with two convolutional blocks and four identity blocks and Branch-2 with one convolutional block and two identity blocks), both incorporating global average pooling layers; and Stage 3, which utilizes a classification module to generate precise probabilistic maps. This meticulously crafted architecture, complemented by advanced preprocessing and augmentation techniques, ensures the generation of accurate and reliable predictions. Upon rigorous evaluation against benchmark datasets OCT2017, Dataset-101, and Retinal OCT C8 for multi-class and binary classification, respectively, our proposed model demonstrated superior performance in both accuracy and efficiency. Significantly, our approach achieves heightened diagnostic efficacy while substantially reducing computational overheads, underscoring its potential to revolutionize ophthalmology. Looking ahead, our future work will focus on further refining and optimizing our methodology to push the boundaries of eye disease detection accuracy. We envision collaborating closely with healthcare professionals and ophthalmologists to incorporate clinical insights, ensuring the practical relevance and effectiveness of our system in real-world settings. Additionally, ongoing evaluation and validation using larger and more diverse datasets, alongside real-world clinical data, will be pivotal in enhancing the robustness and applicability of our approach.

REFERENCES

[1] M. A. Hossain, T. A. Asa, F. Huq, and M. A. Moni, "Eye disorders in bangladesh: A hospital-based descriptive study," *J. Biomed. Anal.*, vol. 2, no. 1, pp. 27–40, Mar. 2019.

[2] I. Sutradhar, P. Gayen, M. Hasan, R. D. Gupta, T. Roy, and M. Sarker, "Eye diseases: The neglected health condition among urban slum population of Dhaka, Bangladesh," *BMC Ophthalmology*, vol. 19, no. 1, pp. 1–22, Dec. 2019.

[3] M. Yusufu, J. Bukhari, X. Yu, T. P. H. Lin, D. S. C. Lam, and N. Wang, "Challenges in eye care in the Asia–Pacific region," *Asia–Pacific J. Ophthalmology*, vol. 10, no. 5, pp. 423–429, Sep. 2021.

[4] *Dataset for Different Eye Disease*. Accessed: Feb. 1, 2024. [Online]. Available: https://www.kaggle.com/datasets/dhirajmwagh1111/dataset-for-different-eye-disease

[5] D. Kermany, K. Zhang, and M. Goldbaum. (2018). *Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-ray Images*. [Online]. Available: https://data.mendeley.com/datasets/rscbjbr9sj/3

[6] L. Fang, Y. Jin, L. Huang, S. Guo, G. Zhao, and X. Chen, "Iterative fusion convolutional neural networks for classification of optical coherence tomography images," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 327–333, Feb. 2019.

[7] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, Feb. 2018.

[8] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2017, *arXiv:1605.07146*.

[9] V. Das, S. Dandapat, and P. K. Bora, "Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images," *Biomed. Signal Process. Control*, vol. 54, Sep. 2019, Art. no. 101605.

[10] Healthcare Radius. (2040). *Glaucoma Eye Disorder is Expected To Double in India By 2040*. Accessed: Oct. 2, 2024. [Online]. Available: https://www.healthcareradius.in/clinical/glaucoma-eye-disorder-is-expected-to-double-in-india-by-2040

[11] A. S. M. Miah, Md. R. Islam, and M. K. I. Molla, "Motor imagery classification using subband tangent space mapping," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2017, pp. 1–5.

[12] A. S. M. Miah, S. R. A. Ahmed, M. R. Ahmed, O. Bayat, A. D. Duru, and Md. K. I. Molla, "Motor-imagery BCI task classification using Riemannian geometry and averaging with mean absolute deviation," in *Proc. Sci. Meeting Elect.-Electron. Biomed. Eng. Comput. Sci. (EBBT)*, Apr. 2019, pp. 1–7.

[13] A. S. M. Miah, M. R. Islam, and M. K. I. Molla, "EEG classification for MI-BCI using CSP with averaging covariance matrices: An experimental study," in *Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng.*, Jul. 2019, pp. 1–5.

[14] M. R. Rahman, M. T. Hossain, N. Nawal, M. S. Sujon, A. S. M. Miah, and M. M. Rashid, "A comparative review of detecting Alzheimer's disease using various methodologies," *BAUST J.*, vol. 2, pp. 1–26, Jun. 2020.

[15] K. Atai Kibria, A. Sarker Noman, M. Abir Hossain, M. Shohidul Islam Bulbul, M. Mamunur Rashid, and A. Saleh Musa Miah, "Creation of a cost-efficient and effective personal assistant robot using Arduino & machine learning algorithm," in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 477–482.

[16] T. Zobaed, S. R. A. Ahmed, A. S. M. Miah, S. M. Binta, M. R. A. Ahmed, and M. Rashid, "Real time sleep onset detection from single channel EEG signal using block sample entropy," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 928, May 2020, Art. no. 032021.

[17] M. S. Ali, J. Mahmud, S. M. F. Shahriar, S. Rahmatullah, and A. S. M. Miah, "Potential disease detection using naive Bayes and random forest approach," *BAUST J.*, vol. 1, no. 1, pp. 1–23, 2022.

[18] M. A. Rahim, F. A. Farid, A. S. M. Miah, A. K. Puza, M. N. Alam, M. N. Hossain, S. Mansor, and H. A. Karim, "An enhanced hybrid model based on CNN and BiLSTM for identifying individuals via handwriting analysis," *Comput. Model. Eng. Sci.*, vol. 140, no. 2, pp. 1689–1710, 2024.

[19] M. M. Hossain, A. S. Noman, M. M. Begum, W. A. Warka, M. M. Hossain, and A. S. Musa Miah, "Exploring Bangladesh's soil moisture dynamics via multispectral remote sensing satellite image," *Eur. J. Environ. Earth Sci.*, vol. 4, no. 5, pp. 10–16, Oct. 2023.

[20] M. M. Hossain, Z. R. Chowdhury, S. M. R. H. Akib, M. S. Ahmed, M. M. Hossain, and A. S. M. Miah, "Crime text classification and drug modeling from Bengali news articles: A transformer network-based deep learning approach," in *Proc. 26th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2023, pp. 1–6.

[21] M. A. R. Tushar and A. S. M. Miah, "Bantrafficnet: Bangladeshi traffic sign recognition using a lightweight deep learning approach," *Comput. Vis. Pattern Recognit.*, vol. 1, pp. 1–16, Jun. 2024.

[22] M. M. R. Tusher, F. A. Farid, M. Al-Hasan, A. S. M. Miah, S. R. Rinky, M. H. Jim, S. Mansor, M. A. Rahim, and H. A. Karim, "Development of a lightweight model for handwritten dataset recognition: Bangladeshi city names in Bangla script," *Comput., Mater. Continua*, vol. 80, no. 2, pp. 2633–2656, 2024.

[23] T. He, Q. Zhou, and Y. Zou, "Automatic detection of age-related macular degeneration based on deep learning and local outlier factor algorithm," *Diagnostics*, vol. 12, no. 2, p. 532, Feb. 2022.

[24] S. Naz, A. Ahmed, M. U. Akram, and S. A. Khan, "Automated segmentation of RPE layer for the detection of age macular degeneration using OCT images," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Dec. 2016, pp. 1–4.

[25] P. M. Arabi, N. Krishna, V. Ashwini, and H. M. Prathibha, "Identification of age-related macular degeneration using OCT images," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 310, pp. 1–18, Jun. 2018.

[26] A. Thomas, P. M. Harikrishnan, R. Ramachandran, S. Ramachandran, R. Manoj, P. Palanisamy, and V. P. Gopi, "A novel multiscale and multipath convolutional neural network based age-related macular degeneration detection using OCT images," *Comput. Methods Programs Biomed.*, vol. 209, Sep. 2021, Art. no. 106294.
[27] M. M. Sharif, M. Usman Akram, and A. W. Malik, "Extraction and analysis of RPE layer from OCT images for detection of age related macular degeneration," in *Proc. IEEE 20th Int. Conf. E-Health Netw., Appl. Services (Healthcom)*, Sep. 2018, pp. 1–6.
[28] V. Melnychuk, E. Faerman, I. Manakov, and T. Seidl, "Matching the clinical reality: Accurate OCT-based diagnosis from few labels," 2020, *arXiv:2010.12316*.
[29] Kaggle. (2021). *Retinal Oct-C8*. Accessed: Aug. 19, 2024. [Online]. Available: https://www.kaggle.com/datasets/obulisainaren/retinal-oct-c8/data
[30] Z. Li, Y. Han, and X. Yang, "Multi-fundus diseases classification using retinal optical coherence tomography images with Swin transformer V2," *J. Imag.*, vol. 9, no. 10, p. 203, Sep. 2023.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
[32] M. Chetoui and M. A. Akhloufi, "Deep retinal diseases detection and explainability using OCT images," in *Proc. Int. Conf. Image Anal. Recognit.*, 2020, pp. 1–22.
[33] T. Tsuji, Y. Hirose, K. Fujimori, T. Hirose, A. Oyama, Y. Saikawa, T. Mimura, K. Shiraishi, T. Kobayashi, A. Mizota, and J. Kotoku, "Classification of optical coherence tomography images using a capsule network," *BMC Ophthalmology*, vol. 20, no. 1, pp. 1–17, Dec. 2020.
[34] S. Kaymak and A. Serener, "Automated age-related macular degeneration and diabetic macular edema detection on OCT images using deep learning," in *Proc. IEEE 14th Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Sep. 2018, pp. 265–269.
[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

MD ZAHIN MUNTAQIM received the Bachelor of Science (B.Sc.) (Eng.) degree in computer science and engineering (CSE) from Bangladesh Army University of Science and Technology (BAUST), Saidpur, Nilphamari, in January 2024, where he is currently pursuing the master's degree in artificial intelligence (AI).

He is also proficient in multiple programming languages, particularly in Java and C++. He has expertise in data structures, algorithms, and software engineering principles. His strong teamwork and leadership abilities are evidenced by his experience in project management and collaborative development. His strong teamwork and leadership abilities are evidenced by his experience in project management and collaborative development. He aspires to work as a Research Scientist or a Software Engineer, contributing to cutting-edge technologies that have the potential to transform industries and improve lives.

TANGIN AMIR SMRITY received the Bachelor of Science (B.Sc.) (Eng.) degree in computer science and engineering (CSE) from Bangladesh Army University of Science and Technology (BAUST), Saidpur, Nilphamari, in January 2024, where she is currently pursuing the master's degree in computer science and engineering. She possesses a profound passion for utilizing her programming skills, algorithmic thinking, and problem-solving abilities. She is also proficient in several programming languages, including Python, C++, and C. Her research interests include deep learning, image processing, computer vision, machine learning, and natural language processing (NLP). She loves tackling challenges and has worked on projects & thesis. She enjoys tackling complex challenges and has gained valuable experience through various projects and her thesis work, which have enhanced her adaptability and creativity. She is also passionate about fostering collaboration and aims to contribute meaningfully to the ever-evolving field of computer science, both as an Educator and a Researcher.

ABU SALEH MUSA MIAH (Member, IEEE) received the B.Sc. (Eng.) and M.Sc. (Eng.) degrees (Hons.) in computer science and engineering from the Department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh, in 2014 and 2015, respectively, and the Ph.D. degree in computer science and engineering from The University of Aizu, Japan, in 2024, under a scholarship from Japanese Government (MEXT). He assumed the positions of a Lecturer and an Assistant Professor with the Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology (BAUST), Saidpur, Bangladesh, in 2018 and 2021, respectively. He has been a Visiting Researcher (Postdoctoral Researcher) with The University of Aizu, since April 2024. His research interests include AI, ML, DL, human activity recognition (HCR), hand gesture recognition (HGR), movement disorder detection, Parkinson's disease (PD), HCI, BCI, and neurological disorder detection. He has authored or co-authored more than 57 publications in widely cited journals and conferences.

HASAN MUHAMMAD KAFI (Member, IEEE) received the B.Sc. (Eng.) degree in computer science and engineering from the Department of Computer Science and Engineering, Chittagong University of Engineering and Technology (CUET), Chattogram, Bangladesh, in 2010, and the M.Sc. (Eng.) degree in information and communication technology from the Institute of Information and Communication Technology, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2017, where he is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering. He began his career as a Software Engineer with L2N Software Ltd., in 2010, and was promoted to a Senior Software Engineer, in 2013. In 2017, he was appointed as a Lecturer with the Department of Computer Science and Engineering, Sheikh Fazilatunnesa Mujib University, Jamalpur, Bangladesh. Then, he became a Lecturer with the Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology (BAUST), Saidpur, in 2019. He was later promoted to an Assistant Professor with the Institute of Information and Communication Technology. His research interests include machine learning and deep learning, computer vision, and data science.

TAOSIN TAMANNA received the Bachelor of Science degree in computer science and engineering (CSE) from Bangladesh Army University of Science and Technology (BAUST), Saidpur, Nilphamari, in January 2024, where she is currently pursuing the Graduate degree. Both the Secondary School Certificate and the Higher Secondary School Certificate completed, in 2016 and 2019. She has completed the computer security, digital image processing, data warehouse and data mining, artificial neural networks and fuzzy logic systems, and machine learning courses. Her research interests include machine learning, image processing, and networking.

FAHMID AL FARID (Member, IEEE) received the B.S. degree in computer science and engineering from the University of Chittagong, Bangladesh, in 2010, the M.S. degree from the Faculty of Computer Science and Electrical Engineering, University of Ulsan (UOU), South Korea, in 2015, and the Doctor of Philosophy (Ph.D.) (by Research) degree in information technology from Multimedia University, Cyberjaya, Malaysia. He is currently a Postdoctoral Scientist with the Faculty of Engineering, Multimedia University. From 2013 to 2014, he was a Research Assistant with the Embedded System Laboratory, UOU. In 2015, he was a Research Assistant with the Ubiquitous Computing Technology Research Institute (UTRI), Sungkyunkwan University, South Korea. His current research interests include artificial intelligence, algorithm design, computer vision, human–computer interaction, image, video analysis, power generation, and green technology. He received the Korean BK21 PLUS Scholarship, supported by Korean Government in M.S. (2012–2014). He also received an ICT Fellowship from Bangladesh Government, in 2014.

MD ABDUR RAHIM received the Bachelor of Science (Hons.) and Master of Science (M.Sc.) degrees in computer science and engineering from the University of Rajshahi, Bangladesh, in 2008 and 2009, respectively, and the Ph.D. degree from the Graduate School of Computer Science and Engineering, The University of Aizu, Fukushima, Japan, in 2020. He is currently an Associate Professor and the Head of the Department of Computer Science and Engineering, Pabna University of Science and Technology, Pabna, Bangladesh. His research interests include human–computer interaction, pattern recognition, computer vision and image processing, human recognition, and machine intelligence. He has several publications in major journals (SCI and SCIE indexed) and conferences and also serves as a reviewer for several SCI/SCIE indexed journals and international conferences.

HEZERUL ABDUL KARIM (Senior Member, IEEE) received the B.Eng. degree in electronics with communications from the University of Wales Swansea, U.K., in 1998, the M.Eng. degree in science from Multimedia University, Malaysia, in 2003, and the Ph.D. degree from the Center for Communication Systems Research (CCSR), University of Surrey, U.K., in 2008. He is currently a Professor with Multimedia University, where he is also the Deputy Dean of Student Affairs and Alumni with the Faculty of Engineering. He has been teaching multimedia and computing engineering subjects. His research interests include telemetry, 2D/3D image/video coding and transmission, error resilience, algorithms, bioinformatics, deep learning, machine learning, and neural networks.

SARINA MANSOR received the B.Eng. degree (Hons.) in electronics and in electronics and electrical engineering, majoring in computer from University College London, in 1998, the M.Eng.Sc. degree from Multimedia University (MMU), Malaysia, in 2002, and the D.Phil. degree in engineering science from the University of Oxford, U.K., in 2009. She is currently a Senior Lecturer with the Faculty of Engineering, MMU. She is also the Programme Coordinator for B.Eng. degree. Her research interests include signal and image analysis, medical imaging, computer vision, machine learning, and the Internet of Things.