International Journal on Robotics, Automation and Sciences

DRD-Net: Diabetic Retinopathy Diagnosis Using A Hybrid Convolutional Neural Network

Muhammad Hassaan Ashraf*, Muhammad Esham Qureshi, Ahmed Khan, Jawaid Iqbal and Musharif Ahmed

Abstract – Diabetic Retinopathy (DR) has become a leading cause of blindness among diabetic patients. Accurate and timely diagnosis of DR is critical to slowing disease progression. This research proposes a Hybrid Convolutional Neural Network (CNN)-based model, named Diabetic Retinopathy Detection Network (DRD-Net). The proposed DRD-Net designed to enhance diagnostic accuracy by addressing key challenges such as gradient vanishing and lesion scale variability in fundus images. Contrast-Limited Adaptive Histogram Equalization (CLAHE) was used to enhance contrast and highlight lesions in fundus images. To increase the diversity of training samples, the proposed framework employs geometric data augmentation techniques. DRD-Net incorporates the Swish activation function along with densely connected blocks to mitigate gradient vanishing and enhancing feature propagation within the network. Additionally, the model integrates two Inception blocks to facilitate multiscale feature extraction, which is essential for detecting small Regions of Interest (RoI) in fundus images. Experimental results demonstrate that DRD-Net achieves a precision of 84.4%, recall of 84.5%, F1-score of 84.1%, and accuracy of 85.1%, outperforming several state-of-the-art models on the IDRiD dataset. These results highlight DRD-Net's potential as an effective solution for automated DR diagnosis, contributing to more efficient and accurate DR screening.

Keywords— Convolutional Neural Networks, Diabetic Retinopathy, Fundus Images, Multiscale Features, Multilevel Features, Swish Activation Function.

I. INTRODUCTION

Chronic Diabetic Mellitus (DM) can develop in DR, a microvascular problem that can damage the retina and impair the vision, possibly leading to blindness. The blood arteries around the retina are gradually harmed by high blood sugar levels in DR, a disorder that affects the eyes and progresses over time. In the last 40 years, the number of adults with DM has increased five-fold globally, from 108 million in 1980 to 537 million in 2021 [1]. The countries with the highest percentage of diabetics in 2021, according to the International Diabetes Federation (IDF), were China with 140.9 million people affected by diabetics, India with 74.2 million, Pakistan with 33.0 million, and the United States with 32.2 million [2]. Type-1 diabetes (T1D) and Type-2 diabetes (T2D) patients, respectively, have DR in 77.3% and 25.1% of cases,

*Corresponding Author email: hassaan.ashraf@riphah.edu.pk, ORCID: 0000-0002-2178-0523

Muhammad Hassaan Ashraf is with Faculty of Computing, Riphah International University, Islamabad, 46000, Pakistan.

Muhammad Esham Qureshi is with Faculty of Computing, Riphah International University, Islamabad, 46000, Pakistan.
(e-mail: 36511@students.riphah.edu.pk), ORCID: 0009-0001-4503-9334

Ahmed Khan is with Faculty of Computing, Riphah International University, Islamabad, 46000, Pakistan.
(e-mail: 37635@students.riphah.edu.pk), ORCID: 0009-0006-8389-2017

Jawaid Iqbal is with Faculty of Computing, Riphah International University, Islamabad, 46000, Pakistan.
(e-mail: jawaid.iqbal@riphah.edu.pk), ORCID: 0000-0002-5045-7485

Musharif Ahmed is with Faculty of Computing, Riphah International University, Islamabad, 46000, Pakistan.
(e-mail musharraf.ahmed@riphah.edu.pk) , ORCID: 0000-0003-0369-1743

and in 25 to 30% of these cases, diabetic macular edema poses a threat to vision [3]. If timely treatment is neglected, those who are still in the early stages of the disease and everyone with some degree of visual impairment brought on by DR risk losing their vision. Regular screening is essential to identify DR early and enable action to arrest progressive degeneration using the right medications.

DR typically progresses through a series of distinct stages, from minor non-proliferative problems with the increased vascular permeability to non-proliferative diabetic retinopathy (NPDR) problems with vascular closure to PDR with retinal and posterior vitreous surface neovascularization [4]. There are four further different levels of severity of NP-DR: stage zero, stage mild, stage moderate, and stage severe [5]. A Microaneurysm (MA), a microscopic, dark red lesion that appears close to the end of a blood vessel, is one of the typical DR symptoms, as seen in Figure 1. Retinal vein obstruction, which can resemble MAs, and retinal Hemorrhages (HMs), which are caused by hypertension, are other DR-related consequences. Exudates, which are yellow deposits made of protein and lipid residues, develop when capillaries are damaged. In the later stages of DR, it becomes even more challenging to cure. The mild stage of NPDR only exhibits a small number of microaneurysms. In contrast, the moderate stage of NPDR is distinguished by a great deal of MAs, HMs, and venous beading, all of which impair blood flow to the retina. Along with the previously stated symptoms, the PDR stage also results in the creation of new blood vessels (i.e., neovascularization) [6].

DR can be effectively diagnosed using fundus camera, a noninvasive tools in ophthalmology. That captures detailed images of eye structures, such as the retinal blood vessels, optic nerve head, retina, choroid, vitreous, and macula [7]. Fundus photography allows for in-depth evaluations of the eye, enabling timely diagnoses and effective treatment planning, ultimately improving patient outcomes and preserving vision.

Diagnosing DR from color fundus images requires clinicians to recognize subtle features, a complex process due to the intricate grading system.

Automated systems for DR detection have evolved into traditional hand-crafted feature-based methods and modern deep learning approaches. Different classical computer vision and machine learning methods, such as Speeded-Up Robust Features (SURF) and Scale-Invariant Feature Transform (SIFT) which help in identifying DR, with classifiers such as Naive Bayes and Support Vector Machines (SVM) determining its presence. However, these methods face challenges like domain dependence, which limits generalization across different imaging conditions, and difficulties with high-dimensional spaces that can drag down detection performance [8].

Based on the limitations of traditional feature extraction methods, deep learning methodologies using CNNs are preferred for their capability to automatically identify spatial hierarchies of features in images [9]. CNNs excel in medical imaging due to the capability to automatically extract features from raw images, eliminating the manual features extraction. They capture a range of representations, from simple edges to complex patterns, enhancing detection accuracy. Despite challenges like the gradient vanishing issue and varying scales of retinopathy regions in fundus images, the objective of this study is to automatically detected Diabetic Retinopathy (DR) using CNNs.

A. Research Contributions

The research contributions are summarized below:

This study presents the Diabetic Retinopathy Detection Network (DRD-Net), a CNN framework that integrates elements from VGG16, DenseNet, and Inception Net. DRD-Net leverages the depth of VGG16, DenseNet's connectivity, and Inception Net's efficiency, incorporating a "Dense Block" modeled by DenseNet-201, which includes features specialized
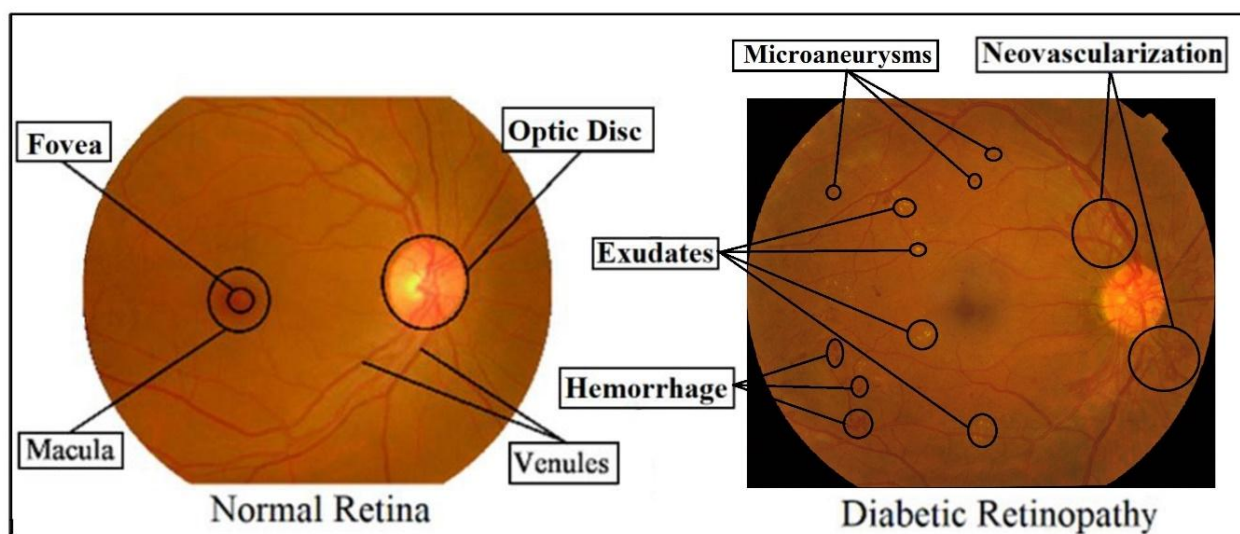

FIGURE 1. Normal retina vs. diabetic retinopathy affected retina

transition and convolutional layers. The architecture involves multiple convolutional and max-pooling layers, alongside two Inception sections for the extraction of multi-scale features. To enhance feature flow and mitigate the gradient vanishing problem, dense connections are integrated. The network incorporates Global Average Pooling (GAP) for feature aggregation, a dropout layer for regularization, and concludes with a softmax activated output layer for classification. The Swish activation function is included to enhance convergence speed and overall performance [8].

Performance evaluations are carried out to compare the proposed diabetic retinopathy detection method against prominent CNN architectures designed for this DR detection: AlexNet [10], Light-CNN [11], DenseNet-121[12] , ResNet-50 [13], Inception V3 [14], and MobileNet V2 [15].

B. Section Organization and Outline

This study is organized as follows: Current state-of-the-art is discussed in Section 2, Section 3 outlines the proposed architecture of diabetic retinopathy detection, experimental findings are discussed in Section 4, and in Section 5 we discussed the conclusion and future work.

II. CURRENT STATE-OF-THE-ART

The rising incidence of diabetes and the serious effects of advanced diabetic retinopathy (DR) emphasize the importance of early detection. Timely diagnosis is vital for effective treatment. With advancements in Artificial Intelligence (AI), particularly in convolutional neural networks (CNNs), interest in their application for DR diagnostics is on the rise.

Discusses the difficulties in manually interpreting color fundus images for diabetic retinopathy (DR) diagnosis, as it is time-consuming and prone to errors. To address these challenges, the author proposed a CNN model that efficiently classifies DR images using a streamlined set of learnable parameters. This model integrates VGG16 with a spatial pyramid pooling (SPP) layer, improving nonlinearity and robustness against scale variations. In [11], the authors analyzed several CNN-based method for DR detection, and proposed a lightweight feed-forward CNN. Using the IDRiD dataset, their results demonstrated the effectiveness of CNN approaches in DR diagnosis, highlighting CNN's significant impact on medical diagnostics.

In [16] the authors surveyed VGG16 and VGG19 CNN frameworks for fundus image analysis. Using 5,000 images, they showed VGG16's superior performance in segmenting DR across severity levels and suggested that genetic algorithms and 5-fold cross-validation could enhance diagnostic accuracy. On the other hand, authors [4] addresses challenges in DR image analysis, such as low contrast, inadequate lighting, and noise, which affect diagnostic accuracy. The authors improved the VGG16 architecture to enhance classification precision, employing augmentation and class balancing techniques during training of model, which strengthens its robustness.

Ali et al.,[17] introduced Incremental Modular Networks (IMNets), which consist of small subnet modules enhanced to detect specific features in images. This design promotes efficient feature capture while minimizing computational demands, improving performance. In another study Gangwar et al., [18] introduces a novel CNN framework for classifying DR images. With ten layers of 3x3 convolutions inspired by VGGNet and leaky ReLU activation, the model identifies vascular features like hemorrhages, exudates, and microaneurysms, categorizing them from 'absent' to 'proliferative' based on severity.

In [19], introduced a deep learning framework that incorporates three integrated CNNs into a single meta-learner for DR diagnosis. Utilizing the EyePACS dataset, their model achieved 97.92% accuracy in binary classification, outperforming models like VGG16 and ResNet-50. However, further enhancements in image quality and noise reduction are still needed. Turning to [20], the authors combined deep learning and machine learning by using features from ResNet-50 to enhance a Random Forest classifier for DR image classification. This method achieved 96% accuracy on Messidor2 dataset and 75.09% on EyePACS dataset, showcasing the effectiveness of integrating these approaches into DR diagnostics.

In [21], the authors outlined a two-phase technique for DR detection from fundus images, focusing on feature extraction and classification. Their approach involves color space transformation, optic disc removal, and segmentation of blood vessels. Utilizing a Deep Convolutional Neural Network (DCNN) and SVM with Genetic Algorithm (SVMGA), they achieved 98.80% accuracy but lacked detailed severity classification.

In [22] , the authors presented the Coarse-to-Fine Network (CF-DRNet), which includes a coarse network for binary classification and a fine network for classifying four levels DR severity. Using the IDRiD and Kaggle datasets with data augmentation, CF-DRNet outperformed ResNet, achieving an accuracy of 83.10% on Kaggle dataset and 56.19% on the IDRiD dataset.

In [23], the authors presented a deep learning method to categorize images as either DR or no-DR. Their process involves pre-processing, Hessian matrix-based segmentation, and feature extraction via CNN. Tested their model on DIARETDB1 dataset and achieved a precision of 97.2% and an accuracy of 98.7%, surpassing traditional techniques.

To combat limited data availability, [24] developed a deep learning architecture using MESSIDOR dataset, refining AlexNet by adjusting convolution and max-pooling configurations in the starting layers and enhancing the final fully connected layers, resulting notable accuracy.

In [25], the authors explored deep neural networks like VGG16, Inception, and ResNet, focusing on transfer learning and employing Gaussian preprocessing to reduce noise. InceptionV3 emerged

as the top performer with a 81.2% accuracy in training and 79.4% validation accuracy.

In [26], the authors presented the DAG network model to extract complex DR features for accurate categorization. Their approach, validated with medical institution data and the DIARETDB1 dataset, demonstrated effective classification.

In [27], the authors proposed a CNN-based DR diagnostic system that accurately categorized fundus images by DR presence and severity, achieving improved detection rates using a public Kaggle dataset.

In parallel, [28] investigated integrating CNN with meta-plasticity for DR detection, showing enhanced performance during backpropagation with the InceptionV3 model on a widely available dataset.

Recognizing the challenges of attaining peak accuracy in DR classification with singular-view fundus datasets, [29] developed an automated DR detection system using multi-view fundus images, employing attention mechanisms to extract lesion attributes and improve diagnostic accuracy, outperforming traditional models on a multi-view dataset.

Progress in diabetic retinopathy (DR) lesion identification faces challenges, especially with small lesions. Accurate detection relies on high-resolution imaging, highlighting the need for quality retinal images. The tiny size of some lesions can complicate differentiation between closely located DR sites. Furthermore, recognizing severe lesions with large receptive fields may encounter gradient issues that hinder model learning. Handling high-definition images also demands substantial computational resources, making training both resource-intensive and costly.

A. Research Gap Analysis

A detailed review of the existing literature reveals numerous unexplored areas and opportunities for further innovation: Research focused on straightforward feed-forward CNNs [30], [24], including architectures like VGG16 and AlexNet, highlights their challenges with the gradient vanishing problem This issue becomes especially prominent during the training of deep networks, ultimately affecting the model's accuracy and effectiveness [12]. While densely interconnected CNNs for DR detection [31] appear as a more promising solution compared to

basic feed-forward CNNs, the extensive incorporation of dense connection blocks in some design conductors in considerable computational overheads [9]. Although dense connection blocks are designed to enhance feature transmission, they also significantly increase the model's complexity, making training more challenging in resource-limited environments. Despite the merits of DenseNet-based DR detection models [31], there appears to be an oversight concerning the importance of multiscale features [32]. Recognizing scale differences is critical, particularly in images that display Regions of Interest (RoI) of varying sizes. Overlooking this multiscale dimension could undermine the network's proficiency in generalizing across varied datasets characterized by important RoI scale variations [8]. While Inception-Net stands as a remarkable CNN architecture adept at extracting multiscale features [33], it tends to neglect the importance of multilevel features. Such an oversight can be detrimental in combating the gradient vanishing challenge [9]. Neglecting to leverage multilevel features could hinder the network's optimization efforts, leading to subpar results. A discernible pattern across various CNN designs [27], [34], [35], [36] is their reliance on conventional activation functions. At times, this leads to a 'dead neuron' predicament. When a neuron stagnates during training and regularly emits a uniform value, and it becomes inert or non-contributory, which hinders the network's ability to learn effectively [32]. These issues make training more difficult and limit the model's effectiveness. By addressing these gaps, we can pave the way for developing DR detection models that are both strong and adaptable. Such improvements could lead to greater accuracy, better training methods, and improved handling of complex datasets.

III. MATERIALS AND METHODS

A. Data Acquisition

The IDRiD (Indian DRImage Dataset) and APTOS19 is a key resource for diabetic retinopathy (DR) research. Both datasets contain five classes: Normal, Mild, Moderate, Severe, and Proliferative. The IDRiD dataset comprises 516 high-resolution color fundus images, while APTOS19 contains 3,662 images that capture various aspects of the retina and DR stages. In this research, these datasets were combined to validate the effectiveness of the proposed model. The primary goal is to advance and

TABLE 1. Dataset distribution before and after

Classes	Dataset			Post-Augmentation		
	IDRiD	APTOS19	Total	Train	Test	Total
Normal	168	1805	1973	1578	395	1973
Mild	25	370	395	1264	316	1580
Moderate	168	999	1167	933	234	1167
Severe	93	193	286	915	229	1144
Proliferative	62	295	357	1142	286	1428

benchmark image classification algorithms for early DR detection.

The original dataset distribution is as follows: Normal with 1,973 samples, Mild with 395 samples, Moderate with 1,167 samples, Severe with 286 samples, and Proliferative with 357 samples. However, it was observed that the Mild, Severe, and Proliferative classes had fewer samples, leading to a class imbalance issue. To address this limitation, geometric augmentation techniques were applied on these classes, expanding the total dataset size to 7,292 images. Of these, Mild with 1,580 images, Moderate with 1,167 images, Severe with 1,144 images, Proliferative with 1,428 images, Normal with 1,973 images representing healthy retinas. For machine learning applications, the dataset is split into two sets training and evaluation, with 80% (5,832 images) allocated for training which helps the model to learn the intricate patterns of DR and remaining 20% (1,460 images) are reserved for testing, providing an unbiased assessment of model performance. All images are in the universal JPG format, enhancing accessibility and compatibility across different computational platforms.

B. Image Preprocessing and Augmentation

Image preprocessing is key for improving DR classification. Images are resized to 224x224 pixels for consistency, then pixel values are normalized between 0 and 1 to facilitate smoother and faster model training, reducing issues like gradients vanishing or exploding. To further enhance lesion visibility and subtle retinal features CLAHE is applied. CLAHE performs localized histogram equalization within small regions of the image, enhancing local contrast while preventing over-amplification of noise. We set the clip limit to 2.0 to preserve critical retinal structures like microaneurysms and hemorrhages without introducing excessive noise or distorting the image. The 8x8 tile grid size ensures effective contrast enhancement across the image, preserving both large-scale structures and intricate details.

To expand dataset diversity and improve model adaptability, image augmentation techniques are applied. This includes 180-degree rotations to account for orientation changes and horizontal and vertical flips to help the model recognize retinal features from various spatial perspectives. Figure 2 shows visuals of geometrically augmented samples.

C. The Proposed CNN Architecture

In the field of deep learning for biomedical image classification, improving established architectures to create better models is a common approach. Our proposed architecture combines the strengths of VGG16, DenseNet, and Inception Net. By combining VGG16's ability to extract rich semantic features, DenseNet's dense connections to improve gradient flow and training efficiency, and Inception Net's multiscale feature extraction to handle various lesion sizes, we aim to develop a more effective DR detection

model. This architecture enhances feature extraction, computational efficiency, and overall accuracy in identifying complex DR patterns.
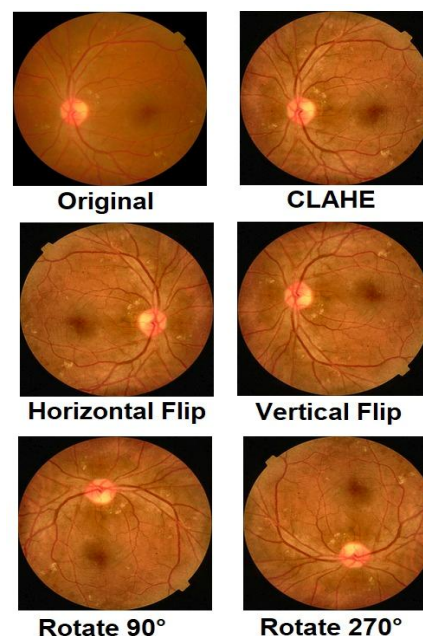

FIGURE 2. Data augmentation on original image.

The CNN architecture shown in Figure 3 starts with a convolutional and max-pooling block designed to capture detailed patterns from color input fundus images. The first two layers have 3x3 filters, with the initial layer containing 64 filters and the next with 128. This gradual increase in filter count enables the model to recognize increasingly complex features as it goes deeper, while the compact 3x3 filter size helps keep computational requirements manageable.

Following the convolutional layers, a max pooling layer is added with a stride of 2, reducing the feature map dimensions from 224x224 to 112x112. Max pooling decreases spatial dimensions, lowering computational complexity and mitigating overfitting. The second convolutional and max-pooling block builds on the first, using layers with 128 and 256 filters to capture increasingly abstract features. After this, a max-pooling layer reduces the feature map size to 56x56, preparing it for more complex analysis. Next is an Inception block, designed to capture features at multiple scales. It includes four 1x1 convolution layers, a 3x3 convolution layer containing 192 filters, a 5x5 convolution layer containing 96 filters, and a 3x3 max pooling layer with stride 1. This arrangement supports simultaneous feature extraction at different levels of detail and improves computational efficiency by reducing dimensionality before applying larger filters.

After the Inception block, a 1x1 convolutional layer containing 256 filters decreases dimensionality and enhances non-linearity while decreasing computational demands. A subsequent max pooling layer with a stride of 2 further lowers feature map dimensions. The third convolutional and max pooling block follows, extracting 256 and 512 features through
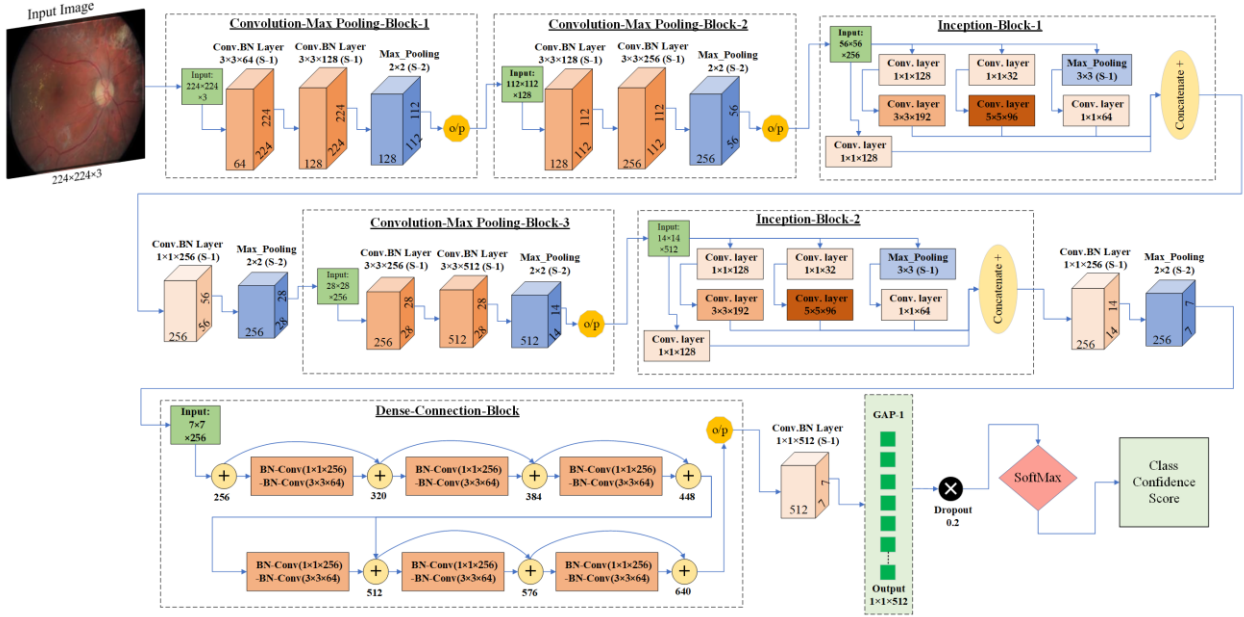
FIGURE 3. Proposed DRD-Net backbone architecture.

two convolutional layers, which are then processed by another max pooling layer. Finally, a second Inception block is introduced, similar to the first, to extract multiscale features at a deeper level.

D. The Activation and Loss Function

To avoid the need for one-hot encoding, we used the sparse categorical cross-entropy loss function in this framework for multi-class classification with integer target labels. This method maintains computational resources and works effectively in the final layers by using SoftMax activation, optimizing model to maximize the probability of the correct class. The loss is mathematically defined as,

The Activation function Swish is defined as:

$$f(x) = x \times \sigma(x) \qquad (1)$$

Sigmoid function is denoted as σ(x), and it's given by the equation:

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad (2)$$

Mathematically, the Loss is defined as follows:

$$Loss = -\sum_{i=1}^{n} t_i \log(p_i) \qquad (3)$$

Where pi represents the Softmax probability for the ith class, n is number of classes, and ti is the actual labels.

E. Technologies

This research utilizes Google Colab for tasks ranging from preprocessing of data to model evaluation. Colab, a cloud-based Jupyter Notebook platform, offers an interactive environment for visualization and analysis of data. Its integration with powerful GPUs, such as the NVIDIA Tesla K80, significantly speeds up computation and reduces training time compared to non-GPU systems.

F. Training and Testing

For automated DR detection from retinal images, we utilized TensorFlow and Keras frameworks. The dataset was categorized into five classes, with each image resized to 224x224 pixels for computational efficiency and detail preservation. We used Keras's ImageDataGenerator for image normalization and allocated 20% of the data for validation to evaluate the model on new samples.

The Adam optimizer was applied alongside with sparse categorical cross-entropy loss, suitable for sparse labels. Early stopping was used to monitor validation loss and stop training if no improvement was monitored for 15 epochs, preventing overfitting. Then 60 epochs was set for the training phase, with visual plots of accuracy and loss providing insights into model performance.

G. Performance Evaluation Metrics

In conclusion, we evaluated the model's ability in detecting and categorizing diabetic retinopathy using performance indicators, which includes accuracy, recall, precision, and the F1-score.

$$Precision = \frac{TP}{TP+FP} \qquad (4)$$

$$Recall = \frac{TP}{TP+FN} \qquad (5)$$

$$F1-score = 2 \times \frac{(Recall \times Precision)}{(Recall+Precision)} \qquad (6)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \qquad (7)$$

IV. EXPERIMENTS

We introduced DRD-Net to improve DR diagnosis using deep learning and compared it with six different models such as: AlexNet, DenseNet-121, Light CNN,

ResNet-50, MobileNet and InceptionNet. Experiments included evaluating activation functions, and testing each architecture's performance, with DRD-Net demonstrating its effectiveness for detecting DR.

A. Behavior comparison of ReLU and Swish Activation Functions

In the initial phase, we compared the effects of the traditional ReLU activation function with the newer Swish function using a CNN framework which consists of two convolutional and max pooling blocks, as depicted in the Figure 4.
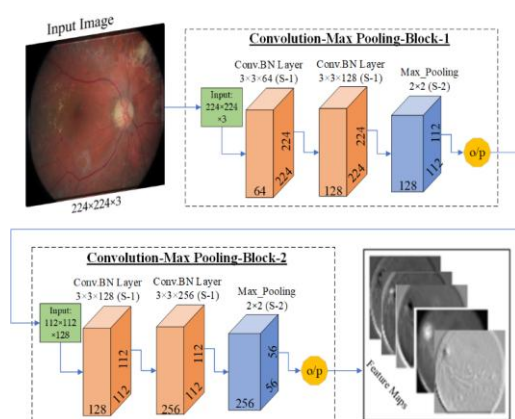
This framework utilized two consecutive convolutional layers with 3x3 filters to efficiently capture localized patterns in the input images. A max pooling layer of stride 2 subsequently reduced the feature map dimensions from 224x224 to 112x112, enhancing computational efficiency and providing translational invariance to mitigate overfitting. The second block followed a similar structure but focused on more complex image details. However, the study highlighted that the ReLU activation function suffered from "dead neurons" resulting in empty feature maps and a loss of critical information, as illustrated in Figure 5.

In contrast, the Swish-activated feature maps in Figure 6 showed no blank areas, emphasizing its superiority over ReLU. Swish consistently activated features, effectively preserving critical details essential for diabetic retinopathy detection.
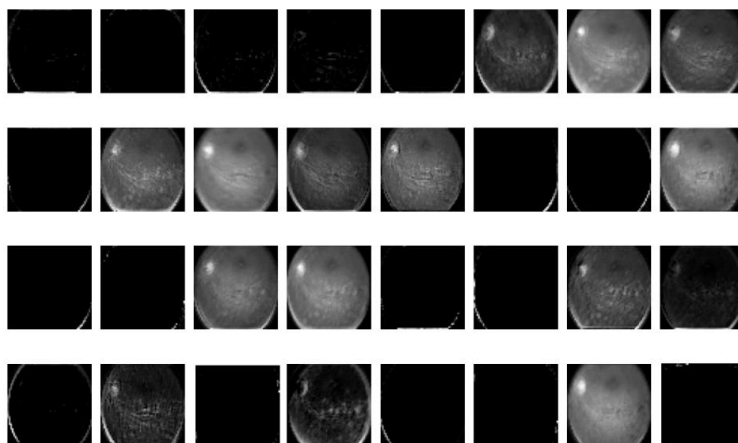

FIGURE 4. Max pooling and two consecutive conv. blocks.

FIGURE 5. Feature maps generated with ReLU activation, highlighting the "Dead Neuron" issue shown through visibly blank maps.
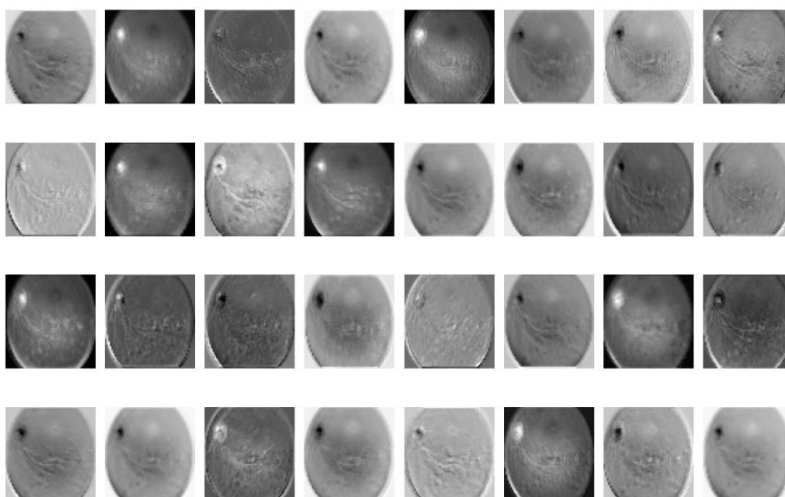

FIGURE 6. Feature maps generated with SWISH activation, showcasing the robustness of SWISH by demonstrating consistent activation without blank maps.

B. DRD-Net Performance Evaluation and Comparison

We analyze the performance of our customized CNN architecture (i.e., DRD-Net). Experimental results demonstrate the model's effectiveness. Figure 7 and Figure 8 depict the training and validation loss and accuracy, illustrating the model's performance progression.
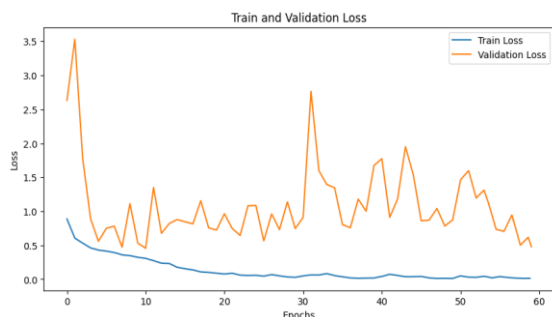

FIGURE 7. Training and validation loss for the proposed Multi-scale DRD-Net model.

FIGURE 8. Training and validation accuracy for the proposed Multi-scale DRD-Net.

The confusion matrix in Figure 9 illustrates the performance of the proposed model across the five classes. For the Normal class, the model correctly classified 382 out of 395 samples, only misclassified 13 samples as Moderate, For the Mild class, the model correctly classified 259 out of 316 samples, with 23 samples misclassified as Severe, 33 as Proliferative, and only one as Moderate. The model performed exceptionally well for the Moderate class, misclassifying just 12 out of 234 samples, with 9 samples classified as Mild and 1 as Severe, correctly classifying 222 samples. However, the model faced challenges with the Severe and Proliferative classes due to their similar and complex patterns. For the Severe class, the model correctly classified 188 out of 229 samples, with 39 samples misclassified as Proliferative and 2 as Moderate. For the Proliferative class, the model correctly classified 192 out of 286 samples, with 78 samples misclassified as Severe, 10 as Normal, 3 as Mild, and 3 as Moderate. Despite these challenges, the model demonstrates strong overall performance in DR diagnosis, particularly for the Normal, Mild, and Moderate classes.
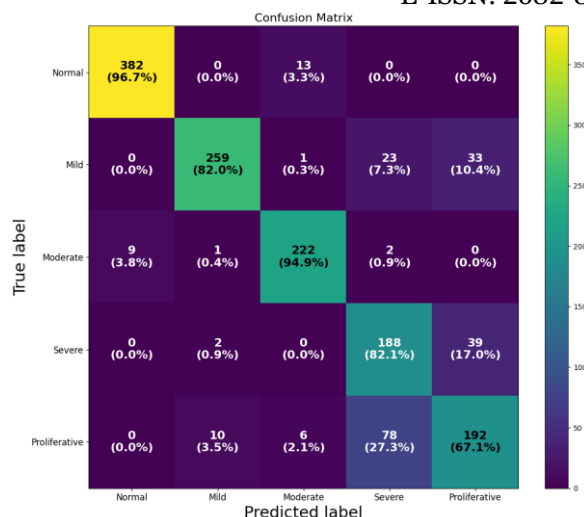

FIGURE 9. Proposed Multi-scale DRD-Net's confusion matrix.

Table 4 presents a comparison of DRD-Net's performance against state-of-the-art classifiers for diagnosing diabetic retinopathy. AlexNet, a lightweight foundational CNN, performs well in the Normal class (90.6% precision, 95.4% recall, 92.97% F1-score) and Moderate class (85.89% F1-score). However, due to inter-class similarity and intra-class variation, the model struggles in the Severe and Proliferative classes, achieving F1-scores of 57.79% and 49.89%, respectively. ResNet-50, a deep model with residual connections, outperforms AlexNet and achieves strong results in Normal (94.94% F1-score) and Moderate (88.15% F1-score), but performs poorly in Proliferative (43.98% F1-score). The computationally efficient Light Weight CNN performs well in Normal (95.87% F1-score) and Moderate (91.03% F1-score) but struggles in Severe (47.75% F1-score). With dense connectivity for feature reuse, DenseNet-201 struggles in Severe (52.33% F1-score) but performs well in Normal (94.66% F1-score) and Moderate (89.61% F1-score), with high Mild recall (97.2% recall). Inception V3, utilizing multi-scale feature extraction, performs well in Normal (96.76% F1-score) and Moderate (93.01% F1-score) but struggles in Severe (45.76% F1-score). MobileNet V2, optimized for mobile applications, matches DRD-Net in Severe (72.3% F1-score) by striking a balance between Normal (94.5% F1-score) and Proliferative (72.7% F1-score). For proposed DRD-Net's model, Normal class outperforms all other models with 97.7% precision, 96.7% recall, and 97.20% F1-score. The Mild class achieves a notable precision of 95.2%, recall of 82.0%, and F1-score of 88.10%. While the Proliferative class ranks second among all models with 72.7% precision, 67.1% recall, and 69.82% F1-score, the Severe class excels in recall with 64.6% precision, 82.1% recall, and 72.31% F1-score. These results demonstrate DRD-Net's strong performance across all classes.

TABLE 2. Comparative analysis of DRD-Net with state-of-the-art schemes.

CNN architecture	Class	TP	FP	FN	TN	Precision (%)	Recall (%)	F1-Score (%)
AlexNet [10]	Normal	377	39	18	1083	90.6	95.4	92.97
	Mild	264	94	52	1196	73.7	83.5	78.34
	Moderate	207	41	27	1253	83.5	88.5	85.89
	Severe	141	118	88	1319	54.4	61.6	57.79
	Proliferative	116	63	170	1344	64.8	40.6	49.89
ResNet-50 [13]	Normal	394	41	1	1066	90.6	99.7	94.94
	Mild	247	51	69	1213	82.9	78.2	80.46
	Moderate	186	2	48	1274	98.9	79.5	88.15
	Severe	189	204	40	1271	48.1	82.5	60.77
	Proliferative	95	51	191	1365	65.1	33.2	43.98
Light Weight CNN [11]	Normal	383	21	12	1077	94.8	97.0	95.87
	Mild	243	65	73	1217	78.9	76.9	77.88
	Moderate	208	15	26	1252	93.3	88.9	91.03
	Severe	101	93	128	1359	52.1	44.1	47.75
	Proliferative	178	153	108	1282	53.8	62.2	57.70
DenseNet-201 [12]	Normal	390	39	5	1070	90.9	98.7	94.66
	Mild	307	159	9	1153	65.9	97.2	78.52
	Moderate	194	5	40	1266	97.5	82.9	89.61
	Severe	101	56	128	1359	64.3	44.1	52.33
	Proliferative	132	77	154	1328	63.2	46.2	53.33
Inception Net V3 [33]	Normal	388	19	7	1072	95.3	98.2	96.76
	Mild	305	82	11	1155	78.8	96.5	86.77
	Moderate	213	11	21	1247	95.1	91.0	93.01
	Severe	81	44	148	1379	64.8	35.4	45.76
	Proliferative	198	119	88	1262	62.5	69.2	65.67
MobileNet V2 [15]	Normal	382	31	13	1078	92.5	96.7	94.5
	Mild	239	32	77	1221	88.2	75.6	81.4
	Moderate	185	50	49	1275	78.7	79.1	78.8
	Severe	162	57	67	1298	74.0	70.7	72.3
	Proliferative	221	101	65	1239	68.6	77.3	72.7
DRD-Net	Normal	382	9	13	1078	97.7	96.7	97.20
	Mild	259	13	57	1201	95.2	82.0	88.10
	Moderate	222	20	12	1238	91.7	94.9	93.28
	Severe	188	103	41	1272	64.6	82.1	72.31
	Proliferative	192	72	94	1268	72.7	67.1	69.82

AlexNet records an average of 73.4% precision, 73.9% recall, 72.9% F1-score, and 75.7% accuracy, while ResNet-50 shows a slight improvement with an average of 77.1% precision, 74.6% recall, 73.6% F1-score, and 76.1% accuracy, reflecting better precision. Light Weight CNN achieves an average of 74.5% precision, 73.8% recall, 74.1% F1-score, and 76.2% accuracy, improving over ResNet-50 in precision and accuracy.

DenseNet-201 records an average of 76.3% precision, 73.8% recall, 73.6% F1-score, and 76.9% accuracy, showing a notable gain in recall compared to Light Weight CNN. Inception Net V3 performs at an average of 79.3% precision, 78.1% recall, 77.6% F1-score, and 81.1% accuracy, with better recall than DenseNet-201.

MobileNet V2 reaches an average of 80.4% precision, 79.8% recall, 79.9% F1-score, and 81.4% accuracy, slightly improving over Inception Net V3 in

F1-score. The proposed DRD-Net model leads with an average of 84.4% precision, 84.5% recall, 84.1% F1-score, and 85.1% accuracy, demonstrating the highest performance across all metrics. These results are illustrated in Figure 10.
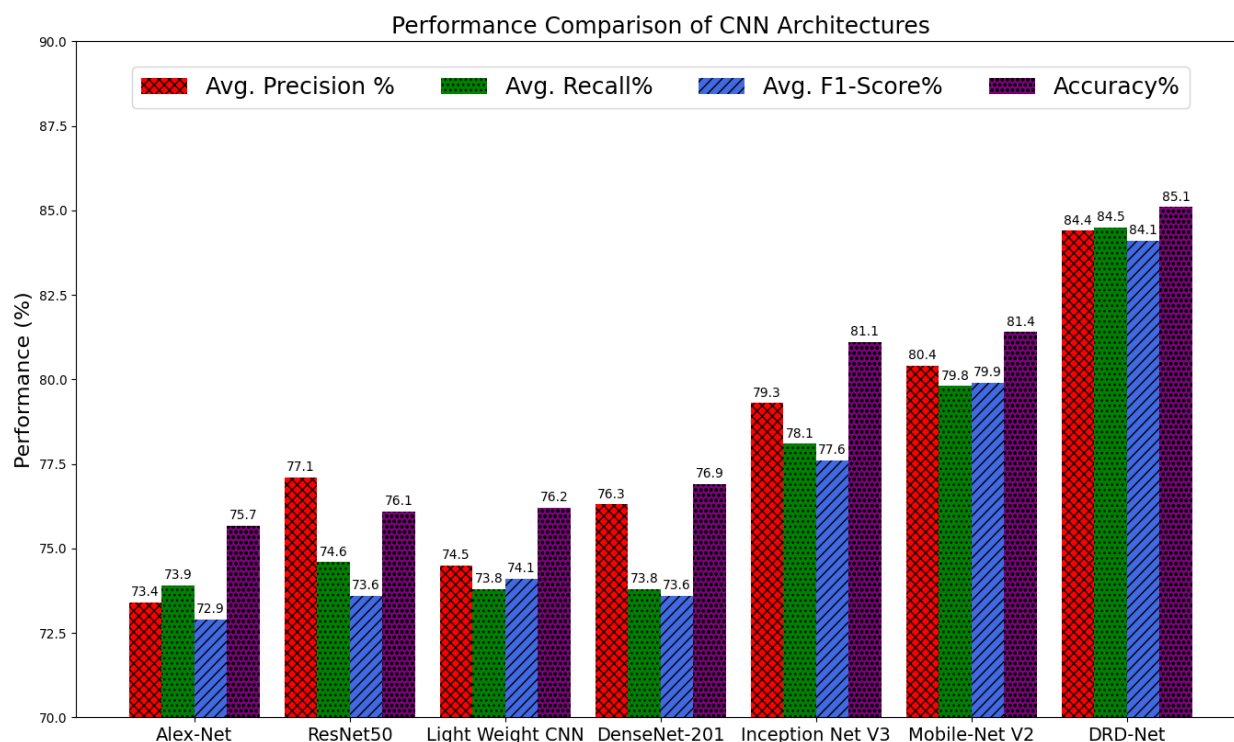
FIGURE 10. Overall Comparative analysis of DRD-Net with state-of-the-art schemes

V. CONCLUSION

The increasing prevalence of diabetes necessitates an urgent need for accurate and rapid diabetic retinopathy (DR) diagnosis. This study introduces the DRD-Net framework, which automates DR identification and grading by integrating features from established architectures like VGG16, DenseNet, and Inception Net. Our analysis underscores the importance of combining both multi-level features (MLF) and multi-scale features (MSF) for enhanced performance. DRD-Net merges feedforward CNN, Inception, and dense connection blocks, achieving an accuracy of 85.1%. With superior precision, recall, and F1-score, it outperforms models such as ResNet-50, AlexNet, Light-CNN, DenseNet-121, MobileNet V2 and Inception V3, demonstrating its effectiveness in DR detection. As diabetes cases rise, the automation provided by DRD-Net can alleviate the diagnostic burden on healthcare professionals, representing a significant advancement in the integration of medical imaging and deep learning.

Future work will include testing the proposed DRD-Net framework on more diverse datasets, refining its activation functions, and optimizing strategies to further improve performance.

AUTHOR CONTRIBUTIONS

Muhammad Hassaan Ashraf: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation, Visualization, Validation, Supervision, Software, Resources, Project Administration, Investigation, Formal analysis;

Muhammad Esham Qureshi: Conceptualization, Data Curation, Methodology, Validation, Writing – Review & Editing, Software, Resources, Investigation, Formal analysis;

Ahmed Khan: Conceptualization, Data Curation, Validation, Software, Resources, Investigation, Formal analysis;

Jawaid Iqbal: Project Administration, Supervision, Writing – Review & Editing;

Musharif Ahmed: Project Administration, Supervision, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Ethical approval was not required for this study as it utilized publicly available and anonymized dataset (i.e., IDRiD). All data were obtained in compliance with the source's terms of use, and no new data were collected from human participants. The study did not involve any experiments on humans or animals.

REFERENCES

[1]   W. Ling, Y. Huang, Y.M. Huang, R.R. Fan, Y. Sui, H.L. Zhao, "Global Trend of Diabetes Mortality Attributed to

Vascular Complications, 2000–2016," *Cardiovasc Diabetol*, vol. 19, no. 1, 2020.
DOI: https://doi.org/10.1186/s12933-020-01159-5

[2] M.H. Soomro and A. Jabbar, "Diabetes Etiopathology, Classification, Diagnosis, and Epidemiology," in *BIDE's Diabetes Desk Book: For Healthcare Professionals*, Elsevier, pp. 19–42, 2024.
DOI: https://doi.org/10.1016/B978-0-443-22106-4.00022-X

[3] M. Kąpa, I. Koryciarz, N. Kustosik, P. Jurowski and Z. Pniakowska, "Future Directions in Diabetic Retinopathy Treatment: Stem Cell Therapy, Nanotechnology, and PPARα Modulation," *Journal of Clinical Medicine*, vol. 14, no. 3, p. 683, 2025.
DOI: https://doi.org/10.3390/JCM14030683

[4] D.A. da Rocha, F.M.F. Ferreira and Z.M.A. Peixoto, "Diabetic Retinopathy Classification Using VGG16 Neural Network," *Research on Biomedical Engineering*, vol. 38, no. 2, pp. 761–772, 2022.
DOI: https://doi.org/10.1007/S42600-022-00200-8

[5] R. Mumtaz, M. Hussain, S. Sarwar, K. Khan, S. Mumtaz and M. Mumtaz, "Automatic Detection of Retinal Hemorrhages by Exploiting Image Processing Techniques for Screening Retinal Diseases in Diabetic Patients," *International Journal of Diabetes in Developing Countries, vol. 38, no. 1, pp. 80-87, 2018*.
DOI: https://doi.org/10.1007/S13410-017-0561-6

[6] P. Chudzik, S. Majumdar, F. Calivá, B. Al-Diri and A. Hunter, "Microaneurysm Detection Using Fully Convolutional Neural Networks," *Computer Methods and Programs in Biomedicine, vol. 158, pp. 185-192, 2018*.
DOI: https://doi.org/10.1016/J.CMPB.2018.02.016

[7] G. Lim, V. Bellemo, Y. Xie, X. Q. Lee, M. Y. T. Yip and D.S.W. Ting, "Different Fundus Imaging Modalities and Technical Factors in AI Screening for Diabetic Retinopathy: A Review," *Eye and Vision*, vol. 7, no. 1, 2020.
DOI: https://doi.org/10.1186/S40662-020-00182-7

[8] M.H. Ashraf and H. Alghamdi, "HFF-Net: A Hybrid Convolutional Neural Network for Diabetic Retinopathy Screening and Grading," *Biomedical Technology*, vol. 8, pp. 50–64, 2024.
DOI: https://doi.org/10.1016/J.BMT.2024.09.004

[9] M.H. Ashraf, F. Jabeen, H. Alghamdi, M.S. Zia and M.S. Almutari, "HVD-Net: A Hybrid Vehicle Detection Network for Vision-Based Vehicle Tracking and Speed Estimation," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, 2023.
DOI: https://doi.org/10.1016/j.jksuci.2023.101657

[10] A. Krizhevsky, I. Sutskever and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
URL: https://papers.nips.cc/paper_files/paper/2012/hash/c39986 2d3b9d6b76c8436e924a68c45b-Abstract.html
(Accessed 8 June, 2025)

[11] N. Jiwani, K. Gupta and N. Afreen, "A Convolutional Neural Network Approach for Diabetic Retinopathy Classification," *IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 637–642, 2022.
DOI: https://doi.org/10.1109/CSNT54456.2022.9787577

[12] G. Huang, Z. Liu, L.V.D. Maaten and K.Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.
DOI: https://doi.org/10.1109/CVPR.2017.243

[13] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
DOI: https://doi.org/10.1109/CVPR.2016.90

[14] G. Kurup, J.A.A. Jothi and A. Kanadath, "Diabetic Retinopathy Detection and Classification Using Pretrained Inception-v3," *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON 2021)*, pp. 1–6, 2021.

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.
DOI: https://doi.org/10.1109/CVPR.2018.00474

[16] S.K.R. Meruva, V.G.S. Tulasi, N. Vinnakota and V. Bhavana, "Risk Level Prediction of Diabetic Retinopathy Based on Retinal Images Using Deep Learning Algorithm," *Procedia Computer Science,* vol. 215, pp. 722-730, 2022.
DOI: https://doi.org/10.1016/J.PROCS.2022.12.074

[17] R. Ali, R.C. Hardie, B.N. Narayanan and T.M. Kebede, "IMNets: Deep Learning Using an Incremental Modular Network Synthesis Approach for Medical Imaging Applications," *Applied Sciences*, vol. 12, no. 11, p. 5500, 2022.
https://doi.org/10.3390/APP12115500

[18] A.K. Gangwar and V. Ravi, "Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning," *Advances in Intelligent Systems and Computing*, vol. 1176, pp. 679–689, 2021.
DOI: https://doi.org/10.1007/978-981-15-5788-0_64

[19] H. Kaushik, D. Singh, M. Kaur, H. Alshazly, A. Zaguia and H. Hamam, "Diabetic Retinopathy Diagnosis From Fundus Images Using Stacked Generalization of Deep Models," *IEEE Access*, vol. 9, pp. 108276–108292, 2021.
DOI: https://doi.org/10.1109/ACCESS.2021.3101142

[20] M.K. Yaqoob, S.F. Ali, M. Bilal, M.S. Hanif and U.M. Al-Saggaf, "ResNet Based Deep Features and Random Forest Classifier for Diabetic Retinopathy Detection," *Sensors*, vol. 21, no. 11, p. 3883, 2021.
DOI: https://doi.org/10.3390/S21113883

[21] O.M. Al-hazaimeh, A.A. Abu-Ein, N.M. Tahat, M.A. Al-Smadi, M.M. Al-Nawashi, "Combining Artificial Intelligence and Image Processing for Diagnosing Diabetic Retinopathy in Retinal Fundus Images," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 13, pp. 131–151, 2022.
DOI: https://doi.org/10.3991/IJOE.V18I13.33985

[22] Z. Wu, "Coarse-to-Fine Classification for Diabetic Retinopathy Grading Using Convolutional Neural Network," *Artificial Intelligence in Medicine,* vol. 108, pp. 101936, 2020.
DOI: https://doi.org/10.1016/J.ARTMED.2020.101936

[23] S. Das, K. Kharbanda, S. M, R. Raman and E.D. D, "Deep Learning Architecture Based on Segmented Fundus Image Features for Classification of Diabetic Retinopathy," *Biomedical Signal Process Control*, vol. 68, p. 102600, 2021.
DOI: https://doi.org/10.1016/J.BSPC.2021.102600

[24] T. Shanthi, R.S. Sabeenian, "Modified Alexnet Architecture for Classification of Diabetic Retinopathy Images," *Computers & Electrical Engineering*, vol. 76, pp. 56–64, 2019.
DOI: https://doi.org/10.1016/J.COMPELECENG.2019.03.004

[25] A. Khan, N. Kulkarni, A. Kumar and A. Kamat, "D-CNN and Image Processing Based Approach for Diabetic Retinopathy Classification," *Advances in Intelligent Systems and Computing,* pp. 283–291, 2022.
DOI: https://doi.org/10.1007/978-981-16-2008-9_27

[26] L. Fang and H. Qiao, "A Novel DAG Network Based on Multi-Feature Fusion of Fundus Images for Multi-Classification of Diabetic Retinopathy," *Multimedia Tools and Applications*, vol. 82, no. 30, pp. 47669–47693, 2023.
DOI: https://doi.org/10.1007/S11042-023-15296-Y

[27] S. Sridhar, J.P. Kandhasamy, M. Sinthuja and T.N.S. Minish, "WITHDRAWN: Diabetic Retinopathy Detection Using Convolutional Nueral Networks Algorithm," *Materials Today: Proceedings*, 2021.
DOI: https://doi.org/10.1016/J.MATPR.2021.01.059

[28] V. Vives-Boix and D. Ruiz-Fernández, "Diabetic Retinopathy Detection Through Convolutional Neural Networks With Synaptic Metaplasticity," *Computer Methods and Programs in Biomedicine*, vol. 206, p. 106094, 2021.

DOI:https://doi.org/10.1109/SMARTGENCON51891.2021.9645914

DOI: https://doi.org/10.1016/J.CMPB.2021.106094

[29]   X. Luo, "MVDRNet: Multi-View Diabetic Retinopathy Detection by Combining DCNNs and Attention Mechanisms," *Pattern Recognition*, vol. 120, p. 108104, 2021.
DOI: https://doi.org/10.1016/J.PATCOG.2021.108104

[30]   Z. Khan, "Diabetic Retinopathy Detection Using VGG-NIN a Deep Learning Architecture," *IEEE Access*, vol. 9, pp. 61408–61416, 2021.
DOI: https://doi.org/10.1109/ACCESS.2021.3074422

[31]   S.G. Kobat, "Automated Diabetic Retinopathy Detection Using Horizontal and Vertical Patch Division-Based Pre-Trained DenseNET With Digital Fundus Images," *Diagnostics*, vol. 12, no. 8, p. 1975, 2022.
DOI: https://doi.org/10.3390/DIAGNOSTICS12081975

[32]   H. Alghamdi and T. Turki, "PDD-Net: Plant Disease Diagnoses Using Multilevel and Multiscale Convolutional Neural Network Features," *Agriculture*, vol. 13, no. 5, p. 1072, 2023.
DOI: https://doi.org/10.3390/AGRICULTURE13051072

[33]   C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.
DOI: https://doi.org/10.1109/CVPR.2016.308

[34]   S. Kanakaprabha, D. Radha and S. Santhanalakshmi, "Diabetic Retinopathy Detection Using Deep Learning Models," *Smart Innovation, Systems and Technologies*, vol. 302, pp. 75–90, 2022.
DOI: https://doi.org/10.1007/978-981-19-2541-2_7

[35]   T.J. Jie and M.S. Sayeed, "Review on Detecting Pneumonia in Deep Learning," *International Journal on Robotics, Automation and Sciences*, vol. 6, no. 1, pp. 70–77, 2024.
DOI: https://doi.org/10.33093/ijoras.2024.6.1.10

[36]   K.B. Gan and Y.E. Teoh, "An Edge Convolution Neural Network Model for Plant Health Classification Using Camera," *International Journal on Robotics, Automation and Sciences*, vol. 7, no. 1, pp. 1–6, 2025.
DOI: https://doi.org/10.33093/ijoras.2025.7.1.1