Healthcare Analytics 5 (2024) 100303

Contents lists available at ScienceDirect

Healthcare Analytics

journal homepage: www.elsevier.com/locate/health

An advanced deep neural network for fundus image analysis and enhancing diabetic retinopathy detection

F M Javed Mehedi Shamrat [a], Rashiduzzaman Shakil [b], Sharmin [a], Nazmul Hoque ovy [c], Bonna Akter [b], Md Zunayed Ahmed [d], Kawsar Ahmed [e,f,g,*], Francis M. Bui [e], Mohammad Ali Moni [h,**]

[a] Department of Computer System and Technology, University of Malaya, Kuala Lumpur, 50603, Malaysia
[b] Department of Computer Science and Engineering, Daffodil International University, Dhaka, 1216, Bangladesh
[c] Department of Software Engineering, Daffodil International University, Dhaka, 1216, Bangladesh
[d] Department of Computer Science and Engineering, European University of Bangladesh, Dhaka, 1216, Bangladesh
[e] Department of Electrical and Computer Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, S7N 5A9, Canada
[f] Group of Biophotomatiχ, Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail, 1902, Bangladesh
[g] Health Informatics Research Lab, Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City (DSC), Birulia, Savar, Dhaka, 1216, Bangladesh
[h] School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland, St Lucia, QLD, 4072, Australia

ARTICLE INFO

Handling editor: Madijd Tavana

Keywords:
Convolutional neural network
Diabetic retinopathy network
Diagnostic analytics
Classification
Feature map

ABSTRACT

Diabetic retinopathy (DR) involves retina damage due to diabetes, often leading to blindness. It is diagnosed via color fundus injections, but the manual analysis is cumbersome and error-prone. While computer vision techniques can predict DR stages, they are computationally intensive and struggle with complex data extraction. In this research, our prime objective was to automate the process of DR classification into its various stages using convolutional neural network (CNN) models. We employed the performance of fifteen pre-trained models with our novel proposed diabetic retinopathy network (DRNet13) model. We aimed to discern the most efficient model for accurate diabetic retinopathy (DR) staging based on fundus images from five DR classes. We preprocessed the image using a median filter for noise reduction and Gamma correction for image enhancement. We expanded our dataset from 3662 to 7500 images to create a more generalized training model through various augmentation techniques. We also evaluated multiple evaluation metrics, including accuracy, precision, F1-score, Sensitivity, Specificity, Area under the curve (AUC), Mean Squared Error (MSE), False Positive Rate (FPR), False Negative Rate (FNR), in addition to confusion matrices for an in-depth comparison of the performance of these models. Feature maps were employed to illuminate decision making areas in the DRNet13 model, which achieved a 97 % accuracy rate for DR detection, surpassing other CNN architectures in speed and efficiency. Despite a few misclassifications, the model's capability to identify critical features demonstrates its potential as an impactful diagnostic tool for timely and accurate identification of diabetic retinopathy.

1. Introduction

Over the past few decades, diabetes has become one of the most rapidly expanding disease burdens. Currently, 382 million individuals throughout the world suffer from diabetes mellitus (DM), and this number is expected to rise to 592 million by 2025 [1]. Type-I and Type-II DM are the two disease subtypes, respectively, and are distinguished by their respective causes and symptoms. Moreover, the human eye is one of the vital body parts negatively impacted by both forms of DM. Diabetic retinopathy (DR) [2–4] describes a serious eye disease linked to

* Corresponding author. Department of Electrical and Computer Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, S7N 5A9, Canada.
** Corresponding author.
E-mail addresses: javedmehedicom@gmail.com (F.M.J. Mehedi Shamrat), rashiduzzaman.diucse@gmail.com (R. Shakil), sharmin9889@gmail.com (Sharmin), ovy.bjit@gmail.com (N. Hoque ovy), bonna.diucse@gmail.com (B. Akter), zunayedahmed01@gmail.com (M.Z. Ahmed), kawsar.ict@mbstu.ac.bd (K. Ahmed), francis.bui@usask.ca (F.M. Bui), m.moni@uq.edu.au (M.A. Moni).

https://doi.org/10.1016/j.health.2024.100303
Received 6 August 2023; Received in revised form 18 December 2023; Accepted 22 December 2023
Available online 20 January 2024

DM. The symptoms of DR demonstrate that it results in the mutilation of blood vessels in the retina. There have been reports that DR affects 34.6 % of the world's population, which amounts to 382 million people.

To address the aforementioned challenges and the rising frequency of DR, our research demonstrates a mixed-method strategy that leverages the analytical skills of both qualitative evaluations and quantitative [5] deep learning approaches. This method is intended to improve diagnosis accuracy and provide an improved understanding of disease development. In addition to diabetic retinopathy, 7.0 % of the population has proliferative diabetic retinopathy (PDR), and 6.8 % have diabetic macular edema (DME) [6]. When the worldwide scenario is extrapolated from these statistics, the number of DR cases is expected to rise from 126.6 million to 191.1 million by 2030 [3]. Such statistics highlight the critical need for enhanced diagnostic approaches capable of reliably detecting and classifying DR at an early stage in order to prevent permanent vision loss.

Furthermore, the present DR diagnostic landscape is characterized by a wide variety of techniques, each with intrinsic strengths and limitations. Our study aims to critically examine these current approaches, emphasizing the necessity for a complete and automated detection system capable of resolving the inadequacies of manual feature extraction as well as the varied performance of pre-trained models.

On the other hand, reports have demonstrated that DR is responsible for the blindness of 4 % of the total population in the Khyber Pakhtunkhwa (KP) province of Pakistan. In addition, it was found that the most prevalent cause of DR in the region was BD type-I DR (Risk Factors Awareness and Presentation, Pakistan, 2017). Diabetes mellitus was reported by 30 % of the population in KP, with type II diabetes affecting 1.6 % of the patients. In KP, approximately 1.6 % of the overall population has type II diabetes, and it was observed that almost 30 % of diabetic patients were DR patients. Two percent of these individuals have claimed they are very close to being completely blind [1]. Diagnosis can be challenging in the early stages of DR, and patients may report feeling symptom-free. However, Patients in later phases experience decreased visual acuity and symptoms such as blurriness, floaters, and distortions. Although NPDR is the first stage of DR, it is further subdivided into three types.

Microaneurysms, on the other hand, are merely spherical, locally dilated capillaries inside the eye [7]. The severity of NDPR is classified as "mild," "moderate," or "severe." During the early stages of DR, the blood capillary end forms a little, red, spherical spot [8,9]. Flame-shaped haemorrhages can occur in mild instances with more than five MA. When there are more than 20 haemorrhages within the retina, NPDR has progressed to the final stage. The creation of new blood vessels in response to these damage is known as neovascularization, and it affects the whole inner surface of the retina. Early diagnosis of DR can help prevent the worsening of NPDR into PDR. One hundred thirty patients were monitored for DR signs. Two-thirds of a sample size of 130 patients tested positive for DR. Of these DR-positive individuals, 25.8 % had advanced to the PDR stage [8], the terminal DR stage [6]. Because the early signs of this disorder are disguised, it typically stays unnoticed for a long period before causing irreparable vision loss. As a result, detecting DR early is crucial to preventing the complications associated with the condition [10]. To make progress in comprehending the disease and its prognosis, highly qualified medical professionals and specialists armed with cutting-edge diagnostic tools and methodologies are required.

As a result, a reliable automated detection system is required to identify and prevent the impact of DR [11]. Prior work in the DR focused mainly on feature extraction using machine learning approaches [12–15]. However, difficulties with manual feature extraction directed scientists towards deep learning. Data mining, image processing, machine learning, and deep learning are just some of the computer-aided technologies that have benefited from increased medical research [16–18]. However, in recent years, Deep Learning has gained popularity in a variety of disciplines, including sentiment analysis, handwriting recognition, stock market prediction, medical image analysis, and so on. When it comes to image classification, CNN in deep learning typically provides effective outcomes.

In this study, we employ a mixed-method strategy to improve the identification and classification of DR, combining both qualitative insights from medical research and the quantitative capability of deep learning algorithms. This sophisticated approach fills methodological gaps in prior research, which have been limited by excessive dependence on human feature extraction and inadequately adaptable machine learning algorithms. We introduce the DRNet13 model as a model for advancement in medical image analysis by carefully connecting our research with the critical need for early and accurate identification of DR. Our findings have the potential to make a significant contribution to medical diagnostics by transforming the prognosis and treatment of diabetic retinopathy. We have chosen the DRNet13 with purpose and strategy. Unlike other approaches which may provide countable accuracy at the expense of higher processing demand, DRNet13 achieves an optimal balance between accuracy and computational efficiency. This is crucial for practical applications where resources are limited. The decision is straightened further by its unique architectural design, which facilitated complex data extraction, a major challenge in the automated analysis of fundus images associated with DR. The main subject of this research is evaluating the severity of DR fundus images. Fig. 1 illustrates the entire overview of this study. The contributions of this research are as follows:

1. A novel model DRNet13 has been introduced for Diabetic Retinopathy (DR) classification with a validation accuracy of 97 %. It surpassed fifteen pre-trained models in performance.
2. A comprehensive comparative study, including fifteen pre-trained models and DRNet13, has been presented to identify the most efficient CNN architecture for DR classification.
3. An efficient pre-processing pipeline was developed, incorporating Median filtering for noise reduction and Gamma correction for image enhancement.
4. Image augmentation techniques have expanded the dataset from 3662 to 7500 images, enhancing the model's capacity to handle real-world variability.
5. Performance assessment across multiple metrics has provided a holistic view of the model's capabilities, including an in-depth analysis of the misclassification results from the proposed model DRNet13.
6. Feature maps have been used to examine the model's decision-making processes, identifying the ROI and features within the fundus images that the models focused on during the classification task.
7. Among all the models tested, DRNet13 ranked lowest in computational time that proving its efficiency.

2. Literature review

In recent years, the potential deep learning algorithms for detecting and classifying DR have been investigated extensively. Various researchers have made significant strides in automating DR detection process, improving accuracy, and reducing computation time [19–21]. Lin et al. [56] introduce advanced methods for analyzing 3D point clouds, underscoring the wide range of applications and the dynamic development of neural network structures in various fields of image analysis. T. Li and his colleagues [22] used GoogleNet, ResNet, DensNet, and VGG-16 and then compared them to find the best model. Transfer learning networks are pre-trained; therefore, implementation requires a considerably smaller dataset. Before being sent to the transfer learning variant layers, all fundus images were pre-processed and scaled to 224 $\times$ 224 pixels. The Inception model performed best with 82 % accuracy. Aujih et al. [23] proposed EDR-Net, a novel deep neural network architecture featuring depth-wise separable convolution to optimize computational efficiency. The EDR-NET was trained using the

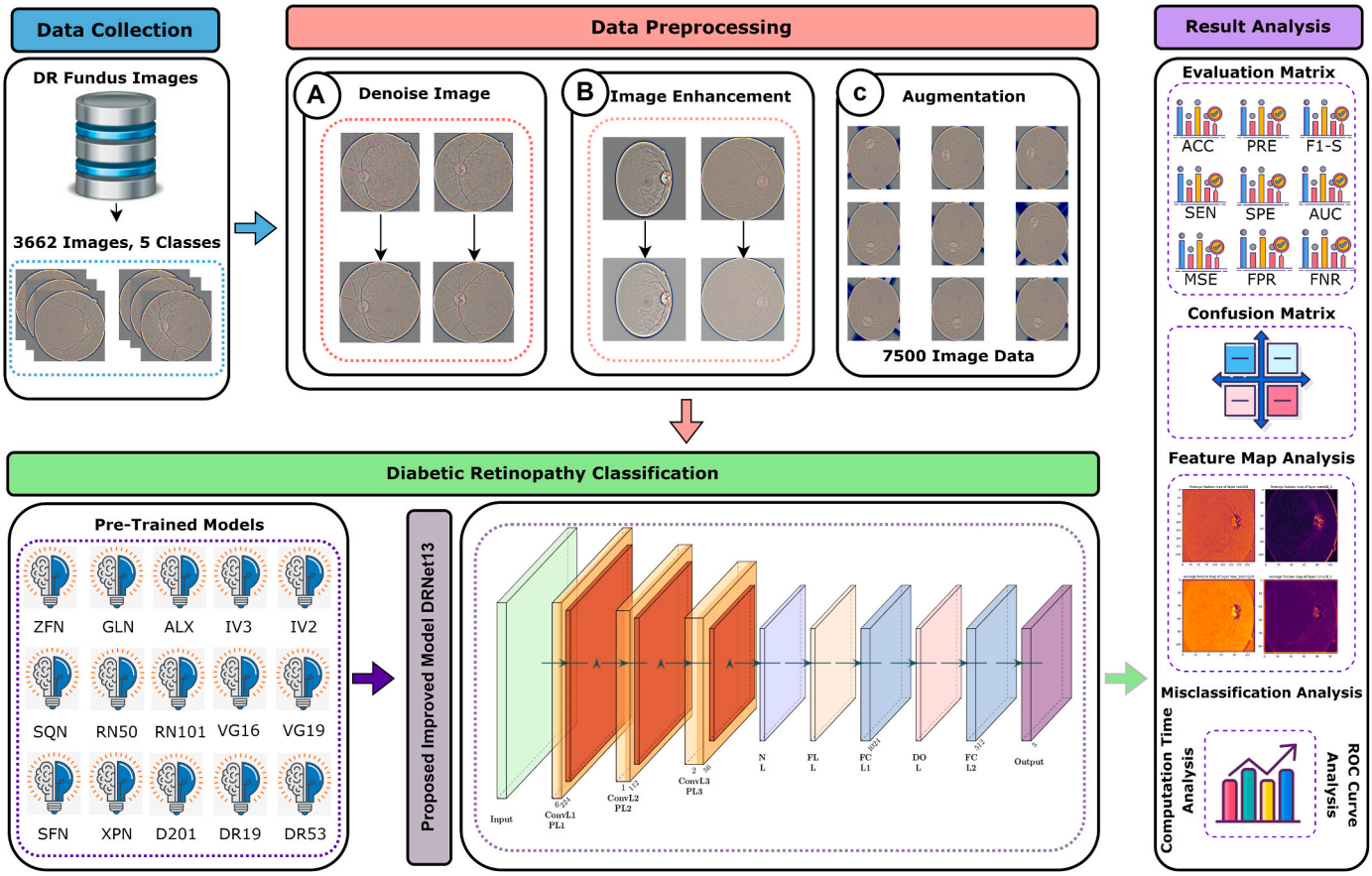

Fig. 1. The overview of the entire system for diabetic retinopathy detection from fundus images, illustrating the steps from image processing to diagnostic output.

DRKaggle-train dataset, encompassing 35,126 images, and tested against the DRKaggle-test and Messidor-2 datasets. This approach reduced computational demands and achieved predictive performance comparable to existing methods.

Furthermore, Minarno et al. [24] evaluate the EfficientNet-B7 model for classifying diabetic retinopathy using the APTOS 2019 dataset. Through hyperparameter tuning and preprocessing techniques, the study enhances the model performance. It attains 89.1 % training accuracy and 84.36 % test accuracy, showing a comprehensive study with existing methods. This research suggests further exploration of image enhancement techniques to advance diabetic retinopathy detection. Dayana et al. [25] proposed an optimized deep neural network with a Chronological Tunicate Swarm Algorithm for Diabetic retinopathy classification. Utilizing U-Net and sparse Fuzzy C-means-based hybrid entropy model, they effectively segmented the optic blood vessels from low-quality fundus images. The model trained on the DIARETDB0 and DIARETB1 datasets achieved high accuracy with 95.9 % and 95.48 %, respectively, demonstrating its robustness and effectively classifying DR severity. However, the study's limitation lies in its reliance on specific datasets, which may affect its generalizability to other datasets or the real world.

Similarly, Alghamdi et al. [26] evaluated three deep-learning models for automatic Diabetic Retinopathy detection using color fundus images. They optimized these models using early stopping and dropout techniques to mitigate overfitting. The study used a publicly available Kaggle dataset, presenting challenges due to its variability and imbalances. Among the models, VGG-16 showed the highest binary and multistage DR classification accuracy. However, despite high accuracy, the study highlighted a need for improved explainability in these models for medical diagnosis, indicating limitations in their current ability to detect DR-related lesions reliably. Five different classes of diabetic retinopathy are introduced by B. Tymchenko et al. [27]. This study used the APTOS 2019 Blindness Detection Dataset, which included 13000 images. A screening approach for early diagnosis of diabetic retinopathy was presented, with a sensitivity and specificity of 0.99, placing it at position 54 out of 2943 methods.

Additionally, Nasir et al. [28] developed a Convolutional Neural Network model to classify Diabetic Retinopathy stages using the Indian Retinopathy Image Dataset (IDRID). The CNN model, designed with layers including convolutional, subsampling, flattened, dense, and dropout layers, achieved a testing accuracy of 96 % and sensitivity of 82.9 %. While the model showed high efficacy in detecting DR from high-quality retinal images, its performance may not generalize well to real-world conditions due to the dataset's specificities and the varying quality of images in practical settings. Serrano et al. [29] utilized a MTLAB retained AlexNet Convolutional Neural Network to detect glaucoma and diabetic retinopathy using retinal fundus images. Employing transfer learning, they trained the network on diverse datasets like APTOS, HRF, sjchoi86-HRF, LAG, and ODIR. The NetTransfer I–V models demonstrated high validation accuracies ranging from 89.7 % to 94.3 %. These results showcase the effectiveness of retrained CNNs in eye disease detection. The study's dependence on diverse datasets, each with its own unique features, raises concerns about the models' stability and adaptability in various clinical contexts. Moreover, M. Tian et al. [30] and others have developed a fine-grained attention network (FA + KC-Net) for evaluating the severity of diabetic retinopathy. Fine-detailed fundus images and their medical interpretation were extracted using FA-Net and KC-net, respectively; a refining block was built with two decision rules to merge FA-Net and KC-net for final grading. According to the estimated results, the FA + KC-Net also gives superior grading compared to other SOTA deep learning models.

To enhance the efficacy of their research, Mustafa et al. [31]

proposed a multi-stream ensemble deep neural network for the classification of diabetic retinopathy severity. They integrated ResNet-50 and DenseNet-121 for effective feature extraction and utilized principal component analysis to reduce feature dimensionality. The model was further refined with ensemble classifiers like AdaBoost, and the random forest was evaluated on EyePACS, Messidor-2, APTOS, and DDR datasets, achieving up to 95.58 % accuracy on the Messidor-2 dataset. Atwany et al. [32] conducted a comprehensive review and analysis of state-of-the-art deep learning methods for Diabetic Retinopathy (DR) classification. Their paper emphasizes the importance of early detection of DR, a progressive disease caused by high blood glucose levels and a leading cause of blindness in diabetic patients, particularly in the employed communities of developing nations. The study discusses various challenges in the manual diagnosis of DR, including the time, money and effort involved in the process and how machine learning and deep learning algorithms have aided in assessing retinal fundus images for early diagnosis.

M. Nahiduzzaman et al. [33] used a CNN to extract 120 features to identify diabetic retinopathy using a parallel convolutional neural network based on the ELM classifier. Two datasets, Kaggle DR 2015 competition (Dataset 1; 34,984 FIs) and APTOS 2019 (3662 FIs), were used to evaluate the performance of the demonstrated model. K. Oh et al. [34] utilized deep learning and ultra-widefield fundus images to detect diabetic retinopathy. The ResNet-34 model employed ETDRS-7SF and F1–F2 fundus images to evaluate research outcomes and performance. Seeking to advance the field of medical imaging analysis, Jaskari et al. [35] concentrated on improving the automated classification of DR using deep learning networks. Their study highlighted the significance of estimating the variance of classifications, a critical aspect often neglected in modern neural network models. They employed approximate Bayesian neural networks (BNNs) and introduced a classifier using weighted cohen's kappa. Jian et al. [36] introduced Triple-DRNet, a sophisticated triple-cascade convolutional neural network designed to improve the classification of diabetic retinopathy in fundus images. This network is structured into three subnetworks: DR-Net, PDR-Net, and NPDR_Net, each focusing on specific stages of diabetic retinopathy. The model tested on the APTOS 2019 dataset and achieved an accuracy of 92.08 %.

Deep learning medical image analysis has sparked an advancement in the early identification and classification of diseases such as diabetes. As the literature reflects a growing number of automated systems leveraging various CNN models, we have identified the gap in developing models that maintain high accuracy without extensive computational resources. Our decision to adopt and further develop DRNet13 was motivated by a thorough study of current approaches. This model was designed to address the limitations of manual analysis while accommodating the constraints of computational resource availability, an aspect often overlooked in the design of deep learning systems of medical diagnostics. We have also included a comparison of DRNet13 with other pre-trained models to highlight its enhanced performance and efficiency.

3. Methodology

This section gives an extensive explanation of our study approach, which begins with data collection. We gathered high-quality data from Kaggle, a trustworthy open-source platform, concentrating on datasets relevant to diabetic retinopathy (DR) research. Following data collection, we begin processing the data, which entails refining and transforming raw data into a format suitable for our models. This stage includes key steps like image noise removal to eliminate irrelevant details, image enhancement to highlight critical features, and image augmentation to increase the variability and quantity of our data. This intensive pre-processing pipeline ensures that our dataset is primed for subsequent analysis.

Next, we discuss our application of a number of well-known, pre-trained models with impressive image classification capabilities. These models have been thoughtfully chosen to classify DR stages from our pre-processed fundus images, allowing for efficient and precise disease stage identification. Lastly, we present DRNet13, our innovative model. This proposed model is specifically designed to overcome the limitations of existing models. By incorporating cutting-edge methodologies, DRNet13 intends to considerably enhance the performance and accuracy of DR stage classification, making it a promising model in this vital area of medical research.

3.1. Dataset overview

To conduct this analysis, we used the Diabetic Retinopathy 224 × 224 Gaussian Filtered dataset [37]. This dataset was created by compiling 3662 images of filtered retina scans from the Aravind Eye Hospital in India. These images were classified into five groups: No-DR, Mild, Moderate, Severe, and Proliferate-DR. Those images are provided in PNG format with descriptive information. Table 1 comprises details about the data collection sources for each class. The test and validation set each have 733 unique images, whereas the training set has 2197 images and their associated data. The sample of the original dataset is depicted in Fig. 2.

3.2. Data preprocessing

Data preprocessing ensures that the data is in a structure that the deep learning network can understand. The preprocessing of image data is a vital step in preparing it for deep learning models. It entails converting unprocessed image data towards a format that is useable by the model.

3.2.1. Image noise removal

Accurate prediction of diabetic retinopathy is highly dependent on the removal of image disturbance. Our research addresses this difficulty by employing a median filter, an effective noise reduction technique. This method of nonlinear digital filtering works by replacing each pixel value with the median of neighboring pixel values. Its greatest strength is its resistance to extreme outliers, or pixels with atypical intensity values caused by noise or artefacts. By reducing the effect of these outliers, a noise-free image is produced. This image is then optimally positioned for subsequent analysis steps, such as feature extraction or the deployment of machine learning models [38]. The result is a highly reliable and accurate method for predicting diabetic retinopathy, which is facilitated by high-quality, noise reduced images obtained through efficient median filtering.

$$img[x,y] = median\{img1[p,q] \mid (p,q)\epsilon Z_{xv}\} \tag{1}$$

where $img[x,y]$ denotes the pixel value at coordinates $[x,y]$ in the processed image after median filtering has been applied, $median\{img1[p,q]$ refers to the median value calculated from the set of neighboring pixel values surrounding the pixel at coordinates $[p,q]$. Besides, $img1[p,q]$ is a pixel value at coordinates $[p,q]$ in the original, unfiltered image. The (p,q) are the coordinates of the pixels in the neighborhood of the pixel at $[x,y]$, that are considered in calculating the median value. Z_{xv} indicates an area of value for a pixel set by the user and is centered around the coordinates $[x,y]$ in the images of diabatic retinopathy. Fig. 3 illustrating

Table 1
The raw dataset's description.

No.	Classes	Images
1	No DR	1805
2	Mild	370
3	Moderate	999
4	Severe	193
5	Proliferate DR	295

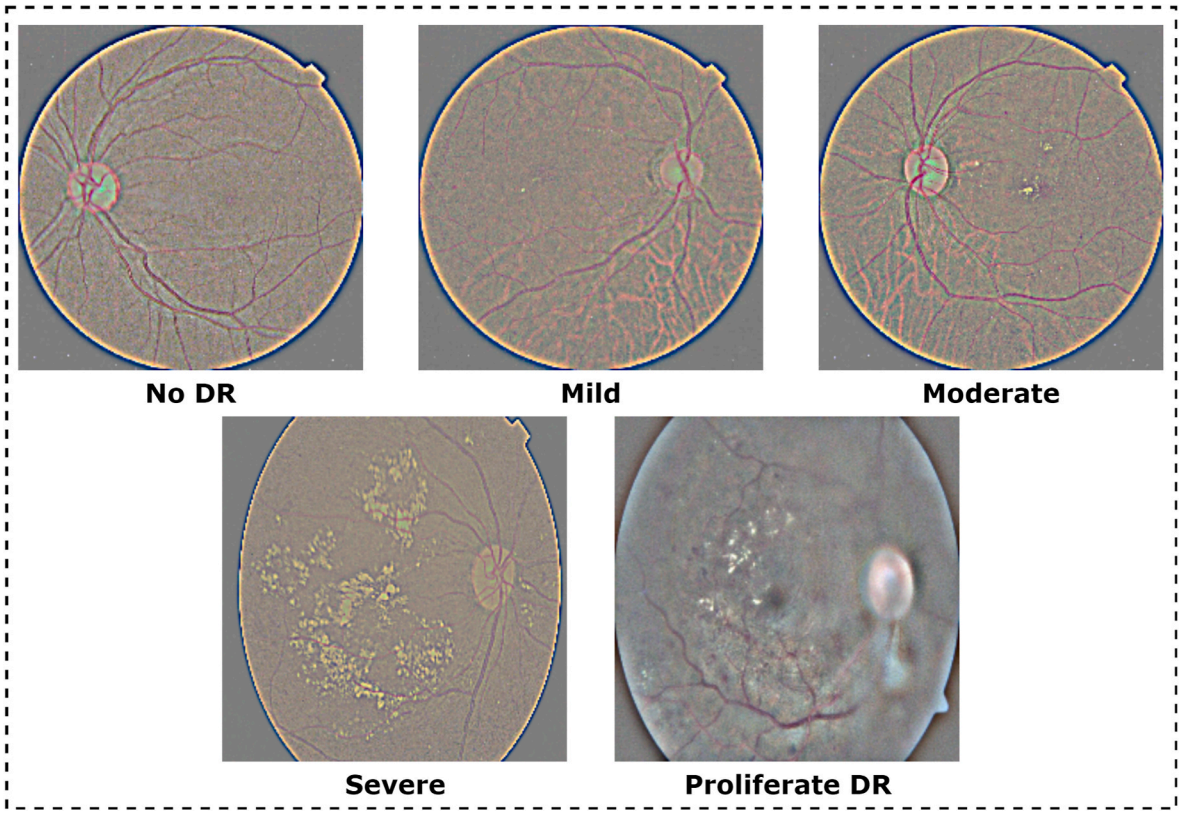

Fig. 2. The dataset under consideration contains five distinct classifications. These categories include the five stages of Diabetic Retinopathy, namely No Diabetic Retinopathy (No DR), Mild, Moderate, and Severe stages, and Proliferative Diabetic Retinopathy (Proliferative DR).

the outcome of the median filter strategies for noise reduction from diabetic retinopathy images.

3.3. Image enhancement

Gamma correction is a non-linear adjustment that is applied to each pixel's value. Typically, linear operations such as addition, subtraction, and multiplication are performed on every pixel [39]. Gamma correction is accountable for applying nonlinear methods to the pixels of the image's input, thereby transforming the image's saturation. Gamma correction has emerged as a technique for adjusting an image's light intensity. Fig. 4 depicted the outcome and improvement of images by applied gamma correction strategy through histogram. Below is the method utilized to obtain a gamma-corrected image:

$$O = \left(\frac{I}{255}\right)^{\frac{1}{\gamma}}.255 \tag{2}$$

where I indicate input pixel value [0,255], O denoted for output pixel value [0,255], γ denoted for gamma that determines image light intensity. If gamma is below 1, the image will be darker; if it is above 1, the imagery will be considered brighter. A value of 1 has no impact at all.

3.4. Image augmentation

A large quantity of input data is required for a deep learning model to perform optimally. In this work, we used a variety of data augmentation procedures to increase the breadth and quality of our dataset. We can improve our proposed algorithms' performance and prediction ability by introducing a varied selection of different sampling into the training datasets. The training dataset's variety and breadth significantly impact the model's accuracy and precision. In this instance, using image enhancing methods leads to an increase in result accuracy. Furthermore, data augmentation techniques are an effective tool for increasing the diversity of datasets. Data augmentation methods are often used to increase the quantity of training sets, providing high-capacity learners with an even more diversified, relevant training environment [40,41]. Mirroring, rotating, zooming, flipping, and cropping are some of the most frequent augmentation methods. In this study, the dataset changes employing the two oversampling and undersampling strategies. Firstly, a random undersampling strategy is used for the excessively prevalent class 'No DR' in the data. This approach randomly removes data from the majority class, reducing the total data per class to 1500 instances. Oversampling (data augmentation) approaches are then used to enhance classes with inadequate data representation ('Mild,' 'Moderate,' 'Severe,' 'Proliferate'). This research involves a variety of augmentation methods. Table 2 provides a complete review of all augmentation strategies used.

Following the implementation of the data augmentation procedures, the dataset volume increased to include a sum of 7500 images. This was divided equally across all classes, with 'No DR,' 'Mild,' 'Moderate,' 'Severe,' and 'Proliferate DR,' each including 1500 images. Fig. 5 depicts the observable outcomes of this data augmentation approach.

3.5. Data Management structure

Before starting the training procedure, the dataset must be correctly partitioned. In our research, we separated the fundus imagine data into three unique subsets using a 60:20:20 ratio, which was then assigned to the training, validation, and test sets. The original dataset, which included 7500 images illustrating different phases of diabetic retinopathy, was methodically divided into these three categories. The training set received the most interest, accounting for 60 % of the total data or 4500 images. This large collection of photos is used to teach the model about the various features and characteristics of the condition. The remaining 40 % of the data, or 3000 images, was divided evenly

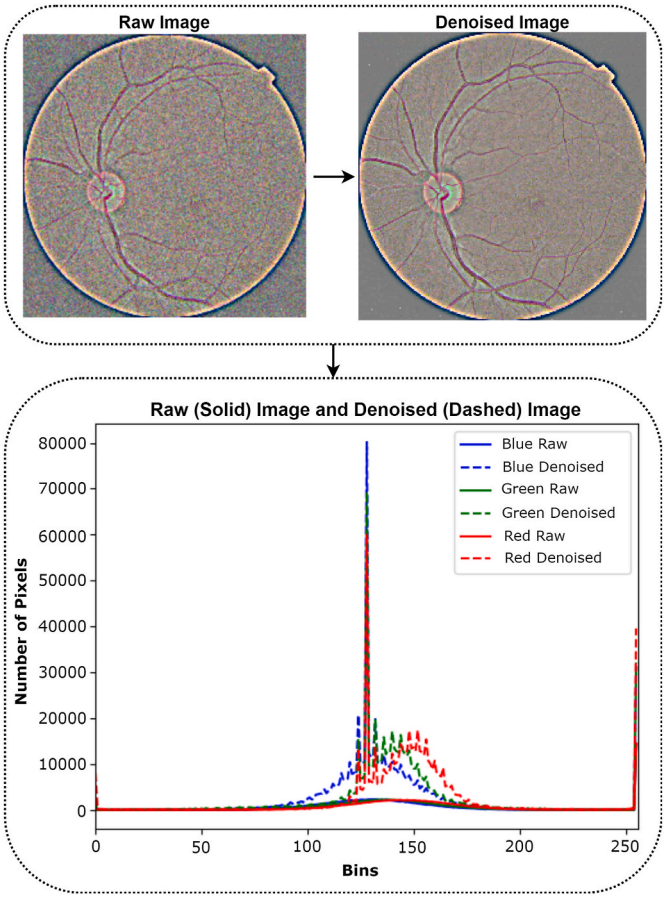

Fig. 3. The figure depicts the improved image quality obtained by using a median filter to reduce noise. The histograms of the original and denoised images are also shown for comparison, demonstrating the significant increase in image clarity and overall quality after filter application.

between the validation and testing sets, with each getting 1500 images. The validation set is essential throughout the training phase because it allows us to fine-tune the parameters of our model and optimize its performance while minimizing the danger of overfitting. The testing set contains 1500 images and is kept aside and used only once the training is completed. This collection is used to assess our model's final performance, offering a credible approximation of how it would perform when provided with fresh, previously unknown data.

Table 3 presents a detailed description of the dataset after the pre-processing processes. It represents the final distribution of the data over the three subsets, training, validation, testing, and the outcomes of all pre-processing and data augmentation activities done on the dataset.

3.6. Applied models selection

This study used fifteen different pre-trained models to simplify comparison with our proposed DRNet13 model. ZFNet, GoogLeNet, AlexNet, InceptionV3, InceptionResV2, SqueezeNet, ResNet50, ResNet101, VGG16, VGG19, ShuffleNet, Xception, DenseNet201, DarkNet19, and DarkNet53 are among the networks that were used in our study. This variety provides a thorough study of performance, accuracy, and efficiency, giving a strong foundation for comparison with our proposed model.

3.6.1. ZFNet network

Deconvolution layers are used in ZFNet to visualize feature maps within the network. It sheds light on the network's image processing and feature learning processes. Also, ZFNet was the first to use "local response normalization" layers. These layers make each neuron's output the same, reducing the effect of internal covariate shift and speeding up the network's learning. It had three fully connected layers, one dropout layer, and three max-pooling layers for five convolutional layers. A 7×7 filter was utilized in the first layer with a lower stride value [42]. The first three convolutional layers are followed by pooling layers during down-sampling, whereas deconvolution layers follow the final two convolutional layers during up-sampling. The normalization layers are positioned between the first and second convolutional layers.

3.6.2. GoogleNet network

The GoogLeNet architecture is a convolutional neural network that takes cues from the popular Inception design. GoogLeNet uses an iterative algorithm as its foundation in an effort to improve computational efficiency [43]. As more layers are added on top of each other, the input from the layers below is filtered in parallel. It uses inception modules, which provide the network with a selection of convolutional filter sizes within each block. These modules can be stacked in two steps to make an inception network that lowers the grid's resolution. GoogLeNet is a deep neural model that only has 22 layers and a smaller number of parameters than the inception architecture. Images of 224×224 pixels or higher quality have been tested and accepted by the pre-trained network. Layers of activation, average pooling, and density make up the architecture. In place of a fully connected layer, GoogLeNet utilized global average pooling.

3.6.3. AlexNet network

The AlexNet model of convolutional neural network is immensely popular. Max pooling, convolutions, and dense layers are all fundamental components of AlexNet. Model fitting occurs over two GPUs with the help of group convolutions. Within every possible combination of rewards, AlexNet has eight layers [44]. The model is made up of five convolutional layers, two normalizing layers, and a softmax layer. The fully connected and max pooling layers are the backbone of the model. Each "layer" is made up of a convolutional layer and a ReLU-based nonlinear activation function. Maximum pooling is achieved by employing pooling layers. As a result of how layers are seen, the input size is capped at $224 \times 224 \times 3$ pixels. When a grayscale image is used as a source, an RGB image is generated by multiplying the individual color channels. The model has 60 million parameters, and the batch size is 128.

3.6.4. InceptionV3 network

InceptionV3 is a conInceptionV3 is a convolutional neural network created for image recognition and classification. It emphasizes 1×1 convolutions and consists of convolutional and pooling layers. During training, the network utilizes inception modules with reduced dimensionality and additional classifiers. The final layers include global average pooling, dropout for regularization, and a softmax layer for classification [45]. For dimensionality reduction, 1×1 convolutions are used, and batch normalization is applied to each convolutional layer. Model weights can be trained from inception or used as pre-trained weights for transferable learning with InceptionV3. Due to its efficient performance, InceptionV3 is frequently used for various image-related duties.

3.6.5. InceptionResV2 network

The InceptionResNetV2 network architecture integrates the Inception module with residual connections. It is composed of multiple inception blocks, each of which contains a unique convolutional layer. The network begins with a root block, followed by multiple inception blocks piled on top of one another. Each inception block contains 1×1 convolutions, 3×3 convolutions, 5×5 convolutions, and pooling layers as its detailed layers [46]. InceptionResNetV2 also consists of skip connections and residual units, which facilitate data transfer between layers. Additional layers, including batch normalization, global average

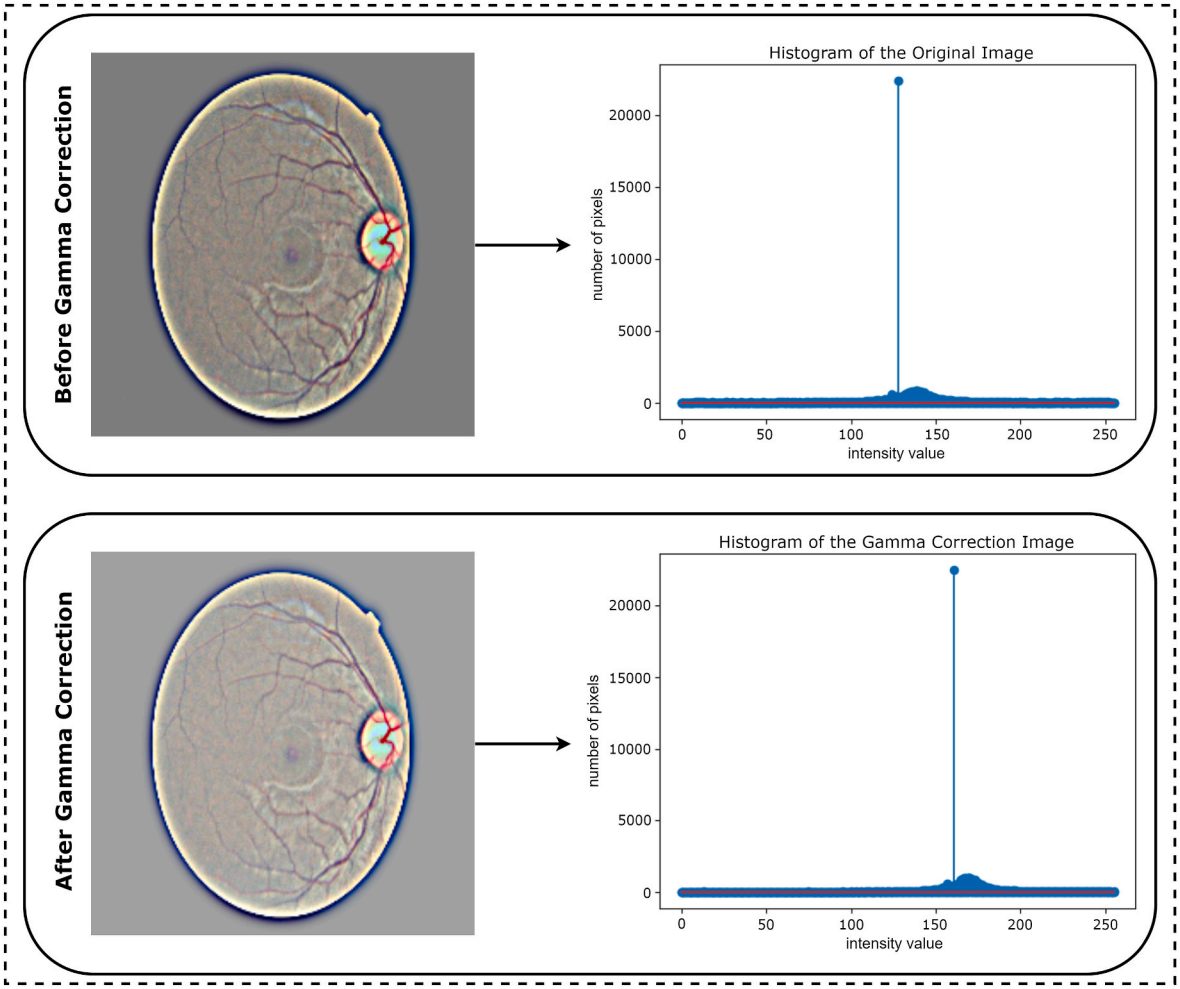

Fig. 4. This demonstrates the substantial improvement in image quality gained by using the gamma correction approach.

Table 2
All parameters for augmented data.

Augmentation Parameters	Applied Values
Rotate	45
Rotate Right	90
Rotate Left	90
Rotate Horizontal	45
Rotate Vertical	45
Horizontal Flip	True
Vertical Flip	True
Translate	x,y (30.0, 15.2)

pooling, and a softmax layer, are employed for classification. InceptionResNetV2 excels at tasks such as image recognition and feature extraction due to its robust architecture and advanced layers.

3.6.6. *SqueezeNet network*

SqueezeNet is a convolutional neural network architecture intended for efficient deep learning on resource-limited devices. It accomplishes high precision while reducing model dimensions by a significant amount. SqueezeNet employs fire modules comprising squeeze layers (1 $\times$ 1 convolutions) to reduce dimensions and expand layers (1 $\times$ 1 and 3 $\times$ 3 convolutions) for collecting additional features. In addition, skip connections are utilized to keep the gradient flow and facilitate information propagation [47]. SqueezeNet's compact model allows quicker inference and deployment on devices with limited memory and computational resources. It is commonly used for real-time applications such as object detection and classification on mobile devices and embedded systems.

3.6.7. *ResNet50 network*

ResNet50 is a well-known convolutional neural network architecture characterized by its deep structure and residual learning. It revolutionized the classification of images along with other computer vision operations with its 50 layers. ResNet50's main characteristics include skip connections, batch normalization, and residual blocks. Skipping connections permits gradients to travel directly across layers, resolving the problem of gradients vanishing. Layer inputs are normalized via batch normalization, which improves training stability and accelerates convergence [48]. Residual blocks enable the successful training of exceedingly deep neural networks by facilitating learning residual functions. In the field of deep learning, ResNet50 is frequently used as a standard for evaluating new models and techniques.

3.6.8. *ResNet101 network*

ResNet101 is an enhanced convolutional neural network design that expands on ResNet50's success. It has 101 layers, which makes it deeper and more effective for complex image recognition tasks [48]. ResNet101, similarly to ResNet50, has skip connections, batch normalization, and residual blocks. The use of skip connections allows for the efficient transfer of gradients between layers, hence overcoming the problem of gradients disappearing. By normalizing layer inputs, batch normalization helps to stabilize training. Residual blocks enable the learning of residual functions, making deep network training easier.

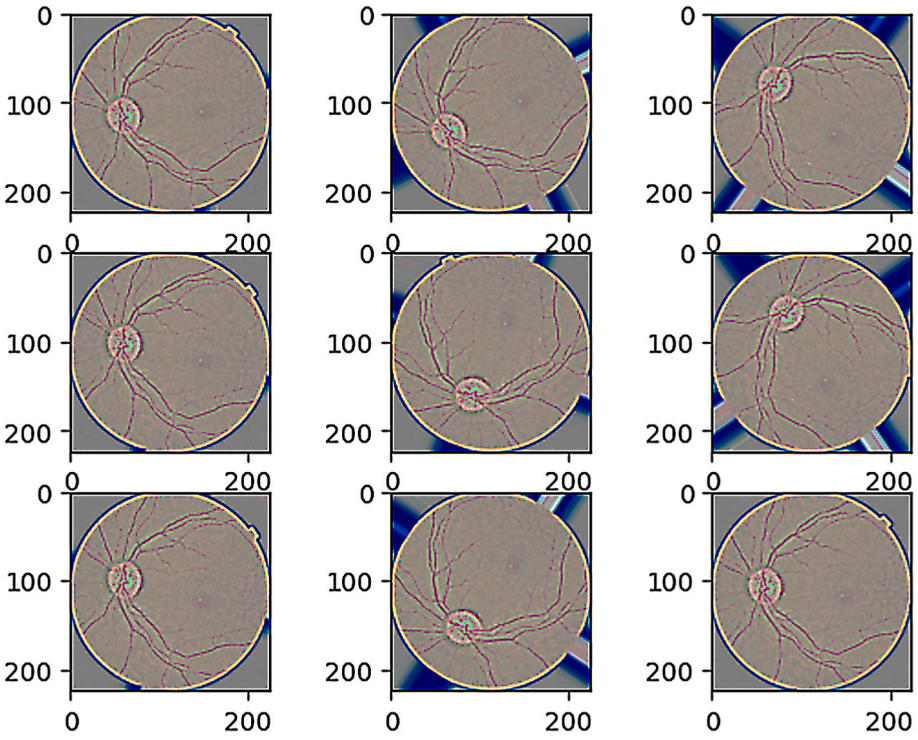

Fig. 5. Various data augmentation techniques, such as 45rotation, 90right rotation, 90left rotation, 45horizontal rotation, 45vertical rotation, horizontal flipping, vertical flipping, and translation in the x and y dimensions by (30.0, 15.2), were used to balance all classes. These approaches are intended to improve the training examples by enhancing the dataset's variety and balance.

Table 3
The final dataset with applied techniques on the raw images of diabetic retinopathy.

Final Features	Values
Total Number of Images	7500
Total Number of Classes	5
Image Noise Removal	Median Filter
Image Enhancement	Gamma Correction
Number of Augmentation Technique	8
No DR	1500
Mild	1500
Moderate	1500
Severe	1500
Proliferate DR	1500
Train Set	4500 (60 %)
Validation Set	1500 (20 %)
Test Set	1500 (20 %)

ResNet101's improved depth provides better representational capacity and has exhibited good performance on various hard computer vision applications, cementing its place as cutting-edge architecture.

3.6.9. VGG16 network

VGG16 is a well-known convolutional neural network architecture that is well-known for its ease of use and performance in image categorization applications. It has 16 layers, which include convolutional and fully connected layers. VGG16 has a sequential structure consisting of successive sets of 3×3 convolutional layers, followed by a 2×2 max-pooling layer [49]. Through these layers, the network gradually learns increasingly complicated traits. It also includes batch normalization for regularization and extensively uses the ReLU activation function. The fully connected layers of VGG16 at the network's conclusion execute categorization. While VGG16 has more parameters than other designs, it has exhibited good performance on numerous image recognition problems and serves as a field benchmark.

3.6.10. VGG19 network

VGG19 is a modified version of the VGG16 network, notable for its deeper structure and enhanced performance. It has 19 layers, including convolutional and fully connected layers. VGG19, like VGG16, has a sequential structure comprising sets of 3×3 convolutional layers and 2×2 max-pooling layers. The new layers in VGG19 boost representational capacity and allow the network to learn more complicated characteristics [49]. It also employs batch normalization for regularization and the ReLU activation function. In the end, VGG19's fully connected layers execute categorization. VGG19's deeper design provides better capabilities for image identification tasks and has been extensively utilized for various computer vision applications.

3.6.11. ShuffleNet network

ShuffleNet is a convolutional neural network design that optimizes computing efficiency while maintaining accuracy. It employs group convolutions and channel shuffling to decrease calculations while retaining performance [50]. ShuffleNet comprises pointwise group convolutions for spatial information, depthwise convolutions for spatial information, and channel shuffling for information flow across channels. It makes use of bottleneck units, which combine 1×1 and 3×3 convolutions. Pointwise convolutions that are grouped facilitate efficient channel interaction. The network comes to a close with global average pooling, which reduces feature mappings to a single vector. The architecture of ShuffleNet allows it to attain high accuracy with fewer parameters, making it suited for resource-constrained applications where computational simplicity is critical.

3.6.12. Xception network

Xception is a convolutional neural network architecture modeled after the Inception network, which is renowned for its superior performance in image classification tasks. It introduces an advanced version of depthwise separable convolutions to improve learning of features and reduce computational complexity. Xception replaces typical convolutional layers with depthwise separable convolutions consisting of

depthwise and pointwise convolutions. These depth-separable convolution layers are interspersed with skip connections [51]. In addition to batch normalization for regularization, the network employs the ReLU activation function. The architecture of Xception enables efficient feature extraction and has demonstrated outstanding performance on various computer vision problems.

3.6.13. DenseNet201 network

DenseNet201 is a convolutional neural network architecture renowned for its exceptional performance in image classification tasks and dense connectivity. It has 201 layers, making it more profound and potent than earlier versions. DenseNet201 consists of dense blocks in which each layer is connected to every other layer. This dense connectivity improves feature reuse and gradient flow, thereby enhancing precision. Transition layers consisting of 1×1 convolutions and pooling are used to control the number of parameters and reduce dimensionality [52]. DenseNet201 also implements batch normalization and the ReLU activation function for regularization. With its dense connectivity and significant structure, DenseNet201 accomplishes cutting-edge performance on various computer vision problems.

3.6.14. DarkNet19 network

DarkNet19 is a convolutional neural network architecture well-known for its ease of use, efficiency, and great performance in various computer vision applications. It has 19 convolutional layers and is equipped with a YOLO object recognition system [53]. DarkNet19 has a simple topology, including 3×3 and 1×1 convolutional layers and max-pooling layers. There are no residual connects or complicated skip connections in the network. It relies on powerful feature extraction skills to recognize objects accurately. DarkNet19 is designed to be small, needing fewer processing resources than more complicated networks. This makes it appropriate for applications that operate in real-time where efficiency is critical while retaining competitive performance.

3.6.15. DarkNet53 network

DarkNet53 is an architecture of convolutional neural networks that functions as the basis of the YOLOv3 object recognition system. It comprises 53 convolutional layers and is known for its intricate structure and high accuracy. DarkNet53 employs primarily 3×3 and 1×1 convolutional layers to achieve a straightforward and efficient architecture. The network contains neither residual nor intricate skip connections [54]. Instead, it concentrates on extensive feature extraction to enable robust object detection. DarkNet53 provides superior detection capabilities and enhanced precision in comparison to its predecessors. Widely utilized for real-time object detection tasks, it offers a decent compromise between computational efficiency and precision.

3.6.16. Proposed DRNet13 network

Diabetic retinopathy is a condition characterized by harm to the retina caused by diabetes, often resulting in vision loss. Early identification of this disease is crucial, as timely, appropriate treatment may stop or delay vision loss. Our proposed DRNet13 model demonstrated outstanding results in the recognition of retinopathy images tasks. The proposed DRNet13 model comprises various components, including the input layer, three sets of convolutional layers and pooling layers, a normalization layer, two fully connected layers (dense layers), a dropout layer, and the output layer. The architecture of the proposed model is given below:

- **Input Layer:** The DRNet13 model takes in retinal images I_{in} with dimension 224×224 and three-color channels (RGB).

$$I_{in} = R^{224\times224\times3} \tag{3}$$

- **Convolutional Layer 1 & Pooling Layer 1:** Initiates feature extraction with 64 filters f_{64}, applying a ReLU activation $\varnothing$, followed by a pooling operation P_{L1} having the dimensionally. Here, $Conv_L$ indicates convolutional layer.

$$Conv_{L1} = \varnothing(f_{64} * I_{in}) \tag{4}$$

$$P_{L1} = pool(Conv_{L1}) \tag{5}$$

- **Convolutional Layer 2 & Pooling Layer 2:** Enhance feature detection with 128 filters f_{128}, followed by pooling P_{L2} to reduce dimensions while preserving essential features.

$$Conv_{L2} = \varnothing(f_{128} * P_{L1}) \tag{6}$$

$$P_{L2} = pool(Conv_{L2}) \tag{7}$$

- **Convolutional Layer 3 & Pooling Layer 3:** Extracts higher-level features with 256 filters f_{256}, then a pooling layer P_{L3} further reduces dimensionality.

$$Conv_{L3} = \varnothing(f_{256} * P_{L2}) \tag{8}$$

$$P_{L3} = pool(Conv_{L3}) \tag{9}$$

- **Normalization Layer:** Normalizes the outputs from P_{L3} to aid model training and robustness. Here, N_L indicates normalization layer.

$$N_L = norm(P_{L3}) \tag{10}$$

- **Flattening Layer**: Transforms the 3D tensor from the normalization layer into 1D vector FL_L. Here, FL_L indicates the flattened layer.

$$FL_L = flatten\ (N_L) \tag{11}$$

- **Fully Connected Layer 1:** Utilizes the flattened vector for initial classification, reducing dimensions to 1024 nodes. Here, FC_L denoted as fully connected layer, W_1 are the learned weights, FL_L is the input vector from the flattened layer, b_1 is the bias term and $\varnothing$ is the activation function.

$$FC_{L1} = \varnothing(W_1.FL_L + b_1) \tag{12}$$

- **Dropout Layer:** Randomly drops out neurons to prevent overfitting while maintaining the dimensionally. Here, DO_L represent dropout layer.

$$DO_L = dropout\ (FC_{L1}) \tag{13}$$

- **Fully Connected Layer 2:** Continues the classification process, decreasing nodes from 1024 to 512.

$$FC_{L2} = \varnothing(W_2.DO_L + b_2) \tag{14}$$

- **Output Layer:** Determines the class probabilities using a softmax function. Here, σ indicates the softmax activation function.

$$O_L = \sigma\ (W_3.FC_{L2} + b_3) \tag{15}$$

The DRNet13 network is a robust convolutional neural network specifically created for the purpose of diabatic retinopathy (DR). It contains a variety of features that collectively improve its effectiveness in diagnosing the disease. The fundamental component of DFNet13's architecture is a specialized input layer designed for analyzing retinal images. The images labelled as I_{in}, have dimensions of 224×224 pixels and include color channels (RGB), allowing for a detailed capture of retinal features essential for DR identification. The network's power is

enhanced by its convolutional and pooling layers. DRNet13 utilizes three sets of these layers, whereas the convolutional layers progressively escalate in intricacy, commencing with 64 filters in the initial layer and expanding to 256 in the third layer. This design allows the network to extract a diverse range of features, ranging from basic to more complex ones, that are necessary for identifying different signs of DR. The pooling layers (P_{L1}, P_{L2}, P_{L3}) are applied after each convolutional layer to decrease the feature maps' spatial dimensions progressively. This helps to reduce the computational burden and the model's vulnerability to overfitting. A normalization layer (N_L) has been included to normalize the outputs of the third pooling layer (P_{L3}). Normalization is essential as it stabilizes the learning process and improves the model's overall resilience, a critical aspect in accurately detecting DR.

The flattening layer (FL_L) is responsible for transforming the multi-dimensional feature maps into a format that is appropriate for classification. It achieves this by reshaping the 3D tensor from the normalization layer into a 1D vector. This transformation has significance for combining the spatial information that has been extracted in the convolutional stages. Furthermore, the network consists of two fully connected layers (FC_{L1}, FC_{L2}) that utilize the flattened vector for classification. The initial layer decreases the dimensionality to 1024 nodes, while the subsequent layer further decreases it to 512 nodes. Each layer in the model incorporates a Rectified Linear Unit (ReLU) activation function to introduce non-linearity. In addition, a dropout layer (DO_L) is strategically included to mitigate overfitting, ensuring the model's efficacy and generalizability to new data. The final stage of this sequential architecture is the output layer, which employs a softmax function to produce the probabilities associated with each class. This is especially appropriate for problems involving several classes, such as the identification of diabetic retinopathy, where the model must accurately categorize whether the condition is present or absent.

The algorithm is significant for the DRNet13 model in particular since it describes a cutting-edge method for detecting diabetic retinopathy. In the discipline of ophthalmology, this may result in better diagnostic equipment and patient outcomes. Fig. 6 visualizing the overall architecture of DRNet13. Where the different layers are indicated as follows: 'ConvL' means for 'Convolutional Layer,' 'PL' stands for 'Pooling Layer,' 'N L′ stands for 'Normalization Layer,' 'FL L′ stands for 'Flatten Layer,' 'FC L′ stands for 'Fully Connected Layer,' and 'DO L′ is for 'Dropout Layer. Each layer serves a distinct and critical role in the operation and performance of the DRNet13 model, adding to its superior capabilities in Diabetic Retinopathy diagnosis. Table 6 described each layer properties with functionalities and the associations throughout the layer and the novel features of the DRNet13 model are described sequentially below:

Algorithm 1 DRNet13 A Deep Neural Model for DR Detection.

Input: Preprocessed retinal images.
Output: Probability of Diabetic Retinopathy Existence.

(*continued on next column*)

(*continued*)

Algorithm 1 DRNet13 A Deep Neural Model for DR Detection.

Conv L1: 64 filters, followed by activation function (ReLU)
Pooling L1: Reduce size to 112 × 112 × 64
Conv L2: 128 filters, followed by activation function (ReLU)
Pooling L2: Reduce size to 56 × 56 × 128
Conv L3: 256 filters, followed by activation function (ReLU)
Pooling L3: Reduce size to 28 × 28 × 256
N L: Normalize the output
FL L: Flatten the output one dimension 200704
FC L1: Reduce dimension to 1024 nodes, apply activation function (ReLU).
DO L: Apply for prevent overfitting
FC L2: Reduce dimension to 512 nodes, apply activation function (ReLU)
Output: Apply softmax function.

Here, Conv L denoted by "Convolutional layer", N L denoted by "Normalization Layer", FL L denoted by "Flatten Layer", FC L denoted by "Fully Connected Layer", DO L denoted by "Dropout Layer".

Table 4 describes the construction of the DRNet13 model. Each row corresponds to a layer. "Stride & Padding" are convolution and pooling layer parameters that regulate the movement and adjustment of the kernels. "Output Shape" specifies the dimensions of the output from each layer. The "Kernel" size denotes the size of the filter utilized. The "Activation Function" describes the function used to create non-linearity, with ReLU employed in intermediate layers and SoftMax in the output layer. "Dropout" is a regularization method used to avoid overfitting. "Parameters" refer to the weights and biases learnt by the model during training. The design begins with input layer and proceeds through convolution, pooling, normalization, and flattening before ending with fully connected and output layers.

3.6.17. *Comparative analysis of model architectures*

Table 5 summarizes the parameters of different deep learning models employed in this study, including ZFNet, GoogLeNet, AlexNet, InceptionV3, InceptionResV2, SqueezeNet, ResNet50, ResNet101, VGG16, VGG19, ShuffleNet, Xception, DenseNet201, DarkNet19, DarkNet53, and the proposed DRNet13. Details such as the number of total layers, the number of learnable layers, the network size in megabytes (MB), the input image size, and the number of parameters is provided for each model. For example, ZFNet is an 8-layer model with all levels learnable. The network is 240 MB in size and needs a 224x224 pixel input image. The model has approximately 60,000,000 parameters. However, the proposed framework, DRNet13, contains 13 layers, 6 of which can be learned. The network is only 35 MB in size, which is far smaller than ZFNet. The required input image size is 224x224 pixels, which is similar to ZFNet. In contrast, DRNet13 has 206,417,249 parameters, which is a substantially bigger quantity. In general, the table enables a comparison of various models based on their complexity and organizational structure.

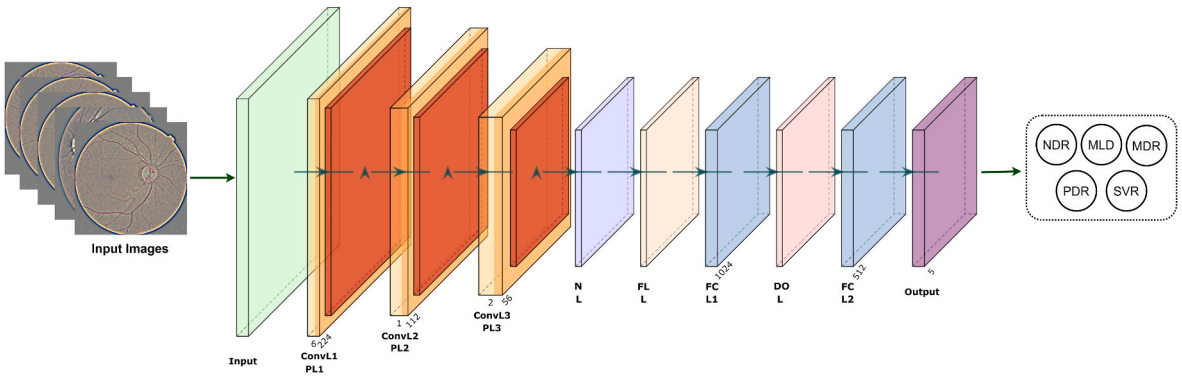

Fig. 6. This diagram depicts the structure of our novel DRNet13 model.

Table 4
Detailed construction of the DRNet13 model.

Layer (Type)	Stride & Padding	Output Shape	Karnel	Activation Function	Dropout	Parameters
Input Layer	–	224 × 224 × 3	–	–	–	0
Conv 1	Stride: 1, Padding: Same	224 × 224 × 64	3 × 3	ReLU	0.5	1792
Pool 1	Stride: 2, Padding: None	112 × 112 × 64	2 × 2	–	–	0
Conv 2	Stride: 1, Padding: Same	112 × 112 × 128	3 × 3	ReLU	0.5	73,856
Pool 2	Stride: 2, Padding: None	56 × 56 × 128	2 × 2	–	–	0
Conv 3	Stride: 1, Padding: Same	56 × 56 × 256	3 × 3	ReLU	0.5	295,168
Pool 3	Stride: 2, Padding: None	28 × 28 × 256	2 × 2	–	–	0
Normalization Layer	–	28 × 28 × 256	–	–	0.5	0
Flatten Layer	–	200704	–	–	0.5	0
FC Layer 1	–	1024	–	ReLU	0.5	205,521,92
Dropout Layer	–	1024	–	–	0.5	0
FC Layer 2	–	512	–	ReLU	0.5	524,800
Output Layer	–	5	–	softmax	–	513

Table 5
Comparing the efficiency of the deployed transfer learning models.

Network	No. of Layers	Learnable Layers	Network Size (MB)	Input Image Size	Parameters
ZFNet	8	8	240	224 × 224	60,000,000
GoogLeNet	144	22	27	224 × 224	7,000,000
AlexNet	25	8	227	227 × 227	61,000,000
InceptionV3	315	48	89	299 × 299	23,900,000
InceptionResV2	824	164	209	299 × 299	55,900,000
SqueezeNet	68	18	5.2	227 × 227	1,240,000
ResNet50	177	50	96	224 × 224	1,240,000
ResNet101	347	101	167	224 × 224	1,240,000
VGG16	41	16	515	224 × 224	20,000,000
VGG19	47	19	535	224 × 224	144,000,000
ShuffleNet	172	50	5.4	224 × 224	1,400,000
Xception	170	71	85	299 × 299	22,900,000
DenseNet201	708	201	77	224 × 224	20,000,000
DarkNet19	64	19	78	256 × 256	20,800,000
DarkNet53	184	53	155	256 × 256	41,600,000
Proposed Model (DRNet13)	**13**	**6**	**35**	**224×224**	**206,417,249**

Table 6
Comparing the efficiency of the deployed transfer learning models.

Network	T_Acc	V_Acc	Ts_Acc	T_Loss	V_Loss	Ts_Loss	Precision	F1–S	SEN	SPE	AUC	MSE	FPR	FNR
ZFNet	89	87	86	0.20	0.22	0.24	0.86	0.87	0.88	0.86	0.92	0.13	0.14	0.12
GoogLeNet	94	92	91	0.92	0.12	0.14	0.16	0.92	0.93	0.91	0.96	0.08	0.09	0.07
AlexNet	92	90	89	0.16	0.18	0.20	0.89	0.90	0.91	0.89	0.94	0.10	0.11	0.09
InceptionV3	96	94	93	0.10	0.12	0.14	0.94	0.94	0.95	0.93	0.97	0.06	0.07	0.05
InceptionResV2	93	91	90	0.15	0.17	0.19	0.90	0.91	0.92	0.90	0.95	0.09	0.10	0.08
SqueezeNet	83	81	80	0.27	0.29	0.31	0.80	0.80	0.81	0.82	0.80	0.85	0.19	0.18
ResNet50	87	85	84	0.23	0.25	0.27	0.85	0.85	0.86	0.84	0.89	0.15	0.16	0.14
ResNet101	91	89	88	0.17	0.19	0.21	0.89	0.89	0.90	0.88	0.93	0.11	0.12	0.10
VGG16	92	90	89	0.16	0.18	0.20	0.89	0.90	0.91	0.89	0.94	0.10	0.11	0.09
VGG19	94	92	91	0.12	0.14	0.16	0.92	0.92	0.93	0.91	0.96	0.08	0.09	0.07
ShuffleNet	88	86	85	0.22	0.24	0.26	0.86	0.86	0.87	0.85	0.90	0.14	0.15	0.13
Xception	93	91	90	0.15	0.17	0.19	0.91	0.91	0.92	0.90	0.95	0.09	0.10	0.08
DenseNet201	95	93	92	0.11	0.13	0.15	0.93	0.93	0.94	0.92	0.97	0.07	0.08	0.06
DarkNet19	94	92	91	0.12	0.14	0.16	0.92	0.92	0.93	0.91	0.96	0.08	0.09	0.07
DarkNet53	96	94	93	0.10	0.12	0.14	0.94	0.94	0.95	0.93	0.97	0.06	0.07	0.05
DRNet13	**99**	**97**	**96**	**0.07**	**0.10**	**0.12**	**0.97**	**0.97**	**0.98**	**0.96**	**0.99**	**0.03**	**0.04**	**0.02**

4. Experimental result

4.1. Environment installation

The studies were conducted on a computer system powered by an AMD Ryzen 7 3800 processor with a base clock speed of 3.90 GHz and 8 cores with a 32 MB L3 cache. The CPU also has 16 threads, which allows for effective multitasking. The machine was also upgraded with 64 GB of RAM to guarantee enough capacity for data processing. Google Collab was utilized as the software environment. The machine was outfitted with an MD Radeon RX 580 series GPU to maximize graphics processing capabilities. This GPU, which is well-known for its performance and efficiency, helped to speed up the calculations necessary for the tests. Using a strong CPU in conjunction with a dedicated GPU enabled efficient and rapid execution of the experimental activities. Among the fifteen pre-trained various models assessed, the unique technique known as DRNet13 displayed the greatest accuracy and was chosen as the best model. Python, a popular programming language for data analysis and machine learning, was used to conduct the study. To handle image processing tasks, the OpenCV package, which has a wide variety of computer vision methods, was used.

4.2. Performance analysis

In this section, extensive experiments are performed to evaluate the putative therapeutic. Within the context of this investigation, the dataset identified as fundus [37]. A total of 7500 fundus eye images of Diabetic retinopathy (DR) data were divided into 5 classes. These are gradually No DR, mild DR, moderate DR, proliferative DR and severe DR. Each class contains 1500 images. Each DR image was first put into one of five groups: healthy DR (DR1), mild DR (DR2, gentle DR with proliferative characteristics), moderate DR (RDR), or severe DR (SDR). The presence of diabetic macular edema and a severity level of mild, non-proliferative DR or above established a diagnosis of referable diabetic retinopathy (DME). After distinguishing it from RDR, the subtype of gentle DR without proliferative characteristics is called DR2.

In this dataset, some image data shapes were different. For this reason, all images are resized to ensure the same scale as required for different models (e.g., for InceptionV3 (299 × 299), DarkNet53 (256 × 256), and Proposed model (224 × 224)). After ensuring the same scale, we separated our dataset into three parts: training, validation, and testing. The ratios among training, validation, and testing are 60:20:20, respectively. At first, fifteen individual pre-train models were developed, verified, and put to the test. ZFNet, GoogLeNet, AlexNet, InceptionV3, InceptionResV2, SqueezeNet, ResNet50, ResNet101, VGG16, VGG19, ShuffleNet, Xception, DenseNet201, DarkNet19, and DarkNet53 are some of the pre-trained models available here. After assessing the effectiveness of existing models, we developed a CNN model named DarkNet13 that outperformed all other pre-trained models.

4.2.1. Evaluation metrics

Several metrics, such as accuracy, precision, F1 Score, recall, Specificity, MSE, FPR, and FNR, are used to assess the efficacy of the proposed approach. This research employs several performance evaluation criteria to ascertain if the proposed investigation would be fruitful. During the entire process, a confusion matrix is produced that may be used to rate the method's effectiveness according to several different indicators. Most approaches rely on the confusion matrix generated during testing for the identification task. Similar to predictability, specificity refers to the percentage of all negative data that was accurately predicted [43]. All the performance evaluations were calculated by using equations (16)–(23).

Accuracy: Accuracy means the percentage of the total accurate value based on the positive and negative classes.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{16}$$

Precision: Precision is the proportion of True Positives relative to the total number of positive samples.

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

F1-Score: The F1-score is an assessment matrix that takes the harmonic mean of both the precision and recall matrices to provide a single score that incorporates both of these metrics into a single measure.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{18}$$

Recall: Recall is also called sensitivity. Recall is the number of true positives divided by the total number of positive samples.

$$Recall \Big/ Sensitivity = \frac{TP}{TP + FN} \tag{19}$$

Specificity: Specificity in the field of medical diagnosis refers to a test's capacity to accurately identify persons who do not have the disease (*TN*). In other words, it is the proportion of properly detected negatives.

$$Specificity = \frac{TN}{TN + FP} \tag{20}$$

Mean Squared Error (MSE): MSE is a statistical metric that represents an average squared variance between actual (observed) and estimated values. It is one of the most often used metrics for assessing prediction error in tasks involving regression, as well as in occasional binary and multi-class classification problems. Here, n is the number of data, y_i represents observed values, $\hat{y}_i$ represents predicted values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{21}$$

False Positive Rate (FPR): This statistic calculates the percentage of true negatives (healthy patients) that are mistakenly recognized as positives (diseased). FPR is calculated as follows:

$$FPR = \frac{FP}{FP + TN} \tag{22}$$

False Negative Rate (FNR): This is a measure for calculating the percentage of true positives (diseased patients) that are misidentified as negatives (healthy). FNR is calculated as follows:

$$FNR = \frac{FN}{FN + TP} \tag{23}$$

in equations (3)–(10), *TP*, *TN*, *FP*, and *FN* mean true positive, true negative, false positive, and false negative, respectively. *TP* means the model predicts DR positive class correctly. And *TN* can predict the negative class successfully. *FP* is a negative class, but that predicts a positive. *FN* occurs when the actual image is DR, but the model predicts not DR.

4.2.2. Result evaluation

Table 6 compares numerous pre-existing and frequently used neural network models namely ZFNet, GoogLeNet, AlexNet, InceptionV3, InceptionResV2, SqueezeNet, ResNet50, ResNet101, VGG16, VGG19, ShuffleNet, Xception, DenseNet201, DarkNet19, DarkNet53, as well as a proposed new model (DRNet13). Each model is tested using a variety of performance indicators. Precision, F1-Score, Sensitivity, Specificity, Area Under Curve (AUC), Mean Squared Error (MSE), False Positive Rate (FPR), and False Negative Rate (FNR) are among these. Here, the training accuracy is labelled as 'T_Acc', validation accuracy labelled as 'V_Acc', test accuracy labelled as 'Ts_Acc', training loss is labelled as 'T_Loss', test loss labelled as 'Ts_loss', validation loss labelled as 'V_Loss', F1-score labelled as 'F1–S', Sensitivity labelled as 'SEN', and Specificity labelled as 'SPE'.

From Table 6 we can observe that our proposed model DRNet13, outperformed the second-best performance, DarkNet53, with 99 %, 97 %, and 96 % accuracies in Training, Validation, and Testing, respectively. This means DRNet13 can learn patterns from data more effectively and generalize better on previously unseen data. DRNet13 also had the lowest losses (0.07 for Training, 0.10 for Validation, and 0.12 for Test), indicating that it performed exceptionally well in terms of decreasing the difference between predicted and actual values. This is a huge improvement above InceptionV3 and DarkNet53, both of which have the second lowest losses (0.10 for Training, 0.12 for Validation, and 0.14 for Test). In terms of Precision, the DRNet13 model outperforms the competition with a value of 0.97, outperforming the next top performers, InceptionV3 and DarkNet53, both of which have Precision values of 0.94. Furthermore, the proposed approach has an F1-Score of 0.97, which is a harmonic mean of Precision and Recall, indicating a balanced performance on both measures.

DRNet13 has the maximum Sensitivity (0.98), surpassing both InceptionV3 and DarkNet53 (0.95). This indicates a superior ability to identify positive classes accurately. It also outperforms all other models in Specificity (0.96), demonstrating its ability to identify negative

classes accurately. The proposed model achieves the maximum AUC score of 0.99, surpassing InceptionV3, DenseNet201, and DarkNet53's scores of 0.97. A greater AUC indicates that the model's sensitivity and specificity are better balanced. Regarding MSE, DRNet13 has the lowest value (0.03), followed by InceptionV3 and DarkNet53 (0.06), indicating a model with fewer squared errors between predicted and actual values.

The DRNet13 model has the lowest False Positive Rate (FPR) and False Negative Rate (FNR), with values of 0.04 and 0.02, respectively. This indicates a lower misclassification rate, demonstrating the model's superior performance in correctly classifying the fundus image data. Fig. 7 illustrated the accuracy (training, validation) and loss (training, validation) of the DRNet13 model. It exhibits the entire learning process and the superiority of our proposed model.

Moreover, Our DRNet13 model exhibits superior performance across all critical metrics. The results demonstrate the model's extraordinary ability to effectively learn from data and generalize to unseen samples, suggesting that it will make accurate predictions with minimal error.

Fig. 8 offers a comprehensive comparative analysis of various transfer learning models, showcasing their performance across several categories of accuracy and loss metrics. The figure is organized in a grid layout, featuring two rows and three columns, with each column representing a different metric.

Training, validation and test accuracies of the models are presented in the top row. DRNet13 performed with remarkable accuracy scores of 99 % in training, 97 % in validation and 96 % in test accuracy. These numbers not only illustrate DRNet13's excellent learning and generalization abilities but also significantly contrast with other famous models like InceptionV3 (96 %, 94 %, 93 %), DenseNet201 (95 %, 93 %, 92 %), and GoogLeNet (94 %, 92 %, 91 %), which although impressive but fall short of DRNet13's performance. Similarly, the training, validation and test losses are detailed on the bottom row. DRNet13 has the lowest losses (0.07, 0.10, and 0.12, respectively), demonstrating its efficiency. Other high-performing models, such as DarkNet53 and InceptionV3 show slightly higher losses (0.10, 0.12 in training and test loss for both), indicating DRNet13's advantage in optimizing loss measures.

Overall, Fig. 8 shows that DRNet13 not only achieved the greatest accuracy but also the lowest loss across all assessment parameters. This significant performance differences particularly when compared to other high performing models, confirms DRNet13 as a standard in effective and dependable model performance within the scope of the research.

4.2.3. Confusion matrix analysis

In the framework of Diabetic Retinopathy (DR) classification utilizing the DRNet13 network, the confusion matrix demonstrates the ratio of the model's predictions to the actual class of each image in the test dataset, which includes a total of 1500 images evenly distributed across five DR categories: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR.

From Fig. 9 its observed that matrix is divided into rows for predicted classes and columns for Diabetic Retinopathy (DR) classes. "No DR" (0), "Mild DR" (1), "Moderate DR" (2), "Severe DR" (3), and "Proliferative DR" (4) are the phases of DR, with the designations "0″ to "4″ indicating the stages of DR. Each cell in the matrix represents the number of cases classified. Off-diagonal cells indicate misclassifications, whereas diagonal cells reflect properly classified cases. For example, the value in the column labelled "1" (Mild DR predicted) and the row labelled "0" (No DR real) shows the number of "No DR" instances misclassified as "Mild DR." In contrast, the cell under "0" (No DR predicted) and the row "0" (No DR actual) indicate correctly classified "No DR" cases.

The confusion matrices created by our different models show the number of correctly diagnosed and misclassified cases throughout five phases of Diabetic Retinopathy (DR), which include "No DR," "Mild DR," "Moderate DR," "Severe DR," and "Proliferative DR." The ZFNet model performed well by correctly classifying 277 instances of "No DR," but it misclassified 11 instances as "Mild DR." GoogLeNet, on the other hand, classified 269 instances of "Severe DR" accurately but wrongly classified 9 instances as "Moderate DR." AlexNet found equally distributed misclassifications across all DR phases, including 12 cases of "No DR" being misclassified as "Mild DR." In contrast, InceptionV3 produced

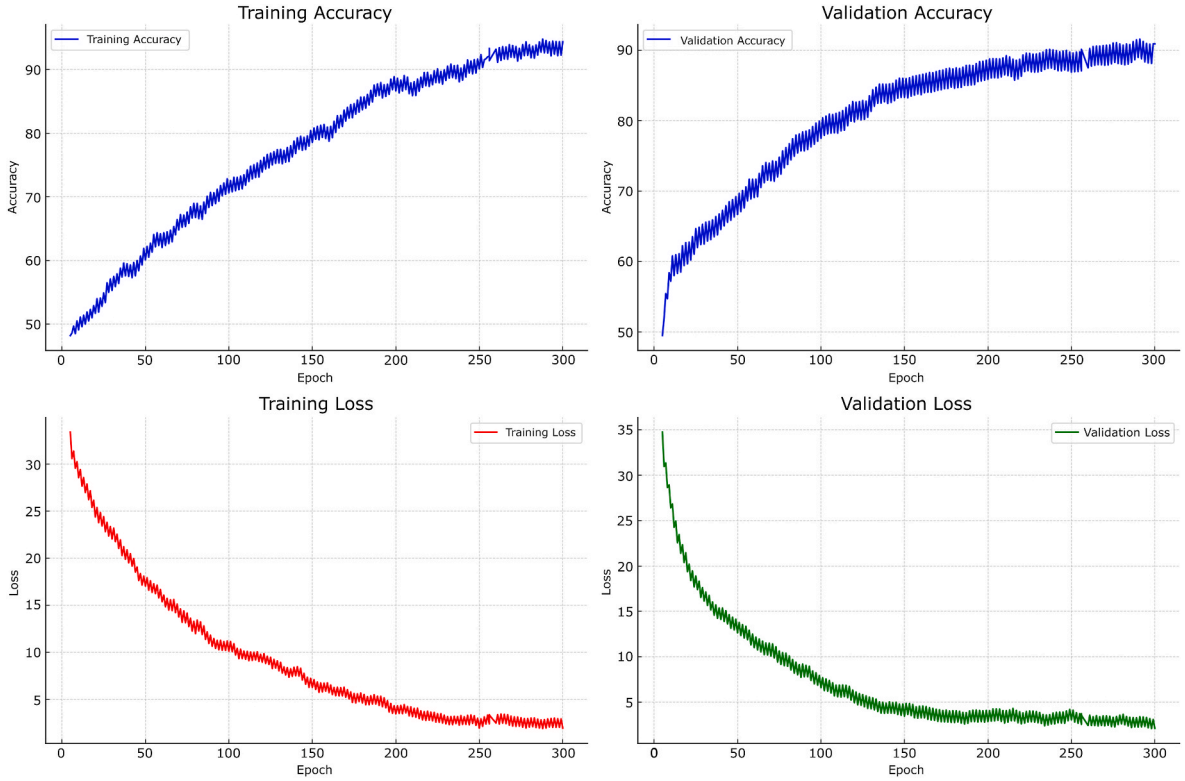

Fig. 7. DRNet13 network accuracy and loss performance over 300 epochs.

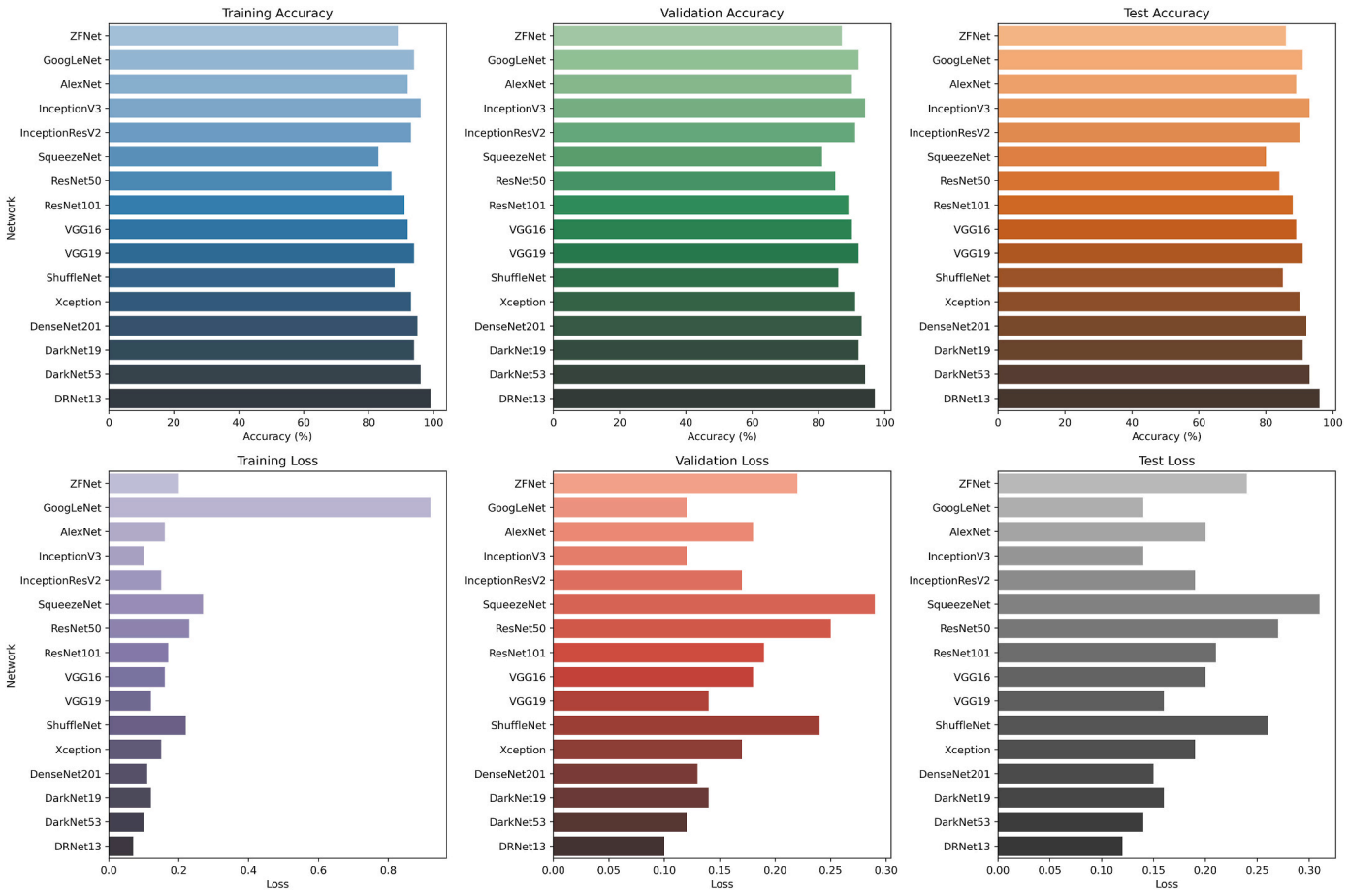

Fig. 8. Performance evaluation for all applied models for diabetic retinopathy detection.

remarkable results, properly classifying 270 instances of "Moderate DR" and misclassifying just 6 instances of "Moderate DR" as "Mild DR." InceptionResV2 has a high level of accuracy, identifying 275 occurrences of "No DR" correctly, but it has misclassified 7 examples as "Mild DR." SqueezeNet correctly classified 269 cases of "Moderate DR" while misclassifying 9 instances as "Mild DR."

ResNet50 and ResNet101 models demonstrated remarkable accuracy. ResNet50 correctly identified 274 cases of "Moderate DR" while incorrectly classifying 6 instances as "Mild DR." ResNet101 performed well in the "No DR" stage, properly classifying 280 cases and misclassifying 5 as "Mild DR." Both the VGG16 and VGG19 models performed well, with the VGG16 correctly identifying 270 occurrences of "Proliferative DR" and misclassifying 9 as "Severe DR." VGG19 scored thoroughly, properly identifying 279 cases of "Proliferative DR" and misclassifying 6 as "Severe DR." ShuffleNet properly classified 256 cases of "Moderate DR" and incorrectly classed 11 instances as "Mild DR." Xception conducted well across all classes, effectively classifying 271 cases of "Moderate DR" and misclassifying 7 as "Mild DR." DenseNet201 worked exceptionally well for the "No DR" and "Proliferative DR" phases, correctly classifying 280 cases in each case and misclassifying 5 as "Mild DR" for "No DR" and 3 as "Severe DR" for "Proliferative DR."

DarkNet19 and DarkNet53 models both demonstrated great accuracy for "Moderate DR" and "No DR" phases. DarkNet19 accurately diagnosed 271 cases of "Moderate DR" and incorrectly identified 7 instances of "Mild DR." DarkNet53, on the other hand, correctly recognized 278 cases of "No DR" and incorrectly identified 5 cases of "Mild DR." DRNet13 performed the best, accurately classifying 288 cases of "No DR" and "Severe DR," misclassifying just 3 and 1, respectively.

4.2.4. *ROC curve analysis of DRNet13*

The performance of our proposed DRNet13 model was assessed using the Receiver Operating Characteristics (ROC) curve and Area Under Curve (AUC) values, which are common metrics for measuring the accuracy of multiclass classification model [55]. The ROC curve plots sensitivity (True Positive Rate) on y-axis against 1-Specificity (False Positive Rate) on the x-axis, providing a graphical representation of a model's diagnostic ability. AUC scores are calculated as the area under the ROC curve with a value ranging from 0 to 1. A higher AUC value indicates a model's improved ability to accurately differentiate across classes.

We compute separate ROC curves for each class in our DRNet13 model, which defines different stages of diabetic retinopathy (DR). This method assesses the model's accuracy in identifying each unique condition by dividing each class and comparing it to all others. The following are the AUC values found in our study for each DR stages: 'No DR' has an AUC of 0.983, indicating excellent model performance for identifying cases without DR. With an AUC of 0.987 for the 'Mild' stage of DR, the model seems to be nearly perfect in identifying the early signs of DR. The AUC values for the 'Moderate' and 'Severe' stages are 0.991 and 0.993 respectively, indicating exceptional model performance for these model development stages. The 'Proliferate DR' stages has an AUC of 0.986, again indicating high accuracy in identifying these advanced stages of disease. Fig. 10 portraying the ROC curve of the proposed model DRNet13.

4.3. *Feature map analysis of DRNet13*

The model comprises of three convolutional layer blocks and a pair of thick layers. Every convolutional block incorporates a MaxPooling

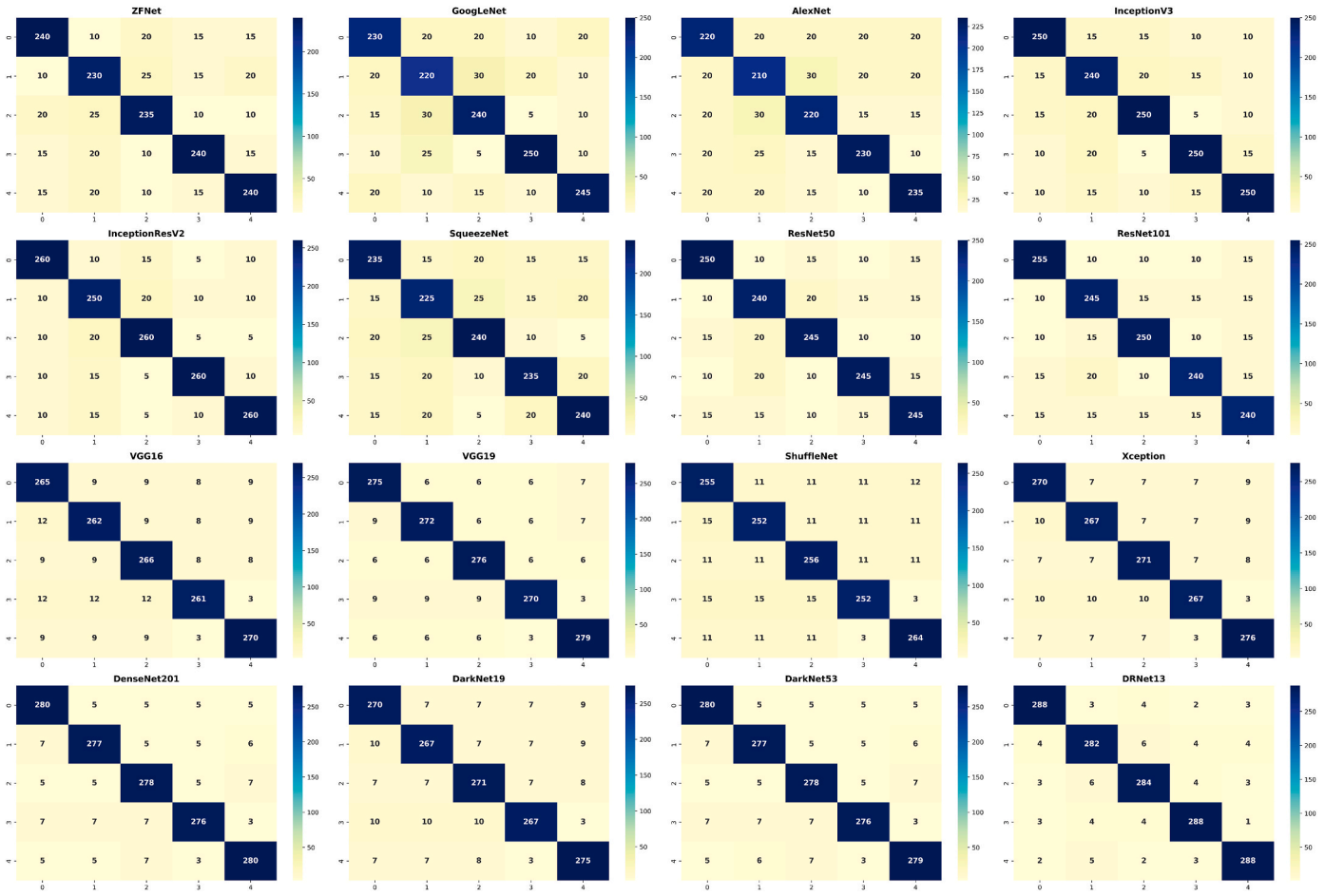

Fig. 9. Confusion matrix for the predictions of the deep learning networks on the test set. Rows represent actual classes, while columns represent predicted classes. The diagonal cells represent instances that were correctly classified. The intensity of color represents the number of occurrences. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

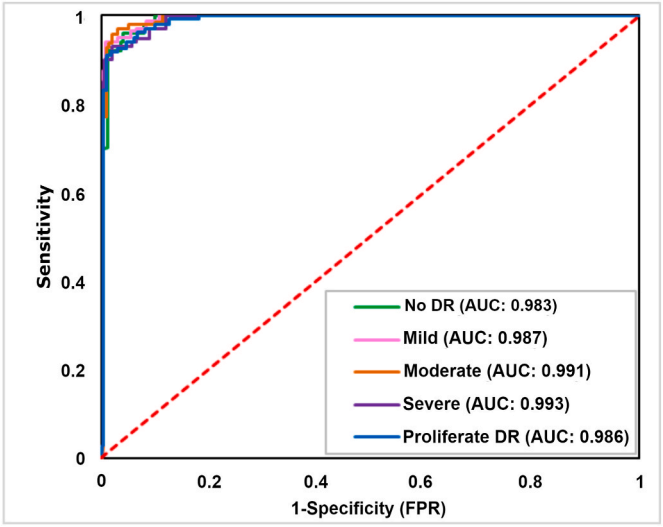

Fig. 10. ROC curve and AUC metrics for DRNet13 model.

layer, which decreases the input's dimensionality. To speed up the learning process, a Batch Normalization layer is implemented. In this model DRNet13, each Conv2D layer produces feature maps as its output. Each feature map reflects the application of a learned filter to the input data. Essentially, the learned filters are what the model learned to detect throughout the training process. The Conv2D layers of the DRNet13 convolutional neural network generate feature maps that indicate the features that the model learned to identify at each stage. In the fundus images, the model typically learns to detect low-level features such as color gradients, boundaries, and fundamental textures in the initial layers with fewer filters. As we proceed deeper into the structure of the network, to layers that include additional filters, these fundamental patterns are combined to identify more complex characteristics. The final layers of the model derive even more abstract characteristics that, despite being difficult for humans to interpret, are crucial for the model to make precise predictions. The DRNet13 feature map is depicted in Fig. 11.

Visualizing these feature maps enables us to comprehend the learning process of the model DRNet13 by providing insight into the categories of features that the model considers significant at each iteration. This is essential for comprehending the model's functions, diagnosing prospective issues, and suggesting enhancements. Adjustments could be made to the model if essential diabetic retinopathy-related characteristics, such as microaneurysms, hemorrhages, or exudates, are not adequately detected. In addition, feature maps provide an interpretive view of the DRNet13 model, illustrating what the neural network is learning and how it makes decisions. This is particularly crucial for medical diagnostic applications, such as diabetic retinopathy detection, where comprehending the model's decision-making process is of the utmost importance.

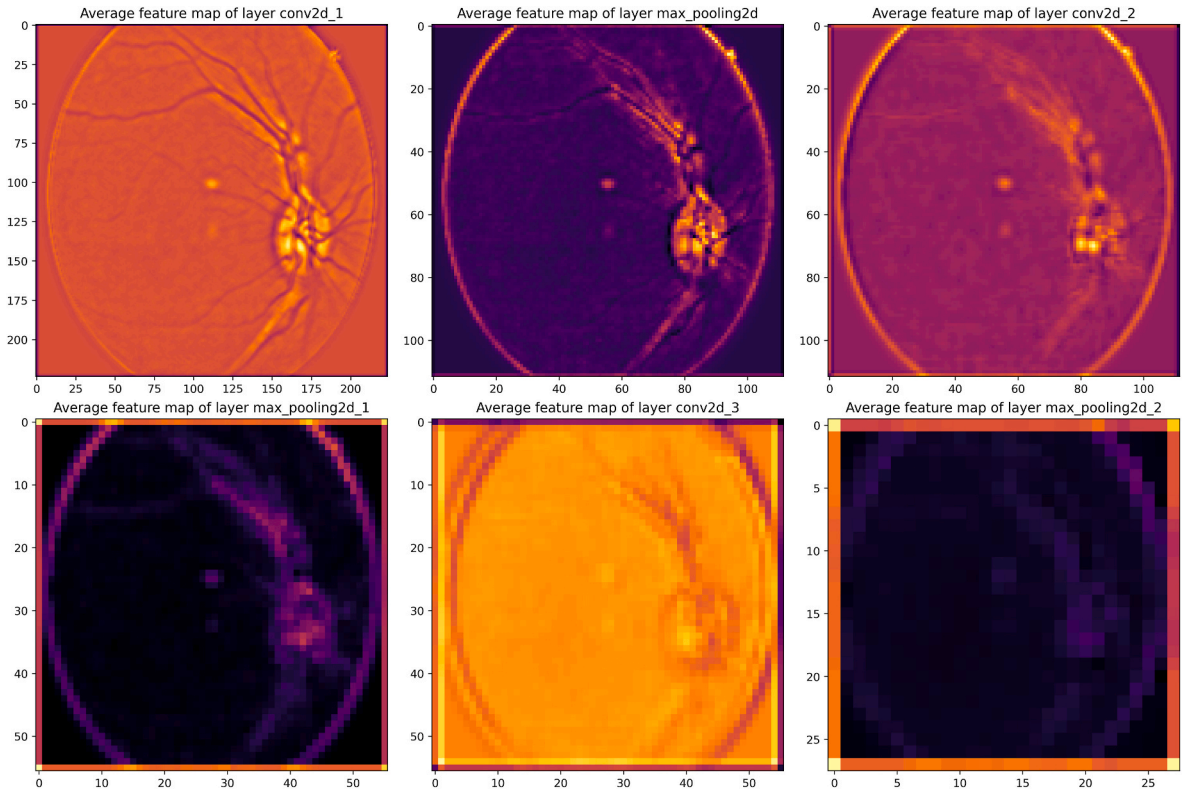

Fig. 11. Displaying the feature map as interpreted by DRNet13 model. Average feature maps from various convolutional and pooling layers of the model. The colors in each feature map represent the intensity of the activation, with cooler colors (blues and purples) indicating lower activations and warmer colors (yellow and reds) indicating higher activations. This color encoding is applied consistently across all layers to facilitate comparison of feature intensity and localization. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

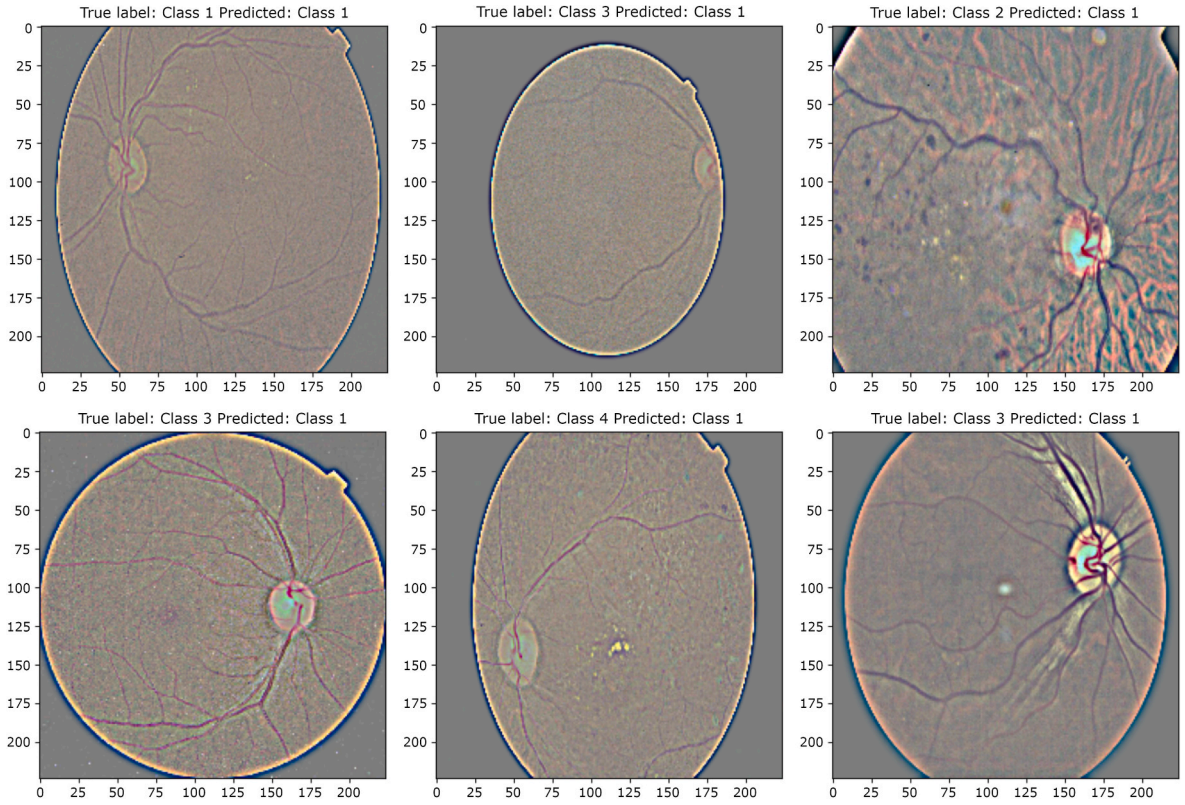

Fig. 12. Visualization of misclassification outcomes derived from the proposed network DRNet13.Here, 0: 'Class 1(No DR)', 1: 'Class 2(Mild)', 2: 'Class 3(Moderate), 3: 'Class 4(Severe)', 4: 'Class 5(Proliferate DR)'.

4.4. Misclassification results analysis

The highest performing proposed model, DRNet13, displayed an outstanding level of accuracy in classifying the phases of diabetic retinopathy, however some cases of misclassification were identified. Fig. 12 depicts these instances. The misclassification is most likely due to imaging similarities across the five types of diabetic retinopathy: No DR, mild, moderate, severe, and proliferative DR. Because of overlapping symptoms, the early phases (No DR and Mild) and later phases (Severe and Proliferative) are especially difficult to discern.

4.5. Comparative compilation complexity and performance analysis

Various deep learning models were compared in the study based on their learning rate, epoch time in seconds, and overall execution time in minutes. Table 7 depicts the comparative study of all model's computation time. The ZFNet model was trained at a learning rate of 0.0001, 54 s per epoch, and a total execution time of 270 min. GoogLeNet, which was likewise trained with a learning rate of 0.0001, took 42 s for each epoch and had a total execution time of 210 min.

In a similar way, AlexNet was trained using a learning rate of 0.0001. Each phase took 48 s for a total execution duration of 240 min. For the InceptionV3 model, a reduced learning rate of 0.00001 was chosen. Each epoch took 36 s, with a total execution duration of 180 min, which is faster than the prior models. The InceptionResV2 and ResNet50 networks were trained at a learning rate of 0.0001. However, while InceptionResV2 took 40 s every epoch (for a total of 200 min), ResNet50 took 51 s each epoch and a total of 255 min. SqueezeNet and ShuffleNet networks used a higher learning rate of 0.001. SqueezeNet had the longest epoch and total execution times, at 60 s and 300 min, respectively. ShuffleNet, on the other hand, required 52 s for every epoch and took 260 min to execute. With a learning rate of 0.0001, models such as ResNet101, VGG16, VGG19, Xception, DenseNet201, DarkNet19, and DarkNet53 had all been trained. These models' execution timeframes varied from 180 min (DarkNet53 and InceptionV3) to 240 min (ResNet101, AlexNet, and DarkNet19).

DRNet13, our proposed model, was the most efficient. With a learning rate of 0.00001, each epoch took 30 s, and the overall execution time was 150 min. This was the quickest time among all models tested. This detailed comparison offers a clear grasp of the execution efficiency of several deep learning models as well as the influence of learning rates, emphasizing the potential of our proposed model, DRNet13.

Furthermore, A comparison of the DRNet13 model's performance with previously used pre-trained and state-of-art methods within the context of diabetic retinopathy detection demonstrates the proposed model's improved efficiency. To appropriately assess DRNet13's effectiveness, it was compared to various different models that have been extensively employed in previous studies. To establish a realistic and unbiased comparison, these comparable models were carefully chosen in accordance with their applicability in diabetic retinopathy diagnosis. Table 8 shows that the DRNet13 model surpasses existing established models in terms of classification accuracy. This comparison is not only numerical; it demonstrates the proposed model's relative strength in solving the difficult issue of recognizing and classifying the phases of diabetic retinopathy.

The DRNet13 model's superior performance score results from its enhanced architecture and methodological advances incorporated into its design. As the comparison shows, this model enhances the state-of-the-art in diabetic retinopathy identification and establishes a new performance standard in the area. The DRNet13 model's enhanced accuracy has huge potential consequences, moving research in automated retinopathy diagnosis ahead and providing novel possibilities for additional investigation and development.

Table 7
Computation time for all employed models.

Models	Learning Rate	Epoch × Seconds	Execution Time (mins)
ZFNet	0.0001	300 × 54	270
GoogLeNet	0.0001	300 × 42	210
AlexNet	0.0001	300 × 48	240
InceptionV3	0.00001	300 × 36	180
InceptionResV2	0.0001	300 × 40	200
SqueezeNet	0.001	300 × 60	300
ResNet50	0.0001	300 × 51	255
ResNet101	0.0001	300 × 48	240
VGG16	0.0001	300 × 46	230
VGG19	0.0001	300 × 44	220
ShuffleNet	0.001	300 × 52	260
Xception	0.0001	300 × 41	205
DenseNet201	0.0001	300 × 39	192
DarkNet19	0.0001	300 × 48	240
DarkNet53	0.0001	300 × 36	180
Proposed Model (DRNet13)	**0.00001**	**300×30**	**150**

5. Summary and conclusions

In our extensive examination, the DRNet13 model achieved a significant milestone in the automated diagnosis and staging of Diabetic Retinopathy (DR). With a phenomenal 97 % accuracy rate, the model establishes a new standard for both efficiency and speed, potentially transforming the diagnosis procedure for this vision-threating disease. Upon critical reflection, the adoption of DRNet13 could significantly improve the clinical workflows, offering rapid and reliable diagnostics that are paramount for early detection. However, we recognize that our study's scope though extensive but bears limitations such as the dataset's representativeness, which may influence the model's universal application. Our findings have implications for ophthalmic healthcare that go beyond immediate clinical application and point to a paradigm change in support of more independent, accurate and patient-centric methods. Furthermore, the model's applicability in environments with limited resources may democratize access to essential diagnostic services, thereby closing gaps in healthcare equity. Looking forward further research will concentrate on improving the robustness of the DRNet13 model using diverse data covers a wide range of DR phases and patient demographic.

Author contributions statement

F.M.J.M., K.A., F.M.B. and M.A.M., provided the idea and designed the experiments; F.M.J.M., R.S., S.S., B.A., and K.A. analyzed the data and wrote the manuscript. F.M.J.M., M.Z.A., N.H.O., and M.A.M. helped perform the experimental analysis with constructive discussions. K.A., F. M.B. and M.A.M., reviewed the work. All authors discussed the results and contributed to the manuscript.

Table 8
An evaluation of the proposed approach with several existing studies.

Authors & Reference	Class	Methodology	Accuracy
Li et al. [22]	6	SE-BN-Inception	82.84 %
Minarno et al. [24]	5	EfficientNet-B7	84.36 %
Dayana et al. [25]	4	UNet	95.9 %
Alghamdi et al. [26]	2	VGG-16	73.04 %
Nasir et al. [28]	2	CNN	96 %
Arias-Serrano et al. [29]	2	AlexNet	94.3 %
Tian et al. [30]	5	FA + KC-Net	86.78 %
Mustafa et al. [31]	2	ResNet-50 + DenseNet-121	95.58 %
Atwany et al. [32]	5	Green-SE-ResNext50	85.7 %
Nahiduzzaman et al. [33]	5	ELM	97.27 %
K. Oh et al. [34]	2	ETDRS 7SF	83.38 %
Jian et al. [36]	5	Triple-DRNet	92.08 %
Shamrat et al.	**5**	**DRNet13**	**97 %**

Funding

This work was supported in part by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Ethics approval

Not Applicable.

Consent to participate

Not Applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] S. Jan, I. Ahmad, S. Karim, Z. Hussain, M. Rehman, M.A. Shah, Status of diabetic retinopathy and its presentation patterns in diabetics at ophthalmogy clinics, J. Postgrad. Med. Inst. 32 (1) (2018).
[2] L. Andersen, P. Andersson, Deep Learning Approach for Diabetic Retinopathy Grading with Transfer Learning, 2020.
[3] L. Math, R. Fatima, Adaptive machine learning classification for diabetic retinopathy, Multimed. Tool. Appl. 80 (4) (2021) 5173–5186.
[4] A. He, T. Li, N. Li, K. Wang, H. Fu, Cabnet: Category attention block for imbalanced diabetic retinopathy grading, IEEE Trans. Med. Imag. 40 (1) (2020) 143–153.
[5] W.X. Lim, Z. Chen, A. Ahmed, The adoption of deep learning interpretability techniques on diabetic retinopathy analysis: a review, Med. Biol. Eng. Comput. 60 (3) (2022) 633–642.
[6] Y. Zheng, M. He, N. Congdon, The worldwide epidemic of diabetic retinopathy, Indian J. Ophthalmol. 60 (5) (2012) 428.
[7] Y. Bai, X. Zhang, C. Wang, H. Gu, M. Zhao, F. Shi, Microaneurysms detection in retinal fundus images based on shape constraint with region-context features, Biomed. Signal Process Control 85 (2023) 104903.
[8] F.L. Ferris, M.D. Davis, L.M. Aiello, Treatment of diabetic retinopathy, N. Engl. J. Med. 341 (9) (1999) 667–678.
[9] R.E. Putra, H. Tjandrasa, N. Suciati, Severity classification of non-proliferative diabetic retinopathy using convolutional support vector machine, International Journal of Intelligent Engineering & Systems 13 (4) (2020).
[10] D. Das, S.K. Biswas, S. Bandyopadhyay, A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning, Multimed. Tool. Appl. 81 (18) (2022) 25613–25655.
[11] M. Zipperle, F. Gottwalt, E. Chang, T. Dillon, Provenance-based intrusion detection systems: a survey, ACM Comput. Surv. 55 (7) (2022) 1–36.
[12] G. Selvachandran, S.G. Quek, R. Paramesran, W. Ding, L.H. Son, Developments in the detection of diabetic retinopathy: a state-of-the-art review of computer-aided diagnosis and machine learning methods, Artif. Intell. Rev. 56 (2) (2023) 915–964.
[13] D. Das, S.K. Biswas, S. Bandyopadhyay, A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning, Multimed. Tool. Appl. 81 (18) (2022) 25613–25655.
[14] M. Nahiduzzaman, M.R. Islam, M.O.F. Goni, M.S. Anower, M. Ahsan, J. Haider, M. Kowalski, Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier, Expert Syst. Appl. 217 (2023) 119557.
[15] K. Gunasekaran, R. Pitchai, G.K. Chaitanya, D. Selvaraj, S. Annie Sheryl, H. S. Almoallim, B.G. Tesemma, A deep learning framework for earlier prediction of diabetic retinopathy from fundus photographs, BioMed Res. Int. (2022), 2022.
[16] V. Narayan, P.K. Mall, A. Alkhayyat, K. Abhishek, S. Kumar, P. Pandey, Enhance-net: an approach to boost the performance of deep learning model based on real-time medical images, J. Sens. (2023), 2023.
[17] N. Sengar, R.C. Joshi, M.K. Dutta, R. Burget, EyeDeep-Net: a multi-class diagnosis of retinal diseases using deep neural network, Neural Comput. Appl. (2023) 1–21.
[18] A. Sutradhar, M. Al Rafi, P. Ghosh, F.J.M. Shamrat, M. Moniruzzaman, K. Ahmed, M.A. Moni, An Intelligent Thyroid Diagnosis System Utilising Multiple Ensemble and Explainable Algorithms with Medical Supported Attributes, IEEE Transactions on Artificial Intelligence, 2023.
[19] R. Thanki, A deep neural network and machine learning approach for retinal fundus image classification, Healthcare Analytics 3 (2023) 100140.
[20] M.S.B. Phridviraj, R. Bhukya, S. Madugula, A. Manjula, S. Vodithala, M.S. Waseem, A bi-directional Long Short-Term Memory-based Diabetic Retinopathy detection model using retinal fundus images, Healthcare Analytics 3 (2023) 100174.
[21] P. Saranya, R. Pranati, S.S. Patro, Detection and classification of red lesions from retinal images for diabetic retinopathy detection using deep learning models, Multimed. Tool. Appl. (2023) 1–21.
[22] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, H. Kang, Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screening, vol. 501, Information Sciences, 2019, pp. 511–522.
[23] A.B. Aujih, M.I. Shapiai, F. Meriaudeau, T.B. Tang, EDR-Net: Lightweight deep neural network architecture for detecting referable diabetic retinopathy, IEEE Transactions on Biomedical Circuits and Systems 16 (3) (2022) 467–478.
[24] A.E. Minarno, M.H.C. Mandiri, Y. Azhar, F. Bimantoro, H.A. Nugroho, Z. Ibrahim, Classification of diabetic retinopathy disease using convolutional neural network, JOIV: International Journal on Informatics Visualization 6 (1) (2022) 12–18.
[25] A.M. Dayana, W.S. Emmanuel, An enhanced swarm optimization-based deep neural network for diabetic retinopathy classification in fundus images, Multimed. Tool. Appl. 81 (15) (2022) 20611–20642.
[26] H.S. Alghamdi, Towards explainable deep neural networks for the automatic detection of diabetic retinopathy, Appl. Sci. 12 (19) (2022) 9435.
[27] B. Tymchenko, P. Marchenko, D. Spodarets, Deep Learning Approach to Diabetic Retinopathy Detection, 2020 arXiv preprint arXiv:2003.02261.
[28] N. Nasir, P. Oswald, O. Alshaltone, F. Barneih, M. Al Shabi, A. Al-Shammaa, Deep DR: detection of diabetic retinopathy using a convolutional neural network, in: 2022 Advances in Science and Engineering Technology International Conferences (ASET), IEEE, 2022, February, pp. 1–5.
[29] I. Arias-Serrano, P.A. Velásquez-López, L.N. Avila-Briones, F.C. Laurido-Mora, F. Villalba-Meneses, A. Tirado-Espin, D. Almeida-Galárraga, Artificial intelligence based glaucoma and diabetic retinopathy detection using MATLAB—retrained AlexNet convolutional neural network, F1000Research 12 (2023) 14.
[30] M. Tian, H. Wang, Y. Sun, S. Wu, Q. Tang, M. Zhang, Fine-Grained Attention & Knowledge-Based Collaborative Network for Diabetic Retinopathy Grading, 2023. Heliyon.
[31] H. Mustafa, S.F. Ali, M. Bilal, M.S. Hanif, Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework, IEEE Access 10 (2022) 113172–113183.
[32] M.Z. Atwany, A.H. Sahyoun, M. Yaqub, Deep learning techniques for diabetic retinopathy classification: a survey, IEEE Access 10 (2022) 28642–28655.
[33] M. Nahiduzzaman, M.R. Islam, M.O.F. Goni, M.S. Anower, M. Ahsan, J. Haider, M. Kowalski, Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and elm classifier, Expert Syst. Appl. 217 (2023) 119557.
[34] K. Oh, H.M. Kang, D. Leem, H. Lee, K.Y. Seo, S. Yoon, Early detection of diabetic retinopathy based on deep learning and ultra-widefield fundus images, Sci. Rep. 11 (1) (2021), 1897.
[35] J. Jaskari, J. Sahlsten, T. Damoulas, J. Knoblauch, S. Särkkä, L. Kärkkäinen, K. K. Kaski, Uncertainty-aware deep learning methods for robust diabetic retinopathy classification, IEEE Access 10 (2022) 76669–76681.
[36] M. Jian, H. Chen, C. Tao, X. Li, G. Wang, Triple-DRNet: a triple-cascade convolution neural network for diabetic retinopathy grading using fundus images, Comput. Biol. Med. 155 (2023) 106631.
[37] "Diabetic retinopathy dataset, available link: https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered.
[38] W. Gouda, N.U. Sama, G. Al-Waakid, M. Humayun, N.Z. Jhanjhi, Detection of skin cancer based on skin lesion images using deep learning, Healthcare 10 (2022) 1183. MDPI.
[39] R. Tharsanee, R. Soundariya, A.S. Kumar, M. Karthiga, S. Sountharrajan, Deep convolutional neural network–based image classification for covid-19 diagnosis, Data Science for COVID- 19 (2021) 117–145. Elsevier.
[40] F.J.M. Shamrat, S. Azam, A. Karim, R. Islam, Z. Tasnim, P. Ghosh, F. De Boer, Lungnet22: a fine-tuned model for multiclass classification and prediction of lung disease using x-ray images, J. Personalized Med. 12 (5) (2022) 680.
[41] S. Akter, F.J.M. Shamrat, S. Chakraborty, A. Karim, S. Azam, Covid-19 detection using deep learning algorithm on chest x-ray images, Biology 10 (11) (2021) 1174.
[42] L. Fu, Y. Feng, Y. Majeed, X. Zhang, J. Zhang, M. Karkee, Q. Zhang, Kiwifruit detection in field images using faster r-cnn with zfnet, IFAC-PapersOnLine 51 (17) (2018) 45–50.
[43] Z. Yu, Y. Dong, J. Cheng, M. Sun, F. Su, Research on Face Recognition Classification Based on Improved Googlenet, Security and Communication Networks, 2022, pp. 1–6, 2022.
[44] Z. Wu, S. He, Improvement of the alexnet networks for large-scale recognition applications, Iranian Journal of Science and Technology, Transactions of Electrical Engineering 45 (2021) 493–503.
[45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[46] S.R. Bose, V.S. Kumar, Efficient inception v2 based deep convolutional neural network for real-time hand action recognition, IET Image Process. 14 (4) (2020) 688–696.
[47] B. Koonce, B. Koonce, SqueezeNet, Springer, 2021.
[48] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[49] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.

[50] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: an extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848–6856.
[51] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
[52] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[53] M. Baygin, O. Yaman, P.D. Barua, S. Dogan, T. Tuncer, U.R. Acharya, Exemplar darknet19 feature generation technique for automated kidney stone detection with coronal ct images, Artif. Intell. Med. 127 (2022) 102274.
[54] H. Wang, F. Zhang, L. Wang, Fruit classification model based on improved darknet53 convolutional neural network. International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), IEEE, 2020, pp. 881–884, 2020.
[55] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (8) (2006) 861–874.
[56] H. Lin, X. Zheng, L. Li, F. Chao, S. Wang, Y. Wang, R. Ji, Meta architecture for point cloud analysis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 17682–17691.