RESEARCH ARTICLE

Enhancing Ocular Healthcare: Deep Learning-Based Multi-Class Diabetic Eye Disease Segmentation and Classification

MANEESHA VADDURI AND **P. KUPPUSAMY**, **(Member, IEEE)**
School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India
Corresponding author: P. Kuppusamy (kuppusamy.p@vitap.ac.in)

ABSTRACT Diabetic Eye Disease (DED) is a serious retinal illness that affects diabetics. The timely identification and precise categorization of multi-class DED within retinal fundus images play a pivotal role in mitigating the risk of vision loss. The development of an effective diagnostic model using retinal fundus images relies significantly on both the quality and quantity of the images. This study proposes a comprehensive approach to enhance and segment retinal fundus images, followed by multi-class classification employing pre-trained and customized Deep Convolutional Neural Network (DCNN) models. The raw retinal fundus dataset was subjected to experimentation using four pre-trained models: ResNet50, VGG-16, Xception, and EfficientNetB7, and the optimal performing model EfficientNetB7 was acquired. Then, image enhancement approaches including the green channel extraction, applying Contrast-Limited Adaptive Histogram Equalization (CLAHE), and illumination correction, were employed on these raw images. Subsequently, image segmentation methods such as the Tyler Coye Algorithm, Otsu thresholding, and Circular Hough Transform are employed to extract essential Region of Interest (ROIs) like optic nerve, Blood Vessels (BV), and the macular region from the raw ocular fundus images. After preprocessing, the model is trained using these images that outperformed the four pre-trained models and the proposed customized DCNN model. The proposed DCNN methodology holds promising results for the Cataract (CA), Diabetic Retinopathy (DR), Glaucoma (GL), and NORMAL detection tasks, achieving accuracies of 96.43%, 98.33%, 97%, and 96%, respectively. The experimental evaluations highlighted the efficacy of the proposed approach in achieving accurate and reliable multi-class DED classification results, showcasing the promising potential for early diagnosis and personalized treatment. This contribution could lead to improved healthcare outcomes for diabetic patients.

INDEX TERMS Deep convolutional neural network, diabetic eye diseases, image enhancement, image segmentation, retinal fundus images.

I. INTRODUCTION

According to the World Health Organization (WHO), around 2.2 billion people throughout the world are limited vision or visually challenged [1]. Among them, at least 1 billion are avoidable. It is believed that diabetes mellitus usually called diabetes has a role in these occurrences of blindness [2]. Most

The associate editor coordinating the review of this manuscript and approving it for publication was Carmen C. Y. Poon.

people with diabetes will eventually develop DED, and due to its high sensitivity in the diagnosis of DED, retinal fundus imaging has become the most widely used technology for detecting DED [2].

DED encompasses CA, DR, GL, and some examples of lesions that must be recognized from retinal images are shown in Fig. 1 These include deterioration of the lens (CA), abnormal BV growth and, narrow bulges or the retina's tiny BV rupturing (microaneurysms), (DR) in its

earliest stages, Low intraocular pressure (GL) is the leading cause of irreversible optic nerve damage and blindness. To effectively treat these conditions, accurate diagnosis and identification are essential [1], [2]. Inspiring proactive solutions for detection and prevention that fulfill many needs associated with retinal diseases and visual disabilities throughout a person's life. The application of Deep Learning (DL) in automated DED diagnostics is crucial for solving these problems [3], [4]. Professional ophthalmologists agree that timely screening for DED is essential for an effective diagnosis, but this screening takes a lot of time and effort [5].
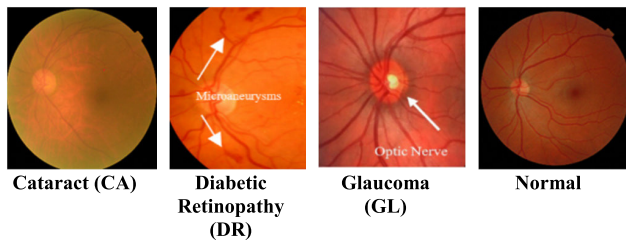

FIGURE 1. Fundus images with problems caused by DED.

While DL has shown outstanding validation accuracies for binary (healthy or diseased) classification, findings for moderate and multi-class classification have been lower striking, especially for mild impairment. Therefore, this study introduces an automatic multi-class DED classification model based on DCNN that can distinguish normal from diseased tissue in images. First, a comparison of diverse Convolutional Neural Network (CNN) architectures is conducted to determine the optimal one for classifying mild and multi-class DED. This model's goal is to improve upon the already impressive performance levels observed in the aforementioned works. Therefore, moderate and multi-class classification models were trained and tested to enhance sensitivity for the different multi-class DED. This involved implementing various pre-processing and augmentation strategies to enhance result accuracy further and ensure a sufficient sample size for the dataset. Treating ocular diseases as soon as possible is crucial, but doing so with the aid of neural networks consumes a significant amount of time and storage space.

Rapid diagnosis and treatment of retinal diseases are essential, but doing so with the use of neural networks is resource-intensive. Because of this, a relatively pre-trained model can improve the process by adjusting the design to cut down on losses. Pre-trained CNN networks are useful in DL because they allow knowledge to be transferred from one task to another with a smaller set of data or less time spent on training [6]. Fine-tuning the pre-trained network is widely recognized as a prominent strategy in transfer learning. It is standard practice to apply various preprocessing techniques to image datasets, including resizing, quantifying, standardizing, and enhancing images. These steps are taken prior to training CNN architectures, regardless of whether the training employs a pre-existing model or a newly developed

model. Improving the CNN model's classification accuracy is an endless pursuit, moreover, the model's accuracy relies heavily on the quality of both the training dataset and the images within it.

A. MOTIVATION
The recent advances in the domains of artificial intelligence, DL, and the computer vision have allowed DL to be applied to produce outstanding outcomes in image categorization and vision applications. Early detection of lesions and anomalies in ocular fundus images is still an outstanding issue. They found that 93% of moderate cases are incorrectly categorized as normal eyes and that deep neural networks have trouble learning enough detailed information to recognize components of mild disease [7]. Therefore, this study presents a system that combines standard image processing methods with the most cutting-edge CNN to assess multi-class DED.

B. CONTRIBUTIONS
The contributions made by this research are as follows:
- Integrate a holistic strategy for the accurate diagnosis of multi-class DED through the utilization of retinal fundus images. This approach encompasses image enhancement, segmentation, and classification techniques to achieve enhanced diagnostic accuracy.
- Employ four pre-trained models, ResNet50, VGG-16, Xception, and EfficientNetB7, and experiment with the raw ocular fundus dataset, and acquire the optimal performing model.
- Develop a new customized DCNN model, and train using images of the retina that have undergone pre-processing and segmentation.
- Investigate and compare the pre-trained optimal model and the new customized DCNN model. This demonstrated the significance of the preprocessing steps in improving the overall classification accuracy.

II. LITERATURE REVIEW
To spot DED in ocular fundus images early on, clinicians need a method that lets them see a full complement of features and pinpoint their precise location within the image [8]. Lens degeneration, dilated BV (microaneurysms), vascular leakage, and impairment of the optic nerve, all need to be present on retinal fundus images to diagnose multi-class DED in diabetic individuals. Fig. 1 depicts the progression of DED.

Previously, automated DED diagnoses were examined to reduce ophthalmologist's workload and improve the consistency of diagnosis [9]. Lesion-based detection has been applied in previous research; for example, a novel model was proposed for identifying microaneurysms in ocular fundus images. Methods such as BV segmentation, localization, and elimination of the fovea are used as part of their preprocessing effort. Following that, a hybrid system comprising neural networks and fuzzy logic models was employed to accomplish the aforementioned tasks of feature extraction and classification [5].

Their research looked at the problem of dividing DR into two groups defined by the presence or absence of microaneurysms. In addition, diagnosis of the DED can be made with a variety of additional features than microaneurysms. Similar to how a classification model based on pixels was presented to evaluate the intensity of ocular illness after segmenting the affected area and pinpointing the anomaly [10]. Used backpropagation neural networks fed data from decision trees and GA-CFS (Genetic Algorithm- Correlation based Feature Selection) methods to identify exudates in DR. Divided healthy eyes and those with exudates into two groups. The achieved results did not give sufficient classification accuracy and did not lead to effective noise removal [11].

Employed a Fuzzy C-Means algorithm and clustering analysis to create a method for identifying exudates. Optic Disc (OD) finding and cauterization of the BV are crucial to their work. The results obtained allow the exudates to be classified without relying on any defining criteria [12]. The technique presented relies on segmenting both the OD and the Optic Cup (OC). The suggested model makes use of two neural networks simultaneously operating with one focusing on the OC and the other on the OD. With the goal of proficiently segmenting, the suggested method targets the OD and the OC within an ocular fundus image. There are no available outcomes from a classification of GL in multiple stages [13]. The use of CNN to recognize DR in the fundus images was presented. They were able to achieve 90% specificity and sensitivity by using larger non-public datasets consisting of 80,000 to 120,000 ocular fundus images for binary classification between "normal," "mild," and "severe" [14], [15].

To identify retina BV 2D matching filters were used [16]. Gabor filter bank outputs were employed to automatically detect and classify anomalies in the vascular network, allowing the recognition of all stages of retinopathy [17]. There are numerous conventional methods for diagnosing and categorizing DED. The majority of methods make use of Fuzzy C-Means clustering, region-of-interest algorithms, mathematical morphology, neural networks, pattern recognition, and Gabor filtering methods [16], [17].

Numerous methods have been suggested to identify OD, and one such approach is the utilization of Principal Component Analysis (PCA) to determine potential optical disc areas by clustering pixels of a similar brightness. Hough Transform was utilized to detect optical discs [18]. An artificial neural network-driven method is employed for exudate identification [19]. Exudate detection was carried out using a method based on Fuzzy C-Means clustering [20]. A computational intelligence-based method was utilized [21]. Automated categorization of DR is attained through the evaluation of distinct attributes, which encompass exudates, hemorrhages, microaneurysms, and BV. This classification process is carried out utilizing a support vector machine [22].

To address the constraints posed by manually crafted features and make them applicable across a range of medical imaging techniques, the adoption of DL-based approaches becomes a feasible option. These approaches entail the acquisition of critical features through learning and then integrating these feature-learning processes into the model development process [23], [24]. A DL approach was investigated to assess the degree of nuclear CA severity from slit-lamp images. This technique involves inputting image patches into a CNN to generate the local filters. Furthermore, higher-order features were extracted using a set of Recursive Neural Networks (RNNs). The grading of CA was achieved using Support Vector Regression [25]. A CA detection experiment was conducted, utilizing the Kaggle dataset of 200 images. In this study, AlexNet the CNN architecture was combined with various common classifiers, including Adaptive Moment Estimation (Adam), SGD, and others. The recommended system achieved a 77% accuracy when employing the Adam optimizer and an impressive 97.5% accuracy when utilizing the Lookahead optimizer with the AlexNet architecture [26]. A unique CNN model architecture ("Cataract Net") was formulated, characterized by its compact size, reduced layers, and training parameters, as well as the use of smaller kernels to enhance computational efficiency. The approach demonstrated a remarkable accuracy of 99.13% for the two classes under study [27]. To identify CA severity from mild to severe, a computer-aided technique using fundus images was proposed. A CNN that had already been trained was transferred to the automated CA classification task as part of this strategy [28]. A classifier employing a Support Vector Machine (SVM) and achieving a four-stage Correct Classification Rate (CCR) of 92.91% was utilized for the classification task. A method for classifying CA disease known as Tournament-based Ranked CNN was introduced. This method employs a tournament structure along with binary CNN models for the classification process [29]. The CNNs and Res-Net-based trained classifier model enabled a system for automated CA identification with an accuracy of 95.78 percent [30].

Recently, a technique utilizing multiple models with attention mechanisms was presented for automated CA disease identification in ultrasound images, achieving an accuracy of 97.5% [31]. Using a pre-trained VGG-19 architecture on a dataset available on KAGGLE, a comparable accuracy of 97.47% was achieved for fundus images [32], [33].

People with diabetes may become limited vision from DR since it has no early warning signs. Yet, DR's effects may be mitigated with early diagnosis. Automated DR diagnosis and classification were suggested [34]. Pre-processing, segmenting of images, extraction of features, and categorization are all rolled into one using this approach. A technique for enhancing local contrast was used on the greyscale images to make the area of interest more visible. Using an adaptive threshold approach and mathematical morphology, the lesion area was accurately segmented. Finally, Enhanced categorization was achieved by merging statistical and geometric characteristics, leading to more accurate outcomes. Those with DR are at risk for developing retinal complications

including blood clots, lesions, and retinal hemorrhages. Retinal images are used to get a DR diagnosis. An approach for DR identification and categorization using a pre-trained CNN was developed. To improve the retrieved characteristics, a data refinement and augmentation technique was first used. Gaussian blur was used on the fundus image to decrease the quantity of noise in the picture. Accuracy was computed in the experimental setting [35].

An approach-based DL was suggested for DR classification, which would include the feature extraction of segmented fundus images. This method began with pre-processing the fundus image and then continued with segmentation. With the advent of the maximum principal curvature model, which prioritizes the greatest Eigenvalues, the branching blood veins can now be eliminated. To enhance the quality and eliminate inaccuracies within the region, morphological opening, and adaptive histogram equalization techniques were employed. Diabetes has been linked to increased optic nerve proliferation. The categorization of DR was carried out using a CNN which consists of three primary functional components: The investigation focused on the pooling layer, convolution layer, and the bottleneck layer. The results demonstrated a precision(pre) rate of 97.2% and an accuracy of 98.7%. Unfortunately, it was not possible to determine the duration of patients' distress [36].

Using an Adaptive machine-learning technique, a DR categorization model was created. By this method, DR pictures may be recognized using their own classifiers and characteristics. Diabetic Retinopathy Estimation (DRE) at the segment level was achieved by using a modified, previously trained CNN. After that, the categorization of DR images was established by connecting lines between all DR maps at each segmentation level. In addition, a learning method was used end-to-end to deal with the non-uniform lesions. Acquired sensitivity of 97% and a specificity of 96.37 %. Also, proliferative diabetic retinopathy was not taken into account by this approach [37].

Screening for DR by ophthalmologists is difficult and time-consuming due to blurred retinal images that make it difficult to see signs like microaneurysm, hemorrhage, etc. Because of this, a machine-learning technique that could automatically identify DR in fundus images. Classification of DR images using DL was made more precise by using a pre-processing improvement strategy. To improve the fundus image's clarity for the viewer, Histogram Equalization (HE), de-haze algorithm, and high pass filter were used. Four-layer convolution was used for image categorization. In the end, a satisfactory level of precision was achieved [38].

A context-aware graph network for tuberculosis detection was presented. Because of training limitations and overfitting issues, the traditional CNN model suffered greatly. For this reason, this study presents transfer learning-based strategies for extracting some of the sample-level characteristics. However, given the large number of pictures used for training, it was urged that EfficientNet be used as a pre-trained model. The categorization benefitted significantly from the

spatial relationship between feature vectors, which helped tremendously. Each of these feature vectors has the potential to supply information that is analogous to that provided by the vectors to either side. Here, a feature graph was provided to keep the image's spatial details intact [39].

Using a chaotic bat algorithm, a refined version of AlexNet as well as Ensemble Learning Model (ELM). Here, a pre-trained AlexNet is used, involving dataset training using images. The process of training the parameters was laborious and time-consuming. To make the AlexNet model more stable, Batch Normalization (BN) was implemented here. As an additional step, the AlexNet model had multiple layers replaced with the ELM. Thus, the model's precision improves as a result [40].

The initial presentation of DR detection through BV and OD segmentation, alongside the identification of retinal anomalies, was introduced. This approach encompasses three fundamental components: pre-processing, segmentation, and the classification, each playing a pivotal role. During the pre-processing stage, the CLAHE method was utilized to process and enhance the green channel component within the Red, Green, Blue (RGB) scale. Once the OD and the BV were primed for segmentation, Subsequent steps involved devising methodologies like the top hat transformation and the Gabor filtering to efficiently identify and isolate anomalies. Throughout the segmentation process, various attributes such as TEM (Texture Energy Measurement), Entropy, and LBP (Local Binary Pattern) were extracted. Moreover, the approach incorporated the Trial-dependent Bypass with an enhanced Dragonfly Algorithm (TB-DA) for optimal feature selection. For distinguishing between different severity levels (light, moderate, and severe), the hybrid neural network method was employed. The experimental outcomes were compared to established methods, assessing various metrics including accuracy, NPV (Negative Predictive Value), precision, FDR (False Discovery Rate), FNR (False Negative Rate), MCC (Matthews Correlation Coefficient), FPR (False Positive Rate), and the F1-score [41].

Recent studies have investigated the feasibility of employing automatic ocular processing of images for GL screening, with results that vary. The techniques covered below span a variety from simpler machine learning methods to more advanced ones, such as DL. GL has been detected using both open and combined datasets. Some research has tried to use a composite of retinal scans from several public sources to diagnose GL. For example, a combination of DRISHTI, and RIMONE V3 publicly available datasets extracted features from the OD and the optic cup to identify GL [42].

An automated GL diagnosis system using three distinct CNN model learning techniques, with results validated by ophthalmologists. The researchers utilized a wide array of neural networks, including Transfer Convolutional Neural Networks (TCNNs), Semi-Supervised Convolutional Neural Networks (SSCNNs) with self-learning, Denoising Auto Encoders (DAE) that relied on both labeled and unlabeled
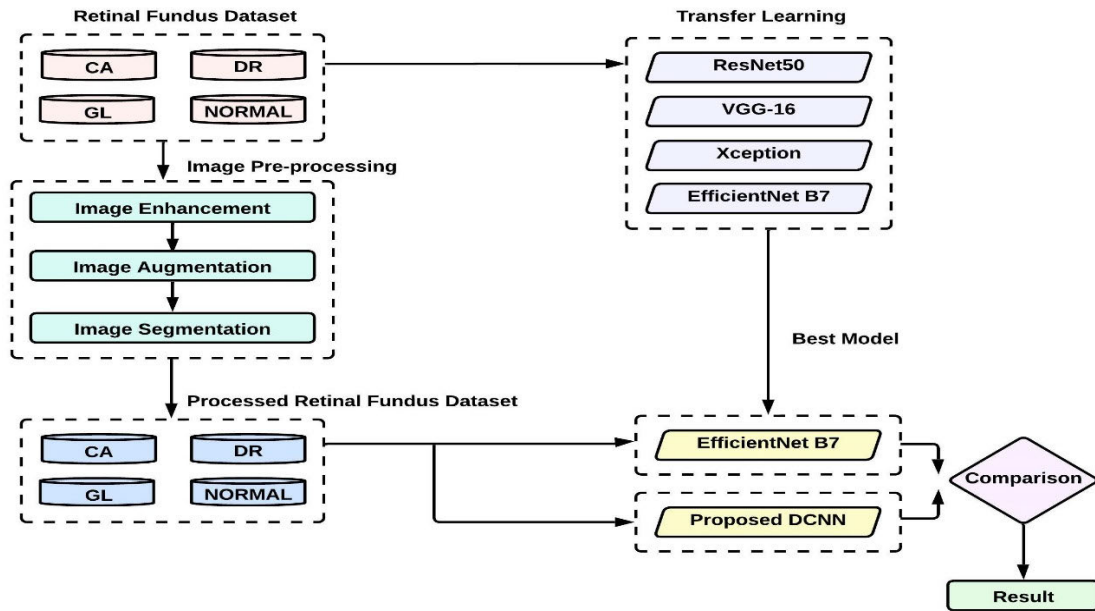
FIGURE 2. The overall process flow.

input data. Their models, when run on the RIMONE and RIGA open-source datasets, showed convincing results and proved that DL models are good at finding GL. The authors say that the TCNN, SSCNN, and SSCNN-DAE all had an overall accuracy of 91.5%, 92.4%, and 93.8%, respectively [43]. A transfer learning-based model was utilized to do automatic GL categorization. Color fundus pictures from the RIM-ONE and DRISHTI-GS databases were used. They added images from two more campaigns in Barcelona, Spain, to their original dataset. Subsequently, using a transfer learning method, they performed image preprocessing and fine-tuned five distinct CNN models. The study revealed that the VGG-19 architecture exhibited the most favorable performance, reaching an Area Under the Curve (AUC) of 94%, accompanied by a sensitivity of 87% and a specificity of 89% [44]. VGG16, VGG19, Xception, InceptionV3, and ResNet50 are pre-trained architectures on ImageNet for GL detection, eliminating the requirement for feature extraction or estimating geometric Optic Nerve Head (ONH) parameters like Cup-to-Disc Ratio (CDR). Combining five publicly accessible datasets of 1,707 fundus pictures created the ACRIMA dataset. The ACRIMA dataset, which contains 396 GL pictures and 309 normal eye images, performed at 0.7678 with an accuracy of 70.2% on the test dataset. The additional open-source datasets in this analysis (HRF, sjchoi86-HRF, RIM-ONE, and DRISHTI-GS1) have AUC values of 0.8354, 0.7739, 0.8575, and 0.8041 [45].

III. MATERIALS AND METHODS

The fundamental objective of this research is to enhance the efficiency of timely identification of multi-class DED utilizing ocular fundus images by experimentally evaluating image preprocessing and classification enhancement strategies.

The aims in this area may be summed up as follows:
- Deploy conventional techniques of image processing like improving image quality, expanding dataset through augmentation, and segmenting images.
- Exploring different model configurations to observe their impact on CNN outcomes.
- Compare the accuracy of the original and pre-processed fundus images using the pre-trained CNN models Xception, ResNet50, VGG-16, and EfficientNet B7.
- Training pre-processed fundus images involves utilizing a DCNN model to improve classification accuracy.
- Performance measures are utilized to assess and compare the outcomes of the pre-trained model with those of the new model.

The pipeline illustration in Fig. 2 illustrates the overall process flow. The dataset consisting of raw retinal fundus images was subjected to testing using four different pre-trained models, namely ResNet50, VGG-16, EfficientNet B7, and Xception, in order to identify the most effective Model. The raw fundus pictures were subjected to normal image processing techniques, and the dataset was then trained using the most effective model identified in the prior experiment. Additionally, a custom-built DCNN was employed to train pre-processed data. Ultimately, a comparison of outcomes was conducted to evaluate whether the execution accuracy of the models improved with the utilization of pre-processed images.

A. DATASET DESCRIPTION

The dataset comprises Retinal images categorized into CA, DR, GL, and NORMAL as shown in Fig. 3.
- Images were collected from publicly available datasets, including IDRiD, Oculur recognition, DRISHTI-GS,

Retinal Dataset on GitHub, Messidor, and the Messidor-2.

- Each of the images is labeled by ophthalmologists, and its lesion grade is determined based on new BV, hemorrhages, and microaneurysms.
- The Messidor Dataset comprises of 1200 ocular fundus images of the back part of the eye's interior, which were taken using a 3CCD color video camera attached to a Topcon TRC NW6 non-retinograph with a 45-degree Field of View (FOV). It was designed to facilitate computer-assisted DED studies.
- The Messidor-2 Dataset is an openly accessible dataset with 1,748 color images of retinas from 874 subjects. Each subject contributes two images, one for each eye. It uses International Clinical Diabetic Retinopathy (ICDR) and Diabetic Macular Edema (DME) grades to assign four disease rates per subject.
- The dataset known as DRISHTI-GS includes 101 ocular images, consisting of 31 normal and 70 showing GL-induced damage. To address limited images, an upsampling technique was used, selecting 1000 images from each class for experimentation.
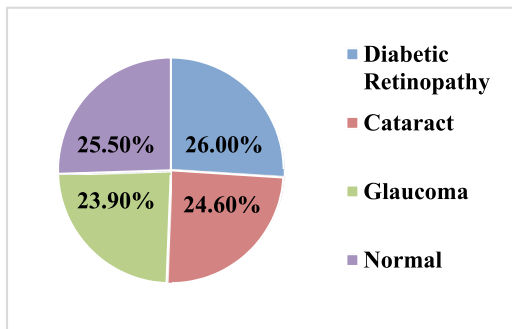

FIGURE 3. Data distribution.

B. IMAGE PRE-PROCESSING

The purpose of the pre-processing phase is to eliminate noise and irregularities from the ocular fundus image, thereby enhancing its quality and contrast. Along with contrast improvement, noise reduction, and image normalization, this pre-processing step can help mitigate irregularities and enhance the accuracy of subsequent stages in the process. In the quest to detect clinical features related to DED, image processing aims to elevate and refine the quality of the ocular fundus image. Fig. 4 shows a flowchart of the method for segmenting and processing images.

Furthermore, DED characteristics from fundus images are localized, retrieved, and segmented for further classification in pre-trained models. This section briefly discusses the preprocessing methods used in this study.

1) IMAGE ENHANCEMENT

Prior to processing, image-enhancing techniques were applied, including contrast enhancement and lighting adjustments, to improve the informational content and visual
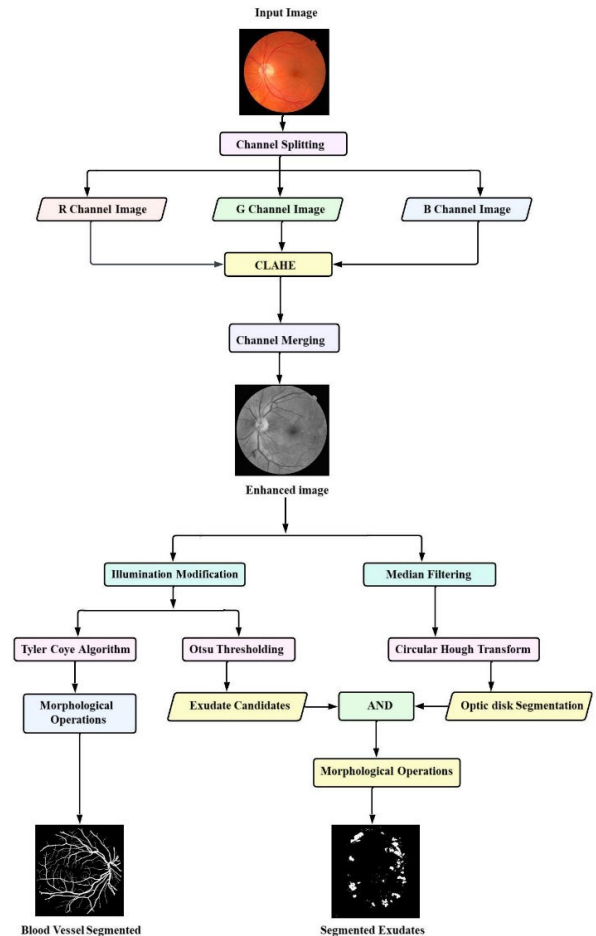

FIGURE 4. The workflow of data preprocessing.

quality of the original images. The technique of CLAHE [19], [61] are used to enhance the visual clarity of the images. The CLAHE technique constitutes a modified component inside the Adaptive Histogram Equalization (AHE) process. The suggested approach encompasses the application of the boosting function to each individual pixel inside the designated region, followed by the identification of the corresponding transformation function. This phenomenon exhibits dissimilarities in comparison to AHE due to its relatively diminished level of contrast. In CLAHE, Contrast-Limited Histogram Equalization (CLHE) is employed as a method to improve an image's contrast. This is achieved by applying CLHE to smaller regions of the image known as tiles, as opposed to the whole image. Bilinear interpolation is then used to put the tiles back together in a perfect way. CLAHE was used on grayscale images of the retina. A function called 'clip limit' is used to limit the amount of noise in an image, clip the histogram, and make a grey-level mapping. In the contextual area, the number of pixels is split evenly between each level of grey, in order to obtain an average pixel value that is grey, as indicated by:

$$N_{avg} = \frac{(N_{cr} - x_p) \times (N_{cr} - y_p)}{N_g} \tag{1}$$

where N_{avg} represents the number of pixels on average, N_g denotes the number of grey levels inside the contextual zone. $N_{cr} - x_p$ represents the amount of pixels in the contextual region's x direction. $N_{cr} - y_p$ represents the amount of pixels in the contextual regions y direction, then figure out the real clip limit.

$$N_{cl} = N_{clip} \times N_{avg} \qquad (2)$$

CLAHE [55] is a helpful method in biological image processing since it effectively highlights the key parts of an image as shown in Fig. 5
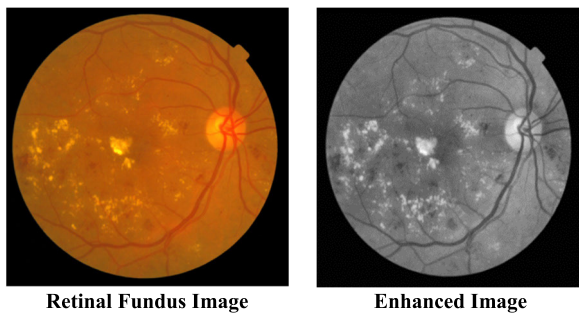

FIGURE 5. Sample retinal fundus image and Enhanced image.

Illumination Modification: This preprocessing approach attempts to minimize the scenario effect introduced by retinal images with inconsistent illumination [48]. The following formula is used to determine the intensity of each pixel:

$$p_i = p_0 + \mu_d - \mu_l \qquad (3)$$

where p_0 and p_i represent the initial and current pixel sizes, μ_d represents the target average intensity, and μ_l represents the local average intensity, respectively [49]. This procedure amplifies the appearance of formatted microaneurysms on the retinal surface.

2) IMAGE AUGMENTATION

DL models exhibit superior performance when provided with substantial volumes of data for learning purposes [50], [51]. Hence, the term "data augmentation" encompasses a group of procedures used to expand the training data size without adding any new examples. As a result, geometric changes including flipping, rotation, mirroring, and cropping are discussed as part of the picture augmentation methods covered in this study. Real-time image augmentation was facilitated using the Keras Image Data Generator class, ensuring that the selected model would obtain image variations during each iteration. In this study, the utilized Image Data Generator class possesses the capability to mitigate overfitting of the selected model by maintaining a consistent dynamic range in the generated images as compared to the originals.

3) IMAGE SEGMENTATION

While designing a classification system for DL-based moderate DED detection, it is critical to consider both the network design and input data quality. For the results to be accurate, the input image quality is a crucial element. The outcome of an automated disease diagnosis method for retinal fundus images is contingent on factors such as the number of images available, the image brightness and contrast, and the presence of anatomical characteristics. Therefore, the process of feature segmentation enhances the utility of images in classification tasks and contributes to the enhancement of accuracy. The procedure is used with the corresponding theoretical framework, is outlined below.

a: EXTRACTION OF BV

for diagnosing DR at its earliest stages, Retinal BV is a key anatomical characteristic in images of the retina. Following these stages accomplishes segmentation of retinal BV: Improved outcomes can be attained by the use of (i) image enhancement, (ii) Tyler Coye algorithm [52], and (iii) morphological operations [46].

After applying the aforementioned image processing methods, the green RGB channel provided the most effective comparison between the vascular network and the backdrop. The methods presented by Zuiderveld [19] and Youssif et al. [49] are used to estimate the contrast and brightness changes in a fundus image's backdrop. ISODATA in the Tyler Coye algorithm is then utilized to retrieve the threshold level once contrast and brightness have been adjusted. Morphological operation (erosion and dilation) was utilized to improve upon the Tyler Coye algorithm's work. These two basic procedures are crucial for eliminating background noise and filling in foreground details. The following equation depicts the process of erosion, which is utilized to eliminate or enhance the border of the region.

$$M \ominus N = \{p | N_p \subseteq M\} \qquad (4)$$
$$M \oplus N = \{x | N_x \cap X \neq 0 \qquad (5)$$
$$M \cdot N = (M \ominus N)N \qquad (6)$$

In which, the dilation is represented by $\ominus$, and the erosion is represented by where M is the structural element, and N is the dilatation of that set's erosion. Unfortunately, Tyler Coye algorithm still has a few gaps. As seen in Fig. 6, this morphological procedure fills up the microscopic gaps, covering a portion of the essential BV areas.
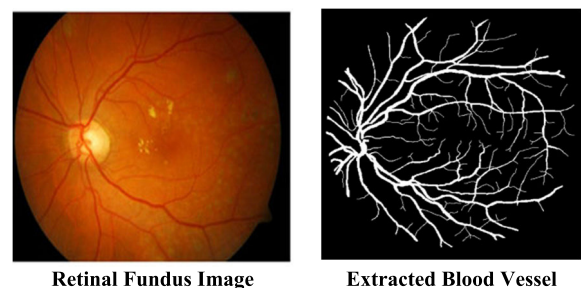

FIGURE 6. Sample retinal fundus image and extracted.

b: IDENTIFICATION AND EXTRACTION OF THE OD

GL is a condition that arises due to optic nerve injury. Segmentation of OD is a useful technique for investigating the sharper anatomical changes in the optic nerve. Fig. 7 displays anatomically accurate retinal fundus images obtained from the data set including the OD. The CHT (Circular Hough Transform) was employed to identify the circular objects, and then the median filter was utilized to reduce the noise, and threshold values were applied to segment the OD, as depicted in Fig. 8, for the purpose of OD segmentation. CLAHE can only be applied on a specified section, or "tile," of the image. It cannot be used on the entire image.
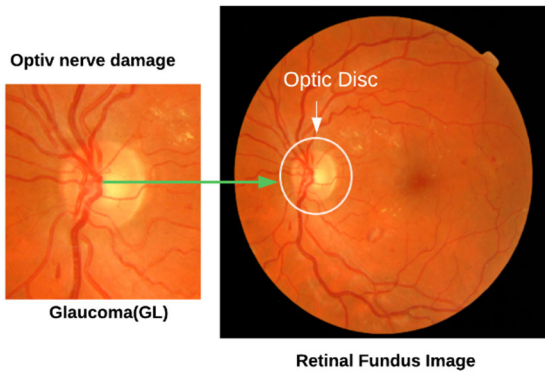

FIGURE 7. Sample retinal fundus image and optical nerve damage in Glaucoma (GL).

Setting the maximum contrast rate to $L, 0 \leq L \leq L$ [53] adapts the image enhancement computation to the user-specified maximum contrast level. Additional contrast enhancement is applied to images with low contrast measured by

$$\phi(a, b) = \left(\frac{\mu(a, b)}{\delta - \Delta} \right) (\Gamma - 1) \quad (7)$$

where, $\phi(a, b)$ and $\mu(a, b)$ denote the pixels after transformation and the pixels before transformation in the (a, b) coordinates, respectively. Δ is the highest pixel value, δ is the lowest pixel value of the input image and Γ is the highest value of the grayscale image.

The use of median filtering is prevalent in the domain of image processing due to its notable efficacy in reducing noise. Using the median filtering, the median pixel value of the window is used to replace the value at the window's midpoint. Median filtering may be expressed mathematically as,

$$F(a, b) = median_{(s,t) \in S_{ab}} \{g(s, t)\} \quad (8)$$

Extraction of objects or segmented areas with similar attributes from the background is the goal of segmentation, using a pixel classification approach [54], [62]. Consequently, the identification of OD was facilitated using the CHT. Circular shapes in images are easy targets for the CHT technique. The CHT method improves compared to alternatives because the model demonstrates a significant level of sensitivity to variations in the feature specification

descriptions while being moderately resistant to the presence of image noise. The computation of the CHT is performed using the following formula:

$$(m - a)^2 + (n - b)^2 = c^2 \quad (9)$$

The following are the stages in the process of circle detection (i) the image's binary edges are extracted, (ii) the parameters 'a' and 'b' are given values, (iii) determine the radius value of 'c', (iv) modify the accumulator in accordance with (a), (b), and (c), (v) within the scope of interest, replace 'a' and 'b' values and proceed to stage that computes 'c'.

Retinal Fundus Image   **Segmented Optic Disk**
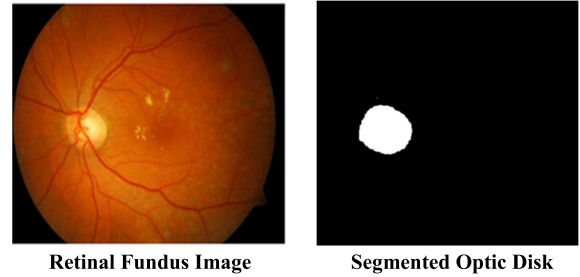
FIGURE 8. Sample retinal fundus image and segmented optic disc.

c: LOCALIZATION AND DETECTION OF EXUDATE

Exudates may be seen as bright patches of varied size, brightness, position, and form in two-dimensional ocular images taken using a digital fundus camera. Accurate exudate segmentation is difficult because of the vast variation in exudate size, intensity, contrast, and shape. Given the wide range of size, intensity, contrast, and form, precise segmentation of exudates is a challenging task. There are three main processing processes: (1) improving image quality; (2) detecting and eliminating the OD; (3) eliminating BV; and (4) extracting exudates. Classification of DR can be accomplished by applying the evaluation standards outlined in the Messidor dataset after exudates have been obtained from the mild dataset. Fundus images may be utilized to identify the existence of the exudates, allowing for a timely diagnosis of early DR. When the OD is found and detached, Otsu thresholding is used to identify potential exudate regions. The Otsu technique can automatically estimate a threshold value of T from the provided input ocular image. Then, (10) represents the histogram uses to calculate its intensity value,

$$P(i) = \frac{n_i}{N}, P(i) \geq 0, \quad \sum_{1}^{256} P(i) = 1 \quad (10)$$

The number of pixel images N, as well as the number of pixels n_i with intensity I. Equations (11) and (12) describe the subject weight and background.

$$W_1(t) = \sum_{i=1}^{t} P(i) \quad (11)$$

$$W_2(t) = \sum_{i=t+1}^{L} P(i) = 1 - W_1(t) \quad (12)$$

Here, the gray level number is L. The background and the object mean is determined by using (13) and (14) respectively.

$$M_1(t) = \sum_{i=1}^{t} i \cdot \frac{P(i)}{W_1(t)} \tag{13}$$

$$M_2(t) = \sum_{i=1}^{t} i \cdot \frac{P(i)}{W_2(t)} \tag{14}$$

Thus, variance is evaluated by (15), (16) respectively, while the (17) represents the expression for the sum of variance.

$$\sigma_1^2(t) \sum_{i=1}^{t} (1 - M_1)^2 \cdot \frac{P(i)}{W_1(t)} \tag{15}$$

$$\sigma_2^2(t) \sum_{i=1}^{t} (1 - M_2)^2 \cdot \frac{P(i)}{W_2(t)} \tag{16}$$

$$\sigma^2(t) = \sigma_W^2(t) + \sigma_B^2(t) \tag{17}$$

Thus, σ_W^2 is referred to as the WVC (Within-Class Variance) and represented in the (18), where σ_B^2 is referred to as the BVC (Between-Class Variance) and is represented in (19). The WVC is the total amount of variance between classes after the probability of each class has been applied to the total amount of variation. Equation (20) is used to compute the average total. The threshold value may be reached by minimizing WVC or maximizing BVC, however BVC requires less computing time.

$$\sigma_W^2(t) = W_1(t) \cdot \sigma_1(t)^2 + W_2(t) \cdot \sigma_2(t)^2 \tag{18}$$

$$\sigma_B^2(t) = W_1 \cdot [M_1(t) - M_T]^2 + W_2 \cdot [M_2(t) - M_T]^2 \tag{19}$$

$$M_T = \sum_{i=1}^{N} i.p(i) \tag{20}$$

Morphology encompasses a group of distinct parameters that pertain to the pixel entity inside an image, using logical operations such as "or", "and". The opening procedure seeks to remove pixel areas that are smaller than structural elements and refine and restore object shape. Equation (21) is used to represent opening operation.

$$MoN = (M \ominus N) \oplus N \tag{21}$$

The segmentation of exudates in the macula is shown in Fig. 9.

C. TRANSFER LEARNING

This study employs CNN-based transfer learning to establish a classification method for DED retinal fundus images. Transfer learning strategies are explored, leveraging pre-trained CNN models to achieve optimal classification outcomes. The following section will provide an in-depth exploration of the specifics related to the pre-trained models. Pan et al. [55] provide the following definition of transfer learning: $D = \Phi, P(X)$ where $X = x_1, x_2, \ldots, x_n \epsilon \Phi$, where D is the domain, Φ is referred to feature space, and $P(X)$ is the

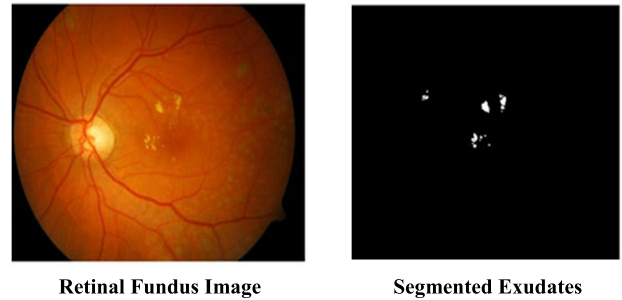**Retinal Fundus Image**   **Segmented Exudates**

FIGURE 9. Sample retinal fundus image and segmented exudates.

marginal distribution of probabilities. $T = Y, F(\cdot)$ is a learnt objective predictive function from the feature vector and label pairs, where T is the task and Y is the label space.

To be more precise, given D_s, a source domain and T_s, a learning task, and D_t, a target domain and T_t, a learning task, transfer learning is a procedure of enhancing the target predictive function learning $F_t(\cdot)$ in D_t based on the knowledge gained from the source domain D_s and the learning task T_s, where $D_s \neq D_t$, or $T_s \neq T_t$. It is important to acknowledge that the aforementioned single source domain has the potential to include a multitude of other source domains.

In image classification, transfer learning is based on the idea that a neural network performs better when it is given a large and varied dataset to learn from, such as ImageNet, it can effectively excel in a particular target task, despite the other having fewer labelled instances than the pre-training dataset. Using these acquired feature maps is advantageous in comparison to building a massive architecture from the scratch using a massive dataset.

In this research, two approaches will be employed to fine-tune existing trained models: (1) Feature extraction, the process of using features discovered in the primary task to draw out pertinent characteristics from the destination task. To adapt the feature mappings learned from the sample data, a new classifier was layered atop the pre-trained network, with the option for training from scratch.

(2) The fine-tuning process, certain previously frozen layers within the base network are unfrozen, allowing training of these unfrozen layers concurrently with the newly introduced classifier layers. This fine-tuning procedure refines the base network's higher-level feature representations to make them better fit to the target task. To accomplish DED image classification, four CNN models that have been pre-trained include ResNet50, VGG-16, Xception, and EfficientNetB7 are fine-tuned. The properties of four Image-Net pre-trained CNN networks are listed in Table 1.

IV. PROPOSED DCNN ARCHITECTURE

In order to classify medical visual abnormalities, CNNs are the most often used DL method [56]. This is because CNN maintains individual characteristics when examining input images. The following discussion highlights the relevance of spatial connections in retinal images, such as the location

TABLE 1. Three CNN models pre-trained using ImageNet and its features.

Model	Size	Parameters
ResNet50	98MB	25,600,000
VGG-16	528MB	138,357,544
Xception	88MB	22,910,480
EfficientNetB7	256MB	66,700,000

of BV rupture or the buildup of a yellowish fluid in the macula. Fig. 10 depicts the whole procedure architecture. Fundus images that have been processed using a DCNN are automatically probed for their feature patterns, using the network's many layers and filters. The suggested framework comprises a set of 2D convolutional layers, max-pooling layers, and the batch normalization layers. These components have been fine-tuned with carefully selected hyperparameters to effectively capture features from input fundus images spanning various categories. To facilitate the diagnosis of eye diseases, we incorporated a fully connected layer to serve as a classifier, which accepts the feature maps generated by the CNN as input. The network consists a total of seventeen weighted layers, constituting the proposed model. This includes fourteen convolutional layers, two fully connected layers, and one classification layer. Additionally, the network is enhanced using batch normalization, max-pooling, dropout, and flattening. In order to classify fundus images, the most crucial and difficult step is to extract features from them. In contrast to numerous manual and machine learning-based methods for extracting features, deep neural networks serve as automated feature extractors. The convolutional operation, represented by ($\times$), is a built-in function and a fundamental component of deep neural networks, essential for feature extraction. Mathematically, it involves multiplying two functions (a and b) to generate a third function ($a \times b$). The use of a $k \times k$ window size or kernel in convolution is preferred, with k ideally being an odd integer for improved symmetry around the origin and reduced aliasing errors. Convolutional layers store high-level extracted features, with the kernel sliding across image pixels to produce feature maps for each of the N filters in every layer. If the input dimension of the fundus image is ($P1 \times P2$) and N kernels with a $k \times k$ window are employed, the resulting image shape will be $N \times ((P1 - K + 1) \times (P2 - K + 1))$. This iterative process continues until precise feature patterns are extracted from the input fundus image. The design parameters proposed in this approach are carefully selected in a systematic manner to fine-tune the DL model and achieve effective results. Several locations from the parameter combination are uniformly picked to provide the best possible hyperparameter combinations. The best parameter for controlling the dataset's complexity is determined by cross-validation for every feasible parameter combination.

The proposed DCNN architecture employed a constant fundus image size of pixels, utilizing a 3×3 filter window size across the entire network. This size choice provides a relatively limited visual field, but it was sufficient for

preserving the image's indications of vertical and horizontal orientation, as well as its $224 \times 224 \times 3$ central features. In the convolutional layers of this proposed network, a stride value of 1 pixel was applied, causing the kernel to shift by 1 pixel when padding was employed to retain information at the image borders. Since the network-wide padding value is uniform, an extra 1 pixel is appended to all four image borders.

Training a deep neural network gets more difficult as the number of parameters increases. To address this issue, pooling layers are commonly employed to decrease the parameter count. One popular pooling technique is max pooling, where a window slides across the feature map generated from ocular fundus images, selecting the highest point value within the window. This method is often favored over other pooling algorithms. In the proposed architecture, five max-pooling layers were incorporated at different points after sets of convolutional blocks. These max-pooling layers utilize a 2×2 pixel window, a stride of 2, and maintain the same padding. After each max-pooling layer is applied, DCNN increases the total number of filters in use from 32 to 512 through a series of weighted block configurations. Since the input feature mapping can change as the network's weights are updated during training, this can add complexity to training a deep neural network. Therefore, the proposed architecture incorporates batch normalization, as it helps mitigate this issue [57]. Batch normalization works by standardizing and normalizing the input to a layer based on mini-batches of data instead of the entire training dataset. This approach enhances the robustness of the neural networks. It effectively tackles the problem of the internal covariate shift by ensuring that the input to each layer maintains a consistent mean and standard deviation, which are representative of a normal distribution. Gradients are less sensitive to changes in their starting values and parameter sizes after being normalized in batches. It initiates training for a deep neural network with the activation function having a Gaussian distribution of one unit. The DCNN was constructed using the optimizer (Adam).

The training loss was minimized by optimizing the learning rate and weights using the Sparse Categorical Cross-Entropy function, which was employed along with the Adam optimization function. The proposed architecture uses dropout for regularization to reduce the possibility of overfitting in situations when the system is required to make a decision relying on an exceptionally extensive set of parameters.

In order to retain neurons that can store patterns related to eye diseases, a more pronounced utilization of dropout is necessary during the classification phase. This differs from the convolutional layer blocks responsible for feature extraction [58]. To ensure training regularization, a dropout value of 0.5 is employed in the initial two fully connected layers, and a batch size of 32 is used.

Rectified Linear Units (ReLU) [59] are used to activate all of the proposed architecture's intermediate layers, and the
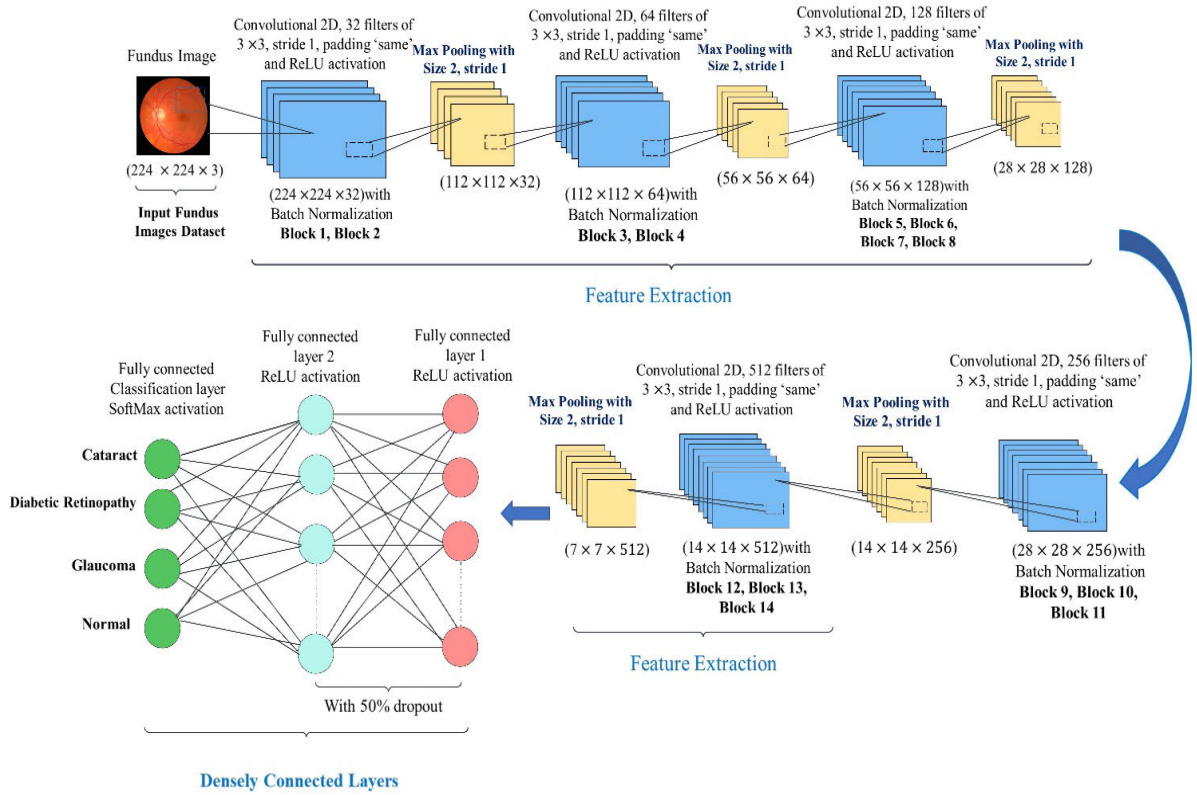
FIGURE 10. The proposed DCNN layered architecture.

Softmax function [47] is used to activate the network's output layer since the dataset is nonlinear. ReLU is a nonlinear activation function that outperforms Sigmoid and Tanh in terms of performance and convergence speed. As shown in (22) below, ReLU rectifies all the values that are negative in the extracted feature map, which boosts accuracy and shortens the training time.

$$\text{ReLU}(a) = \begin{Bmatrix} 0, & if\ a \leq 0 \\ a, & if\ a \geq 0 \end{Bmatrix} \quad (22)$$

In the network's output layer, Softmax activation function was employed to transform the outcome into probabilities for the classification of ocular fundus images into four distinct categories: CA, DR, GL, and NORMAL. The initial convolutional layer used 32 filters for feature extraction, with an input shape of (224 × 224 × 3).

Throughout the proposed architecture, all convolutional layers share common characteristics: they have a kernel size of 3 × 3, use the same padding, employ a stride value of 1, and utilize the ReLU activation function as specified in (23).

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3n} \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} \end{bmatrix} \times \begin{bmatrix} K_1 & K_2 & K_3 \\ K_4 & K_5 & K_6 \\ K_7 & K_8 & K_9 \end{bmatrix}$$

Image Kernel

$$\rightarrow \begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix} \quad (23)$$

Extracted features

After every convolutional layer, batch normalization is used to standardize and normalize the output of each convolutional layer for training, as specified in (24).

$$a \rightarrow \hat{a} = \frac{a - \mu}{\sigma} \rightarrow b = \gamma \hat{a} + \beta \quad (24)$$

In this context, (μ, σ) represents the mean and the standard deviation of a specific parameter within the β-shifted minibatch. Algorithm 1 determines the steps involved in minibatch batch normalization in detail.

The second convolutional layer also received identical configurations, featuring 32 filters. To process the output from the previous layer and decrease the dimensionality of the feature maps, a max-pooling layer with a 2 × 2 kernel size and a stride value of 1 was introduced. This particular pooling layer setup is applied consistently after each pair of convolutional layers within the architecture. The third and fourth convolutional layers employed a set of 64 filters each., arranged in a 112 × 112 × 32 format. The shape of the output from the previous layer was then reduced to (56 × 56 × 64) as a result of the max-pooling layer's operation, which used a 2 × 2 kernel size and a stride value of 1. The fifth

TABLE 2. Various parameters in the proposed architecture.

Layers	Type of Layers	Filters	Activation Function	Output Shape	parameters
1	Convolutional 2D	32	ReLU	224,224,32	896
2	Convolutional 2D	32	ReLU	224,224,32	9248
3	Convolutional 2D	64	ReLU	112,112,64	18,496
4	Convolutional 2D	64	ReLU	112,112,64	36,928
5	Convolutional 2D	128	ReLU	56,56,128	73,856
6	Convolutional 2D	128	ReLU	56,56,128	147,584
7	Convolutional 2D	128	ReLU	56,56,128	147,584
8	Convolutional 2D	128	ReLU	56,56,128	147,584
9	Convolutional 2D	256	ReLU	28,28,256	295,168
10	Convolutional 2D	256	ReLU	28,28,256	590,080
11	Convolutional 2D	256	ReLU	28,28,256	590,080
12	Convolutional 2D	512	ReLU	14,14,512	1,180,160
13	Convolutional 2D	512	ReLU	14,14,512	2,359,808
14	Convolutional 2D	512	ReLU	14,14,512	2,359,808
15	Fully Connected Layer 1		ReLU	4096	102,764,544
16	Fully Connected Layer 2		ReLU	1024	4,195,328
17	Classification layer		Softmax	4	4100
	Trainable parameters				114,927,268
	Non-trainable parameters				6016
	Total Parameters				114,933,284

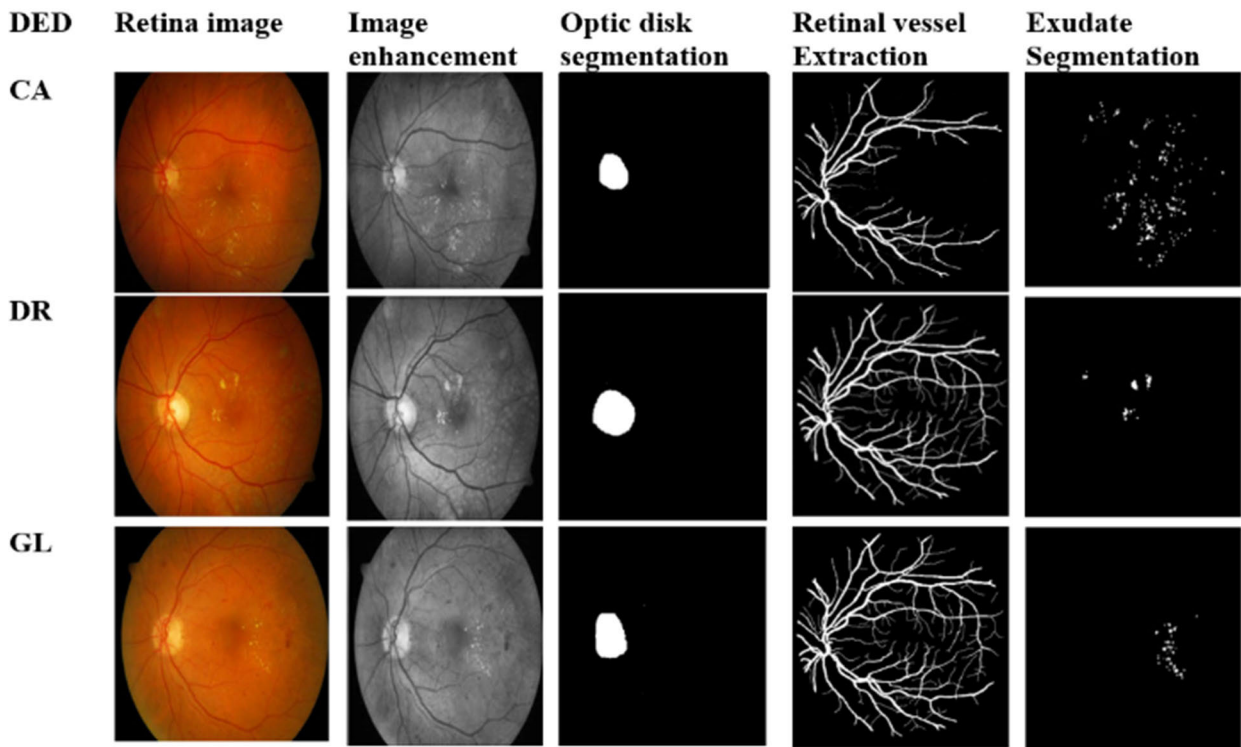**FIGURE 11.** Segmentation Results.

convolutional layer makes use of 128 filters of size 3 × 3. The layer receives a 56 × 56 × 64 input shape, performs a convolutional operation on the feature maps results in an output shape of 56 × 56 × 128. The convolutional layer's output was normalized using batch normalization, which extracted 512 parameters. Having an input shape of 56 × 56 × 128, and then applying batch normalization, sixth, seventh, and the eighth layers are all identical to the fifth. The ninth layer uses 256 filters and a 3 × 3 filter size as input and output shape of the maxpool layer, which has the

---

Algorithm 1 Technique of Batch Normalization Across a Mini-Batch

---

Input: value of **a** over a mini-batch: $\beta = \{a_{1\dots n}\}$; learnable parameters: γ, β

Output: b_i

1: $\mu_\beta \leftarrow \frac{1}{p} \sum_{j=1}^{p} a_j$

2: $\sigma_\beta^2 \leftarrow \frac{1}{p} \sum_{j=1}^{p} (a_j - \mu_\beta)^2$

3: $\hat{a}_j \leftarrow \frac{a_j - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}}$

4: $\gamma_j \leftarrow \gamma \hat{a}_j + \beta \equiv BN_{\mu,\beta}(a_j)$

5: *return* γ_j

---

dimensions of $28 \times 28 \times 128$. After the convolutional layer, 1024 parameters are extracted using batch normalization. With an input shape of $28 \times 28 \times 256$ and then batch normalization, the tenth and eleventh layers are identical to the ninth. The output of the max-pooling layer is a feature map with a shape of $14 \times 14 \times 256$. The input shape of the twelfth layer is $14 \times 14 \times 256$ and 512 filters. After batch normalization, the convolutional layer's output is normalized and 2048 parameters are extracted. The thirteenth and fourteenth layers are identical to the prior layer. Max-pooling serves to decrease the dimensions of the feature map to $7 \times 7 \times 512$ by employing the same window size. Following a DCNN, a flattened layer is introduced to transform the output into a one-dimensional vector. This vector is subsequently employed for classification by means of a fully connected layer. The first and second dense layers use the ReLU activation function. These two dense layers have a dropout layer inserted in between them with a dropout rate of 50%. The output classification layer is built as a densely connected layer, and it consists of four output neurons that reflect the probability of each class during predictions. Table 2 provides comprehensive details of the proposed architecture's configuration and parameters, including the layer-by-layer output structure with weights.

V. EXPERIMENTAL DETAILS
A. HARDWARE AND SOFTWARE
All of the experiments are executed on a Jupyter notebook using Python 3.8, running on hardware with a 2.3 GHz Intel Core i9 processor, 16 GB of 2400 MHz DDR4 RAM, and Intel UHD Graphics 630 with 1536 MB of memory. MatLab was utilized for both the front-end and back-end in the experiments. The data was divided into an 70/10/20 ratio, with 70% used for training, 10% for validation, and the remaining 20% for testing. To ensure a balanced distribution of classes, a generic selection method was employed for data segregation. A mini-batch size of 32 was used, and the categorical cross-entropy loss function was applied. Adam, the default optimizer is used to generate DCNN with 60 epochs. Results were validated using the test dataset's

accuracy, sensitivity, and specificity metrics, which are standard performance assessment metrics.

B. EVALUATION CRITERIA
The most effective DL model has undergone a comprehensive evaluation using various metrics. This evaluation aims to determine the accuracy of classifying DED as either true or false. Initially, we present the confusion matrix in Table 3, obtained through 10-fold cross-validation estimation [60]. This confusion matrix provides predictions for the following outcomes: True Positive (TP): Correct diagnosis with the identification of anomalies, True Negative (TN): Accurate exclusion of periodic instances, False Positives (FP): Instances incorrectly grouped as periodic. The values within the confusion matrix are calculated using the performance metrics outlined below.

1) ACCURACY
Accuracy (Acc) serves as a crucial metric when evaluating the performance of DL classifiers. It is a representation of the correct predictions, encompassing both true positives and true negatives, divided by the total number of elements in the matrix. While a highly accurate model is desirable, it's important to ensure the use of balanced datasets, where false positive and false negative values are approximately equal. To evaluate the effectiveness of the proposed classification model on the DED dataset, we will calculate the elements of the previously mentioned confusion matrix.

$$Acc\,(\%) = \frac{TP + TN}{TP + FN + TN + FP} \quad (25)$$

2) SENSITIVITY
Sensitivity (Sen) is determined by dividing the count of accurate positive predictions by the total count of positive predictions. Sen ranges from 0.0 (lowest) to 1.0 (highest). The following equation is utilized to compute sen:

$$Sen = \frac{TP}{TP + FN} \quad (26)$$

3) SPECIFICITY
Specificity (Spe) is determined by dividing the count of correct negative predictions by the total count of negatives. Spe also ranges from 0.0 (lowest) to 1.0 (highest). The following equation is used to calculate Spe:

$$Spe = \frac{TN}{TN + FP} \quad (27)$$

TABLE 3. Illustration of a confusion matrix.

	Predicted Positive	Predicted Negative	Total
Actual Positive	TP	FN	TP+FN
Actual Negative	FP	TN	FP+TN
Total	Se	Sp	

C. RESULTS

In this study, the performance Acc of three distinct pre-trained DL models, namely ResNet 50, VGG-16, Xception, and EfficientNet B7, was compared and analyzed against the new DCNN model. Large-scale ImageNet data was used to train and evaluate the pre-trained models used in this study. This data includes images of vehicles, animals, flowers, and more. While models are successful in object image categorization, their use is limited to specific domains like medical lesion (DED) detection. Retinal fundus images include a variety of complicated characteristics and lesion localization that influence the prediction of pathological indications. Each CNN layer creates a unique representation of the input image by successively extracting its most salient features. For example, the first layer can learn edges, whereas the last layer can recognize a lesion as a DED classification characteristic. As a consequence, the following conditions were tested: BV, macular areas, and the OD have all been recognized, localized, and segmented as regions of interest.

For each phase of the proposed system, a blend of standard image segmentation methods was employed. All of these algorithms yielded successful segmentation outcomes, demonstrated in Fig. 11, for the specified area of interest. To establish a high-performance system, a series of steps were taken, encompassing the image enhancement, segmentation of BV, OD identification and the extraction, macular region extraction, BV removal, OD elimination, feature extraction, and feature classification. Following segmentation, the image size was optimized to a feasible dimension based on the input specifications of each network. The Image Data Generator class in Keras was used to augment the imbalance dataset in real time, reducing the possibility of model overfitting. Pre-trained models were utilized for fine-tuning after having n layers (CNN layer dependent) discarded and re-trained.

Table 4 and 5 display the conclusive results for each model used for comparison, presenting Acc percentages as the key metric. Among the four fully trained DL models, EfficientNet B7 exhibited superior classification performance, surpassing ResNet 50, VGG-16, and Xception. Similarly, the newly developed CNN model, leveraging pre-processed retinal images, demonstrated exceptional performance, aligning with the proficiency of the other pre-trained models.

Table 4 and 5 display the conclusive results for each model used for comparison, presenting Acc percentages as the key metric. Among the four fully trained DL models, EfficientNet B7 exhibited superior classification performance, surpassing ResNet 50, VGG-16, and Xception. Similarly, the newly developed CNN model, leveraging pre-processed retinal images, demonstrated exceptional performance, aligning with the proficiency of the other pre-trained models.

Ablation Analysis: To explore the effectiveness of the key components in the proposed DCNN structure, an ablation study is conducted and the results are shown in Tables 6 and 7. Initially, the preprocessing components image enhancement filters, and morphological operators are removed from this

TABLE 4. Model's average performance on original images.

DED	Model	Acc%	Sen%	Spe%	Prec%
CA	ResNet50	60.87	67	58	63
	VGG 16	65.87	72	63	68
	Xception	61.67	76.43	58.85	86
	EfficientNetB7	85.43	81.92	90	79
DR	ResNet50	62.07	65	60	62
	VGG 16	67.07	70	65	67
	Xception	56.72	85	56	45
	EfficientNetB7	82.79	85	86	82
GL	ResNet50	63.41	85.71	58.82	80
	VGG 16	68.41	90.71	62.82	85
	Xception	85.49	82	89	81
	EfficientNetB7	90.80	99.12	88.33	90
NORMAL	ResNet50	63.07	66	61	63
	VGG 16	68.07	71	66	68
	Xception	61.67	76.43	68.85	80
	EfficientNetB7	86.49	82	89	81

TABLE 5. The efficientNetB7 Model's average performance on pre-processed images.

DED	Model	Acc%	Sen%	Spe%	Prec%
CA	EfficientNetB7	94.13	90	96	97
DR	EfficientNetB7	88.43	91	90.71	83
GL	EfficientNetB7	93	96	95	95
NORMAL	EfficientNetB7	90	91	93	90.71

TABLE 6. The DCNN model's average performance on original images.

DED	Model	Acc%	Sen%	Spe%	Prec%
CA	DCNN	68.33	58.33	78.33	81
DR	DCNN	87.36	88.33	87.35	86
GL	DCNN	92.77	93	86	88
NORMAL	DCNN	91	88	92.75	89.13

TABLE 7. DCNN model's average performance on pre-processed images.

DED	Model	Acc%	Sen%	Spe%	Prec%
CA	DCNN	96.43	99.46	93.24	99
DR	DCNN	98.33	99.32	91.67	94
GL	DCNN	97	98	88.24	94
NORMAL	DCNN	96	93	94.65	91.71

framework and train the model with original retinal input images. As observations from Tables 6 and 7, the average performance on original images like Acc, Prec, Sen, Spe is downgraded than the performance of the pre-processed retinal images.
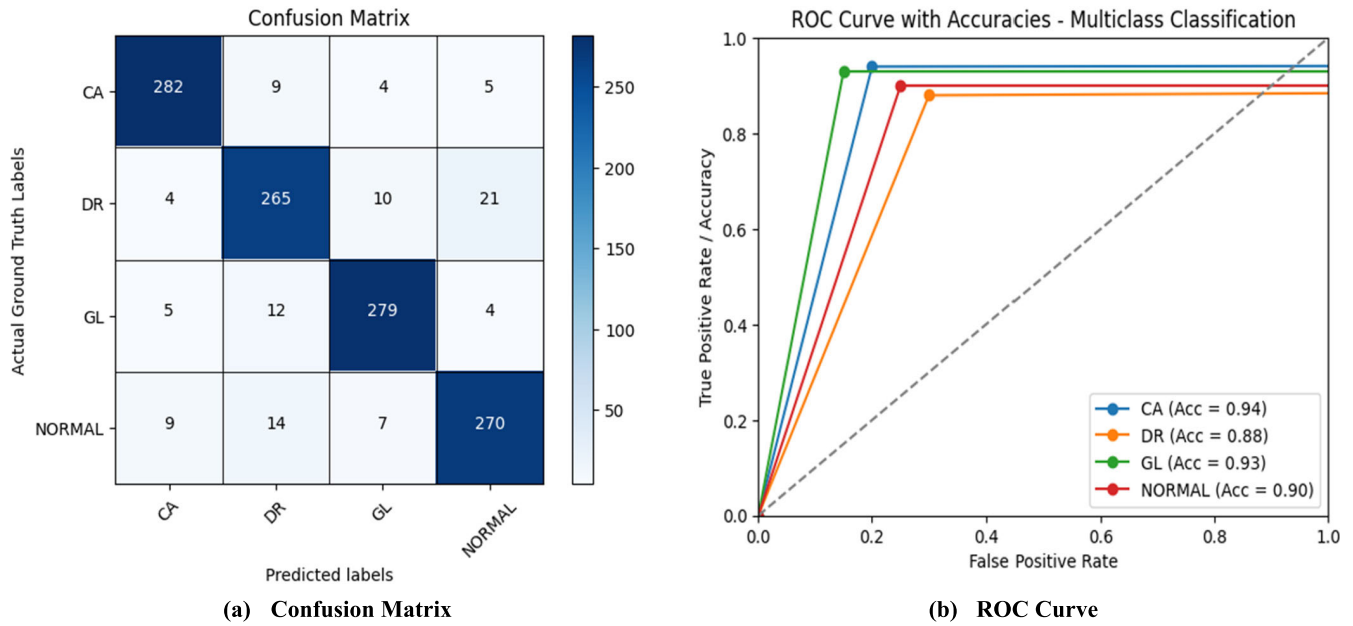
(a) Confusion Matrix

(b) ROC Curve

FIGURE 12. EfficientNet B7 model performance.

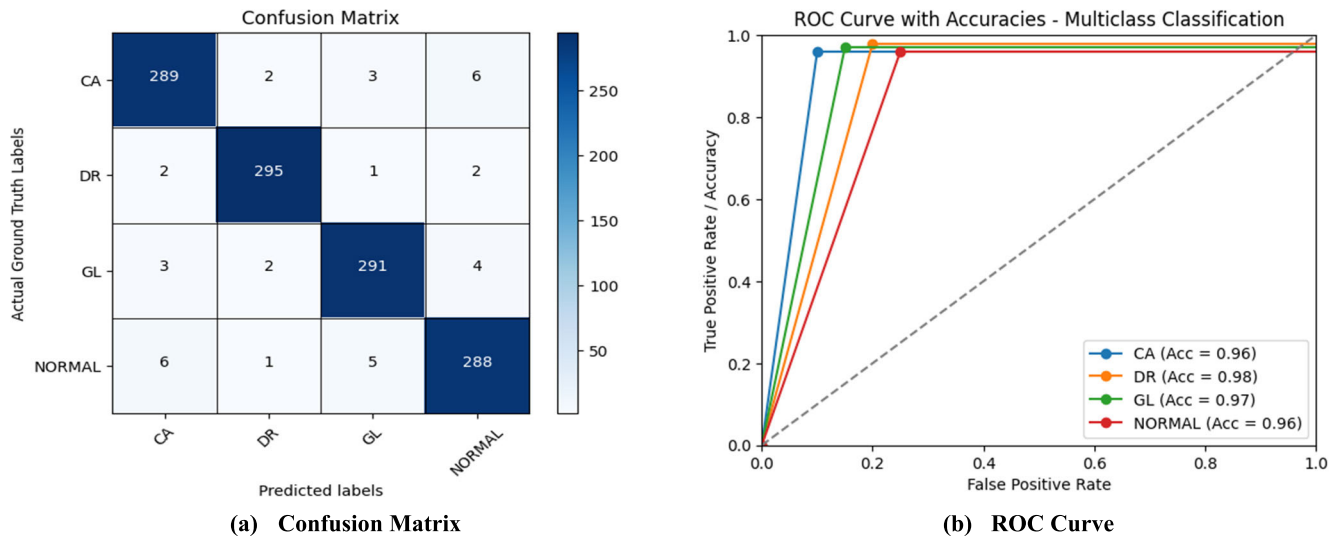**(a) Confusion Matrix**

(b) ROC Curve

FIGURE 13. Proposed DCNN model performance.

For example, the performance of the model in classifying the DED's class CA is downgraded by Acc 28.1%, Sen 41.13%, Spe 14.71%, and Prec 18% than pre-processed retinal fundus images.

It shows the importance of the pre-processing in the DCNN framework. For the multi-class classification of healthy and various DED statuses, the ROC curves and confusion matrices of EfficientNetB7, best performed pre-trained DL model and a built DCNN model are depicted in Fig. 12 and Fig. 13.

D. DISCUSSION

This research investigates the application of multi-class classification using DL techniques to automatically detect

three distinct DEDs. The findings of this study highlight that the intricacy of DL algorithms is primarily affected by the quality and the quantity of available data, specifically ocular fundus images, rather than the inherent method itself. In this study, publicly accessible annotated fundus image data were utilized for experimentation. It is worth noting that labeled hospital fundus images could potentially yield more robust, practical, and realistic results for computer-aided clinical applications. CA, DR, and GL are three of the most common retinal disorders associated with diabetes. Without timely assessment and intervention, these conditions have the potential to cause significant and irreversible visual impairment [1], [2]. Increasing life expectancy, busy lifestyles, and various other variables all point to a rise in

the number of diabetics [1]. Early detection of abnormal symptoms reduces the future progression of the disease, its impact on affected persons, and associated medical expenses. Consequently, the DED identification system has the potential to fully automate or partially automate the eye-screening process. The first approach necessitates a high level of Acc, similar to that of retinal specialists. In line with the guidelines of the British Diabetic Association (BDA), the chosen approach must meet the lowest threshold of 95% Spe, and 80% Sen for the detection of vision-threatening DR method. It condenses the results of massive screening efforts to identify possible DED instances for further study in humans. Both of these alternatives substantially reduce the need for trained ophthalmologists and specialist facilities, opening up the procedure to a much larger population and making it more feasible in areas with limited resources. Additionally, addressing early categorization issues remains a key clinical concern.

Risk Analysis: As observed, DED progression is a risk factor with a long history of diabetes, age over 40, anemia, obesity, and other risk factors. Patients with DED should be screened at least once a year. However, the DED patient's history is required to analyze the disease's progression. This DED is classified into four classes according to its severity. The DCNN-based framework provides a solution but it needs a large dataset to develop the optimal model which includes numerous training parameters. The collection of numerous labeled diabetic retinal fundus images is a challenging task. Because it needs many ophthalmologists to annotate numerous ground truth images. The proposed framework utilizes augmentation methods to deal with the inadequacy of the retinal images. The identification of lines, curves, orientation, and textures in ground truth retinal images is a challenging task. On the other hand, feature maps are generated by convolution layers through the extraction of those features more accurately than handcrafted features. The DCNN fully-connected layer learns the pattern of the features and classifies the disease in the output layer.

Previous studies predominantly centered on binary classification for predicting diabetic eye diseases. It's worth noting that even though Google has developed a DL model that surpasses the performance of ophthalmologists, their 'Inceptionv3' model was specifically optimized for binary classification in the context of DR identification [14]. Due to the very minor signs of the impairment, multi-class DED is sometimes very difficult to distinguish from a normal retina, therefore an enhancement in the data quality was anticipated. To make abnormal features more visible, CNN architecture's top layer was removed and retrained EfficientNetB7, which produced Acc values of 94.13%, 88.43%, 93%, and 90% for each (Table 5). Xception and ResNet50 achieved the lowest performance. The effect of the fine-tuning differed throughout the models. The observed Acc pick-up was minimal, confirming the suitability of networks that are pre-trained by default for DED classification tasks. In simpler terms, even though these CNN networks underwent training

on a diverse range of images from the ImageNet library, they exhibited the capability to differentiate between multi-class DED and a healthy retina. Unfreezing is not recommended if it does not increase in Acc, since this would waste computing resources and time. Comparisons of the study's results with existing works reveal notable achievements. The DCNN model outperforms other models in DR detection with an Acc rate of 98.33% compared to the existing work's range between 93% and 96% [14]. Additionally, for CA detection, the DCNN model achieved an Acc of 96.43%, aligning closely with existing studies reporting rates ranging from 92.91% [29] to 95.78% [30]. The DCNN model also demonstrates superior performance in GL detection with an accuracy of 97%, surpassing existing studies reporting accuracies of 91.5%, 92.4%, and 93.8% [43]. The Acc of the proposed DCNN model was 96.43 %, 98.33 %, 97%, and 96% respectively. The performance of the utilized models was compared using two scenarios: (1) before image preprocessing and (2) after image preprocessing. To address overfitting, models underwent training on a raw dataset without preprocessing, including data augmentation involving geometric transformations applied to the Messidor, Messidor-2, and DRISTI-GS datasets. Post-image preprocessing, the datasets were subjected to various conventional image processing techniques, resulting in an enhanced classification performance of 98.33% (the greatest Acc obtained for DR). After evaluating the high-performing technique on the CA, DR, GL, and NORMAL detection tasks, maximum Sen of 99.46%, 99.32%, 98%, and 93% were achieved, along with maximum Spe of 93.24%, 91.67%, 88.24%, and 94.65%. Therefore, early DED detection met the BDA requirements sufficiently, although Spe was deficient by 9% and 6%.

VI. CONCLUSION

This work presents a method for identifying multi-class DED, which has not been thoroughly described in earlier research. A number of DL performance optimization strategies have been used, including image enhancement methods, like extracting the green channel, CLAHE, and illumination correction, were applied. Subsequently, image segmentation methods such as the Tyler Coye Algorithm, Otsu thresholding, and Circular Hough Transform are applied to extract the essential ROI's such as extraction of features like BV, the macular region, and the optic nerve from the raw ocular fundus images. After preprocessing, these images are trained using EfficientNetB7 model that outperformed among the four pre-trained models ResNet50, VGG-16, Xception, and EfficientNetB7 and the proposed DCNN model. The proposed DCNN methodology holds promising results for the CA, DR, GL, and NORMAL detection tasks, achieving accuracies of 96.43%, 98.33%, 97%, and 96%, respectively. Automatic identification capabilities that are highly selective across categories are another advantage of DL. This approach helps overcome the technical constraints linked to the analytical and frequently subjective process of manual feature extraction. Moreover, the study incorporated

comprehensive datasets from various origins to assess the system's robustness and its capacity to handle real-world scenarios. The proposed model streamlines labor-intensive eye-screening procedures and acts as a supplementary diagnostic tool, minimizing human subjectivity.

ACKNOWLEDGMENT

The authors would like to thank the editors and the reviewers.

REFERENCES

[1] M. Vadduri and P. Kuppusamy, "Diabetic eye diseases detection and classification using deep learning techniques—A survey," in *Information and Communication Technology for Competitive Strategies (ICTCS)* (Lecture Notes in Networks and Systems), vol. 623, A. Joshi, M. Mahmud, and R. G. Ragel, Eds. Singapore: Springer, 2023, pp. 443–454, doi: 10.1007/978-981-19-9638-2_38.

[2] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, J. Ma, and K. Wang, "Image preprocessing in classification and identification of diabetic eye diseases," *Data Sci. Eng.*, vol. 6, no. 4, pp. 455–471, Dec. 2021, doi: 10.1007/s41019-021-00167-z.

[3] M. D. Abramoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer, and G. Quellec, "Automated early detection of diabetic retinopathy," *Ophthalmology*, vol. 117, no. 6, pp. 1147–1154, Jun. 2010, doi: 10.1016/j.ophtha.2010.03.046.

[4] P. Kuppusamy, M. M. Basha, and C.-L. Hung, "Retinal blood vessel segmentation using random forest with Gabor and Canny edge features," in *Proc. Int. Conf. Smart Technol. Syst. Next Gener. Comput. (ICSTSN)*, Villupuram, India, Mar. 2022, pp. 1–4, doi: 10.1109/ICSTSN53084.2022.9761339.

[5] P. Kuppusamy and C.-L. Hung, "Enriching the multi-object detection using convolutional neural network in macro-image," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Coimbatore, India, Jan. 2021, pp. 1–5, doi: 10.1109/ICCCI50826.2021.9402565.

[6] H. Li, Y. Wang, H. Wang, and B. Zhou, "Multi-window based ensemble learning for classification of imbalanced streaming data," *World Wide Web*, vol. 20, no. 6, pp. 1507–1525, Nov. 2017, doi: 10.1007/s11280-017-0449-x.

[7] C. Lam, D. Yi, M. Guo, and T. Lindsey, "Automated detection of diabetic retinopathy using deep learning," *AMIA Joint Summits Transl. Sci. Proc.*, vol. 2017, pp. 147–155, Jan. 2018.

[8] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automatic detection of diabetic eye disease through deep learning using fundus images: A survey," *IEEE Access*, vol. 8, pp. 151133–151149, 2020, doi: 10.1109/ACCESS.2020.3015258.

[9] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Y. K. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2136–2155, Dec. 2013, doi: 10.1016/j.compbiomed.2013.10.007.

[10] M. Kaur and M. Kaur, "A hybrid approach for automatic exudates detection in eye fundus image," *Int. J.*, vol. 5, no. 6, pp. 411–417, 2015.

[11] A. Gowda Karegowda, A. Nasiha, M. A. Jayaram, and A. S. Manjunath, "Exudates detection in retinal images using back propagation neural network," *Int. J. Comput. Appl.*, vol. 25, no. 3, pp. 25–31, Jul. 2011, doi: 10.5120/3011-4062.

[12] A. Sopharak and B. Uyyanonvara, "Automatic exudates detection from diabetic retinopathy retinal image using fuzzy C-means and morphological methods," in *Proc. Conf. Adv. Comput. Sci. Technol.*, 2007, pp. 359–364.

[13] M. Juneja, S. Singh, N. Agarwal, S. Bali, S. Gupta, N. Thakur, and P. Jindal, "Automated detection of glaucoma using deep learning convolution network (G-Net)," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 15531–15553, Jun. 2020, doi: 10.1007/s11042-019-7460-4.

[14] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, p. 2402, Dec. 2016, doi: 10.1001/jama.2016.17216.

[15] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, Jul. 2017, doi: 10.1016/j.ophtha.2017.02.008.

[16] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Med. Imag.*, vol. 8, no. 3, pp. 263–269, Sep. 1989, doi: 10.1109/42.34715.

[17] D. Vallabha, R. Dorairaj, K. Namuduri, and H. Thompson, "Automated detection and classification of vascular abnormalities in diabetic retinopathy," in *Proc. 38th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2004, pp. 1625–1629, doi: 10.1109/acssc.2004.1399432.

[18] K. Noronha, J. Nayak, and S. N. Bhat, "Enhancement of retinal fundus image to highlight the features for detection of abnormal eyes," in *Proc. IEEE Region Conf.*, Hong Kong, 2006, pp. 1–4, doi: 10.1109/ten-con.2006.343793.

[19] G. G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool," *Brit. J. Ophthalmol.*, vol. 80, no. 11, pp. 940–944, Nov. 1996, doi: 10.1136/bjo.80.11.940.

[20] J. C. Bezdek, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Berlin, Germany: Springer, 1999.

[21] R. Markham, A. Osareh, M. Mirmehdi, B. Thomas, and M. Macipe, "Automated identification of diabetic retinal exudates using support vector machines and neural networks," *Investigative Ophthalmol. Vis. Sci.*, vol. 44, no. 13, p. 4000, 2003.

[22] U. R. Acharya, C. M. Lim, E. Y. K. Ng, C. Chee, and T. Tamura, "Computer-based detection of diabetes retinopathy stages using digital fundus images," *Proc. Inst. Mech. Eng., H, J. Eng. Med.*, vol. 223, no. 5, pp. 545–553, Jul. 2009, doi: 10.1243/09544119JEIM486.

[23] E. Uchino, K. Suzuki, N. Sato, R. Kojima, Y. Tamada, S. Hiragi, H. Yokoi, N. Yugami, S. Minamiguchi, H. Haga, M. Yanagita, and Y. Okuno, "Classification of glomerular pathological findings using deep learning and nephrologist-AI collective intelligence approach," *Int. J. Med. Informat.*, vol. 141, Sep. 2020, Art. no. 104231, doi: 10.1016/j.ijmedinf.2020.104231.

[24] M. S. Junayed, A. A. Jeny, S. T. Atik, N. Neehal, A. Karim, S. Azam, and B. Shanmugam, "AcneNet—A deep CNN based classification approach for acne classes," in *Proc. 12th Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, Surabaya, Indonesia, Jul. 2019, pp. 203–208, doi: 10.1109/ICTS.2019.8850935.

[25] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2693–2701, Nov. 2015, doi: 10.1109/TBME.2015.2444389.

[26] M. A. Syarifah, A. Bustamam, and P. P. Tampubolon, "Cataract classification based on fundus image using an optimized convolution neural network with lookahead optimizer," in *Proc. Int. Conf. Sci. Appl. Sci. (ICSAS)*, 2020, Art. no. 020034, doi: 10.1063/5.0030744.

[27] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "Cataract-Net: An automated cataract detection system using deep learning for fundus images," *IEEE Access*, vol. 9, pp. 128799–128808, 2021, doi: 10.1109/ACCESS.2021.3112938.

[28] T. Pratap and P. Kokil, "Computer-aided diagnosis of cataract using deep transfer learning," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101533, doi: 10.1016/j.bspc.2019.04.010.

[29] D. Kim, T. J. Jun, Y. Eom, C. Kim, and D. Kim, "Tournament based ranking CNN for the cataract grading," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 1630–1636, doi: 10.1109/EMBC.2019.8856636.

[30] Md. R. Hossain, S. Afroze, N. Siddique, and M. M. Hoque, "Automatic detection of eye cataract using deep convolution neural networks (DCNNs)," in *Proc. IEEE Region Symp. (TENSYMP)*, Dhaka, Bangladesh, Jun. 2020, pp. 1333–1338, doi: 10.1109/TENSYMP50017.2020.9231045.

[31] X. Zhang, J. Lv, H. Zheng, and Y. Sang, "Attention-based multi-model ensemble for automatic cataract detection in B-scan eye ultrasound images," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–10, doi: 10.1109/IJCNN48605.2020.9207696.

[32] Md. S. Mahmud Khan, M. Ahmed, R. Z. Rasel, and M. M. Khan, "Cataract detection using convolutional neural network with VGG-19 model," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, Seattle, WA, USA, May 2021, pp. 0209–0212, doi: 10.1109/AIIoT52608.2021.9454244.

[33] Kaggle. (2021). *Ocular Disease Recognition*. [Online]. Available: https://www.kaggle.com/andrewmvd/ocular-diseaserecognitionodir5k

[34] J. Amin, M. Sharif, A. Rehman, M. Raza, and M. R. Mufti, "Diabetic retinopathy detection and classification using hybrid feature set," *Microsc. Res. Technique*, vol. 81, no. 9, pp. 990–996, Sep. 2018, doi: 10.1002/jemt.23063.

[35] S. Patel, "Diabetic retinopathy detection and classification using pretrained convolutional neural networks," *Int. J. Emerg. Technol.*, vol. 11, no. 3, pp. 1082–1087, 2020.

[36] S. Das, K. Kharbanda, M. Suchetha, R. Raman, and E. Dhas, "Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102600, doi: 10.1016/j.bspc.2021.102600.

[37] L. Math and R. Fatima, "Adaptive machine learning classification for diabetic retinopathy," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 5173–5186, Feb. 2021, doi: 10.1007/s11042-020-09793-7.

[38] A. H. A. Samah, F. Ahmad, M. K. Osman, N. M. Tahir, M. Idris, and N. A. A. Aziz, "Diabetic retinopathy pathological signs detection using image enhancement technique and deep learning," *J. Electr. Electron. Syst. Res.*, vol. 18, pp. 44–52, Apr. 2021.

[39] S.-Y. Lu, S.-H. Wang, X. Zhang, and Y.-D. Zhang, "TBNet: A context-aware graph network for tuberculosis diagnosis," *Comput. Methods Programs Biomed.*, vol. 214, Feb. 2022, Art. no. 106587, doi: 10.1016/j.cmpb.2021.106587.

[40] S. Lu, S.-H. Wang, and Y.-D. Zhang, "Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm," *Neural Comput. Appl.*, vol. 33, no. 17, pp. 10799–10811, Sep. 2021, doi: 10.1007/s00521-020-05082-4.

[41] A. S. Jadhav, P. B. Patil, and S. Biradar, "Analysis on diagnosing diabetic retinopathy by segmenting blood vessels, optic disc and retinal abnormalities," *J. Med. Eng. Technol.*, vol. 44, no. 6, pp. 299–316, Aug. 2020, doi: 10.1080/03091902.2020.1791986.

[42] J. Civit-Masot, M. J. Domínguez-Morales, S. Vicente-Díaz, and A. Civit, "Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction," *IEEE Access*, vol. 8, pp. 127519–127529, 2020, doi: 10.1109/ACCESS.2020.3008539.

[43] M. Al Ghamdi, M. Li, M. Abdel-Mottaleb, and M. A. El-Ghar, "Multi-task deep learning for glaucoma detection and classification," *Comput. Biol. Med.*, vol. 137, Jul. 2021, Art. no. 104750.

[44] J. J. Gómez-Valverde, A. Antón, G. Fatti, B. Liefers, A. Herranz, A. Santos, C. I. Sánchez, and M. J. Ledesma-Carbayo, "Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning," *Biomed. Opt. Exp.*, vol. 10, no. 2, pp. 892–913, 2019, doi: 10.1364/boe.10.000892.

[45] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J. M. Mossi, and A. Navea, "CNNs for automatic glaucoma assessment using fundus images: An extensive validation," *Biomed. Eng. OnLine*, vol. 18, no. 1, pp. 1–19, Dec. 2019, doi: 10.1186/s12938-019-0649-y.

[46] B. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A Practical Approach with Examples in MATLAB*. Hoboken, NJ, USA: Wiley, 2011.

[47] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*, P. S. Heckbert, Ed., 1st ed. San Diego, CA, USA: Academic Press, 1994, ch. 8.5, pp. 474–485.

[48] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst., Man Cybern., B*, vol. 28, no. 3, pp. 301–315, Jun. 1998, doi: 10.1109/3477.678624.

[49] T. J. Jun, Y. Eom, D. Kim, C. Kim, J.-H. Park, H. M. Nguyen, Y.-H. Kim, and D. Kim, "TRk-CNN: Transferable ranking-CNN for image classification of glaucoma, glaucoma suspect, and normal eyes," *Exp. Syst. Appl.*, vol. 182, Nov. 2021, Art. no. 115211, doi: 10.1016/j.eswa.2021.115211.

[50] Md. S. B. Mostafa, D. Bal, K. A. Sathi, and Md. A. Hossain, "Diagnosis of glaucoma from retinal fundus image using deep transfer learning," in *Proc. 1st Int. Conf. Artif. Intell. Trends Pattern Recognit. (ICAITPR)*, Hyderabad, India, Mar. 2022, pp. 1–4, doi: 10.1109/ICAITPR51569.2022.9844194.

[51] H. Wu, J. Lv, and J. Wang, "Automatic cataract detection with multi-task learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Shenzhen, China, Jul. 2021, pp. 1–8, doi: 10.1109/IJCNN52387.2021.9533424.

[52] T. Mahmud and M. Monirujjaman Khan, "A medical app based automated disease predicting doctor," in *Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Erode, India, Apr. 2021, pp. 478–485, doi: 10.1109/ICCMC51019.2021.9418435.

[53] Kaggle. (2015). *Diabetic Retinopathy Detection*. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection

[54] J. Amin, M. Sharif, M. Yasmin, H. Ali, and S. L. Fernandes, "A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions," *J. Comput. Sci.*, vol. 19, pp. 153–164, Mar. 2017, doi: 10.1016/j.jocs.2017.01.002.

[55] S. Patel et al., "Predicting a risk of diabetes at early stage using machine learning approach," *Turkish J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 10, pp. 5277–5284, Apr. 2021.

[56] S. Lekha and M. Suchetha, "Recent advancements and future prospects on e-nose sensors technology and machine learning approaches for non-invasive diabetes diagnosis: A review," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 127–138, 2021, doi: 10.1109/RBME.2020.2993591.

[57] L. Math and R. Fatima, "Identification of diabetic retinopathy from fundus images using CNNs," *Int. J. Innov. Technol. Exploring Eng.*, vol. 9, no. 1, pp. 3439–3443, Nov. 2019.

[58] A. H. A. Samah, F. Ahmad, M. K. Osman, M. Idris, N. M. Tahir, and N. A. A. Aziz, "Classification of pathological signs for diabetic retinopathy diagnosis using image enhancement technique and convolution neural network," in *Proc. 9th IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Penang, Malaysia, Nov. 2019, pp. 221–225, doi: 10.1109/ICCSCE47578.2019.9068538.

[59] J. Wang, S. Lu, S.-H. Wang, and Y.-D. Zhang, "A review on extreme learning machine," *Multimedia Tools Appl.*, vol. 81, no. 29, pp. 41611–41660, Dec. 2022, doi: 10.1007/s11042-021-11007-7.

[60] S. Wang, M. E. Celebi, Y.-D. Zhang, X. Yu, S. Lu, X. Yao, Q. Zhou, M.-G. Miguel, Y. Tian, J. M. Gorriz, and I. Tyukin, "Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects," *Inf. Fusion*, vol. 76, pp. 376–421, Dec. 2021, doi: 10.1016/j.inffus.2021.07.001.

[61] A. S. Jadhav, P. B. Patil, and S. Biradar, "Optimal feature selection-based diabetic retinopathy detection using improved rider optimization algorithm enabled with deep learning," *Evol. Intell.*, vol. 14, no. 4, pp. 1431–1448, Dec. 2021, doi: 10.1007/s12065-020-00400-0.

[62] J. Civit-Masot, F. Luna-Perejón, S. Vicente-Díaz, J. M. R. Corral, and A. Civit, "TPU cloud-based generalized U-Net for eye fundus image segmentation," *IEEE Access*, vol. 7, pp. 142379–142387, 2019, doi: 10.1109/ACCESS.2019.2944692.
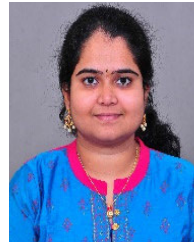
MANEESHA VADDURI received the bachelor's and master's degrees in computer science and engineering from JNTUK, Andhra Pradesh, India, in 2017 and 2019, respectively. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, VIT-AP University, Andhra Pradesh. Her research interests include deep learning, machine learning, and computer vision.

P. KUPPUSAMY (Member, IEEE) received the B.E. degree in computer science and engineering from Madras University, India, in 2002, and the master's degree in CSE and the Ph.D. degree in information and communication engineering from Anna University, Chennai, in 2007 and 2014, respectively. He is currently a Professor with the School of Computer Science & Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India. He has published more than 50 research papers in leading international journals, international conferences, and few books. He is also working on machine learning, deep learning, and the Internet of Things-based research project with sensors, raspberry pi, and Arduino for handling smart devices. He is a member of ISTE, IAENG, and IACSIT. He serves as a reviewer for various journals and international conferences.

● ● ●