Research Paper

OptiCNN: Local thresholding Segmentation and CNN with SVM Approach for Diabetic Retinopathy Detection and Classification of Fundus Images

Anju Mishra[1], Mrinal Pandey[1], Laxman Singh[2]

[1]*Department of Computer Science & Technology, Manav Rachna University, Faridabad, India*
[2]*Department of Computer Science & Technology (AI & ML), KIET Group of Institutions, Ghaziabad (U.P.), India*

Article history
Received: 11 April 2025
Revised: 13 May 2025
Accepted: 28 May 2025

*Corresponding Author:
Anju Mishra
Department of Computer Science & Technology,
Manav Rachna University,
Faridabad, India
Email:
aanjumishra2108@gmail.com

Abstract: Diabetic Retinopathy (DR) is a progressive eye disease that can lead to vision loss and blindness if left untreated. Ophthalmologists diagnose DR using medical imaging modalities such as fundus photography and optical coherence tomography (OCT); however, manual interpretation of these images is time-consuming and subject to inter-observer variability. While DR is irreversible, vision loss can be prevented through early detection and timely intervention. Regular eye examinations and systematic DR monitoring are essential for preventing blindness in diabetic patients. Therefore, there is an urgent need for computer-assisted diagnosis (CAD) systems to support ophthalmologists in detecting and grading DR accurately and efficiently. This paper proposes a novel CAD system for automated DR classification. The proposed methodology consists of three stages: (1) image preprocessing through grayscale conversion and resizing, (2) retinal vessel segmentation using a local thresholding approach, and (3) classification using a hybrid architecture that integrates a Convolutional Neural Network (CNN) with a Support Vector Machine (SVM) classifier. The proposed model, OptiCNN (Optimized CNN), aligns its predictions with the International Clinical Diabetic Retinopathy (ICDR) severity scale. Experimental results demonstrate that OptiCNN achieves an accuracy of 94%, precision of 96%, recall of 91%, F1-score of 93%, and area under the curve (AUC) of 92% on benchmark datasets. The proposed system provides reliable DR staging to assist ophthalmologists in treatment planning and clinical decision-making, potentially reducing diagnostic workload while maintaining high diagnostic accuracy.

Keywords: Diabetic Retinopathy; Convolutional Neural Networks; Computer-Aided Diagnosis; Support Vector Machine; Medical Image Classification; Fundus Image Analysis; Deep Learning

Introduction

Vision is one of the most critical human senses, enabling essential daily activities and quality of life. Diabetic Retinopathy (DR) is a common microvascular complication of diabetes mellitus that, if left untreated, can lead to visual impairment or permanent blindness. The primary pathophysiological mechanism of DR involves progressive damage to retinal blood vessels due to prolonged hyperglycemia (Tilahun, 2019). DR progresses through distinct stages of severity, ranging from mild non-proliferative diabetic retinopathy (NPDR) to severe proliferative diabetic retinopathy (PDR), characterized by varying degrees of microaneurysms, hemorrhages, hard and soft exudates, cotton wool spots, and neovascularization. Accurate staging of DR is essential for timely therapeutic intervention and prevention of vision loss.

DR typically develops 10-15 years after the onset of diabetes mellitus, making early detection critical yet challenging. The primary clinical challenge with DR is that it remains largely asymptomatic in early stages, with patients often unaware of retinal changes until significant, irreversible damage has occurred (Mishra, 2021). By the time symptoms manifest, patients may have progressed to advanced stages where therapeutic options are limited and visual prognosis is poor. This diagnostic delay results in irreversible retinal damage and substantially increases the risk of blindness. Figure 1 illustrates the five-stage International Clinical Diabetic Retinopathy (ICDR) severity scale: (0) no apparent retinopathy, (1) mild NPDR, (2) moderate NPDR, (3) severe NPDR, and (4) PDR.

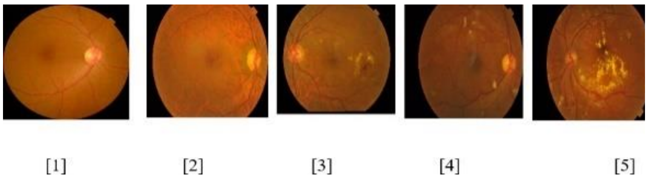

Fig.1: [1] Case 0: No DR [2] Case 1: Mild DR [3] Case 2: Moderate DR [4] Case 3: Severe NPDR [5] Case 4: PDR

Given the asymptomatic nature of early-stage DR, regular ophthalmologic screening of diabetic patients is essential for timely detection and intervention. Current clinical guidelines recommend annual comprehensive eye examinations for all individuals with diabetes. However, the global shortage of ophthalmologists, particularly in developing countries, combined with the increasing prevalence of diabetes worldwide, creates a substantial diagnostic burden that cannot be met through manual screening alone. Therefore, automated computer-aided diagnosis (CAD) systems represent a promising solution to support ophthalmologists in detecting and grading DR accurately, efficiently, and at scale.

DR detection and grading can be formulated as a multi-class image classification problem, where retinal fundus images are categorized into distinct severity stages based on the presence and extent of characteristic lesions, including microaneurysms, hemorrhages, hard exudates, soft exudates, cotton wool spots, neovascularization, and other pathological features. Deep learning (DL) has revolutionized automated DR analysis by offering powerful capabilities in feature extraction, pattern recognition, and classification directly from raw image data (Nazir, 2024). Convolutional Neural Networks (CNNs) have emerged as the dominant architecture for medical image analysis, demonstrating exceptional performance in detecting and classifying DR from retinal fundus photographs.

Deep learning-based approaches offer several advantages over traditional computer vision methods. First, CNNs can automatically learn hierarchical feature representations from raw images without requiring manual feature engineering. Second, deep learning models can capture subtle patterns and complex relationships in retinal images that may be difficult for human experts to articulate explicitly. Third, these models can be trained on large-scale datasets to generalize across diverse patient populations and imaging conditions. Recent advances in DL have also focused on accurate segmentation of retinal structures—including blood vessels, optic discs, microaneurysms, and exudates—which serves as an essential preprocessing step for subsequent classification tasks (Biswas, 2025; Mishra, 2022).

Despite the success of pure CNN-based approaches, hybrid architectures that combine deep learning feature extraction with traditional machine learning classifiers have shown promising results. Specifically, integrating CNNs for automatic feature learning with Support Vector Machines (SVMs) for final classification can leverage the complementary strengths of both approaches: CNNs excel at learning complex hierarchical representations, while SVMs are robust classifiers with strong theoretical foundations and good generalization properties, particularly in scenarios with limited training data.

This study proposes OptiCNN (Optimized CNN), a novel hybrid deep learning framework for automated DR detection and severity grading. The proposed approach integrates three key components: (1) image preprocessing to enhance retinal features and normalize imaging variations, (2) local thresholding-based segmentation to isolate regions of interest, and (3) a hybrid classification architecture combining CNN-based feature extraction with SVM-based classification. The main contributions of this work are as follows:

- Development of an optimized CNN architecture specifically designed for DR classification that balances accuracy and computational efficiency
- Integration of CNN feature extraction with SVM classification to improve generalization and robustness
- Comprehensive evaluation on benchmark datasets demonstrating superior performance compared to existing methods
- Clinical validation of the proposed system's potential to assist ophthalmologists in treatment planning and decision-making

The remainder of this paper is organized as follows: Related Work reviews related work in automated DR detection and deep learning-based medical image analysis. Materials and Methods describe the proposed

OptiCNN methodology, including preprocessing, segmentation, and the hybrid CNN-SVM architecture. Section 4 presents experimental results and comparative analysis with state-of-the-art methods. We discuss the findings, limitations, and clinical implications. Finally, the paper concludes and outlines directions for future research.

Related Work

Over the past decade, researchers have proposed numerous Machine Learning (ML) and Deep Learning (DL) methods for automated detection and classification of diabetic retinopathy. These approaches can be broadly categorized into traditional machine learning techniques, transfer learning-based architectures, hybrid models, and edge-deployed systems. This section reviews representative studies across these categories.

Transfer Learning and Pre-trained CNN Architectures

Transfer learning has emerged as a dominant paradigm in DR classification, leveraging pre-trained convolutional neural networks to address limited annotated medical imaging data. Jabbar et al. (2022) developed a customized model based on VGG-16 architecture with transfer learning for DR recognition. The authors employed data augmentation techniques including rotation, flipping, and scaling to improve the model's generalization capability. While their approach demonstrated competitive performance, the evaluation was conducted on a relatively small dataset, which may limit generalizability to diverse clinical populations.

Abhini and Devi (2025) employed an Inception V3-based transfer learning framework to classify retinal fundus images into five distinct severity categories. The model architecture integrates the pre-trained Inception V3 as a foundational feature extractor with additional stacked layers to enhance representational learning and classification performance. The approach achieved an accuracy of 91.34% and precision of 90.6%, demonstrating the effectiveness of hierarchical feature learning in DR classification.

Yaqoob et al. (2021) combined ResNet-50 for deep feature extraction with a Random Forest classifier for final classification, achieving 75.09% accuracy. This hybrid approach attempts to leverage the feature learning capabilities of deep networks with the interpretability of traditional machine learning classifiers. However, the relatively lower accuracy suggests potential limitations in the integration strategy or dataset characteristics.

Vision Transformers and Hybrid CNN-Transformer Models

Recent advances in computer vision have introduced Vision Transformers (ViTs) as an alternative to traditional CNNs, offering enhanced capability to capture global contextual dependencies through self-attention mechanisms. Liu et al. (2025) proposed a CNN-Vision Mamba model that seamlessly integrates a customized attention mechanism into the conventional Vision Transformer framework. This attention-enhanced architecture demonstrates superior capability in extracting rich feature embeddings and effectively capturing global contextual dependencies within retinal fundus images. The model achieved an accuracy of 90.6% and precision of 90.5%, highlighting the potential of attention-based architectures for medical image analysis.

Amna et al. (2025) introduced ResViT FusionNet, a hybrid model that integrates Convolutional Neural Networks with Vision Transformers to leverage the complementary strengths of both architectures. CNNs excel at extracting local spatial features through convolutional operations, while ViTs capture long-range dependencies through self-attention mechanisms. This fusion enhances feature representation and global contextual understanding, achieving an impressive accuracy of 93.01% in DR classification tasks. However, the combination of CNN and ViT architectures significantly increases computational complexity, which may limit deployment in resource-constrained clinical settings.

Feature Selection and Optimization Approaches

Traditional machine learning methods combined with advanced feature selection and optimization techniques have also been explored for DR classification. Gundluru et al. (2022) presented a model based on Principal Component Analysis (PCA) for dimensionality reduction and the Harris Hawks Optimization (HHO) algorithm for feature selection and optimization. The optimized features were used to classify DR using the UCI Machine Learning Repository dataset. While this approach achieved reasonable performance, it relies on hand-crafted feature engineering, which may not capture the complex hierarchical patterns in retinal images as effectively as end-to-end deep learning methods.

Preprocessing and Hybrid Classification Approaches

Several studies have emphasized the importance of preprocessing and hybrid classification strategies. Swapna et al. (2022) employed Contrast Limited Adaptive Histogram Equalization (CLAHE) as a preprocessing technique to enhance retinal image contrast, followed by deep learning-based feature extraction and Support Vector Machine (SVM) classification. The proposed method achieved satisfactory results compared to contemporary state-of-the-art approaches, demonstrating that appropriate preprocessing can significantly improve classification performance by enhancing relevant features and suppressing noise.

Kipli et al. (2018) adopted a multi-stage pipeline combining preprocessing, segmentation, and deep learning classification. Their methodology first enhanced image quality, then segmented regions of interest such as blood vessels, exudates, and microaneurysms, and finally applied a convolutional neural network for classification. This approach achieved accuracy metrics exceeding 90%, highlighting the value of explicit segmentation in improving DR detection performance.

Multi-Stage and Lesion-Based Classification

Recognizing that DR severity is determined by the presence and extent of specific retinal lesions, several researchers have developed multi-stage frameworks that explicitly identify pathological features before classification. Silva et al. (2021) proposed a multi-stage deep learning framework for DR grading consisting of lesion segmentation followed by severity classification based on detected lesions. Kaushik et al. (2025) developed a residual network-based deep learning framework that achieved an overall accuracy of 83% with a precision of 95% for identifying the absence of diabetic retinopathy, demonstrating particular effectiveness in ruling out non-DR cases. However, both approaches were evaluated on relatively small and imbalanced datasets, which may limit their generalizability.

Hybrid Detection and Localization Systems

Recent work has explored combining classification with object detection to provide both DR severity grading and spatial localization of lesions. Alyoubi et al. (2021) integrated a CNN-based classifier for overall DR grading with YOLOv3 (You Only Look Once version 3) for lesion detection and localization. This dual-task approach achieved 89% classification accuracy while simultaneously providing clinicians with visual explanations through bounding boxes around detected lesions, enhancing model interpretability and clinical trust.

Domain Adaptation and Cross-Dataset Generalization

Addressing the challenge of model generalization across different datasets and imaging conditions, Zhang et al. (2023) developed a Multi-Model Domain Adaptation (MMDA) framework with transfer learning using weighted pseudo-labeling and clustering-based approaches. The model achieved 90.6% accuracy on the APTOS dataset, demonstrating the potential of domain adaptation techniques to improve model robustness across diverse clinical settings and imaging protocols.

Table 1. Comprehensive Review of DR Classification Approaches

Study	Architecture	Dataset	Accuracy	Precision	Key Strengths	Main Limitations
Jabbar et al. (2022)	VGG-16 + Transfer Learning	Small dataset	90.6%	90.5%	Effective data augmentation; proven architecture	Small dataset; limited generalizability
Gundluru et al. (2022)	PCA + HHO + ML	UCI ML Repository	91.64%	N/A	Optimized feature selection	Hand-crafted features; limited scalability
Swapna et al. (2022)	CLAHE + DL + SVM	Not specified	Satisfactory	N/A	Effective preprocessing; hybrid approach	Limited performance details reported
Kipli et al. (2018)	Preprocessing + Segmentation + DL	Small dataset	>90%	N/A	Multi-stage pipeline; explicit segmentation	Small dataset; computationally intensive
Silva et al. (2021)	Lesion Segmentation + ResNet	Small, imbalanced	83%	N/A	Lesion-aware approach	Small, imbalanced dataset
Proposed (OptiCNN)	**CNN + SVM Hybrid**	**Benchmark datasets**	**94%**	**96%**	**High accuracy; computational efficiency; clinical applicability**	**Requires preprocessing pipeline**

Edge Deployment and Resource-Constrained Applications

To enhance accessibility and enable point-of-care diagnosis, several studies have explored deploying DR classification models on edge devices. Silpa et al. (2025) implemented a DenseNet model on a Raspberry Pi 4—a compact, cost-effective, and widely available computing platform. This implementation achieved a classification accuracy of 88%, demonstrating the model's potential for real-world, resource-constrained applications such as remote screening programs in underserved areas. While edge deployment offers significant advantages in accessibility and cost, it often requires careful model optimization and compression to balance accuracy with computational constraints.

Summary and Comparative Analysis

Table 1 summarizes the key characteristics, strengths, limitations, and performance metrics of representative studies in automated DR detection and classification. Several important trends emerge from this review. First, there is a clear progression toward more sophisticated architectures, from traditional CNNs to hybrid CNN-Transformer models, with corresponding improvements in accuracy. Second, transfer learning and pre-trained models have become the dominant approach, addressing the challenge of limited annotated medical imaging data. Third, hybrid approaches combining deep learning with traditional machine learning classifiers (e.g., CNN+SVM, ResNet+Random Forest) offer a balance between automatic feature learning and model interpretability.

However, several limitations persist across existing approaches. Many studies evaluate their models on single datasets, raising concerns about generalization to diverse clinical populations and imaging conditions. The most accurate models, particularly hybrid CNN-Transformer architectures, suffer from high computational complexity that limits real-time deployment and edge computing applications. Furthermore, most studies focus solely on classification accuracy without adequately addressing model interpretability, uncertainty quantification, and clinical workflow integration—all critical factors for clinical adoption.

Materials and Methods

The base model, OptiCNN, employed in this research includes image preprocessing with gray scale conversion, resizing, and segmentation by adaptive mean thresholding. OptiCNN, shown in Figure 2, which includes a sequence of key stages:

1) **Image Preprocessing:** Gray scale conversion followed by resizing to fix the input image size for the system.
2) **Segmentation:** Local Thresholding (adaptive mean thresholding) for refining the results.
3) **CNN with SVM Classifier:** CNN uses four layers, and the SVM classifier is used on the output layer.

After that, the CNN architecture is applied with subsampling layers, complete convolution layers, and an SVM classifier. CNN can automatically extract the relevant features from retinal images and classify them into five stages of DR. However, the OptiCNN model enhances the accuracy with both CNN and SVM benefits. Finally, the detection of DR in the retinal fundus images is computed based on the classification results. The architecture of OptiCNN includes two main phases. Image preprocessing and segmentation are performed to eliminate redundant information and obtain the RoI from the image. Preprocessing of images was further scaled down to reduce the image size, and segmentation effectively extracts the background from the foreground and segments the images. In the second phase, CNN is employed as four convolution layers with ReLU function, and at the output layer, SVM uses a one-vs-rest multiclass classification approach. Subsequently, retinal images are classified into five stages of DR (Aslam T., 2021). This process gives better accuracy in grading the detected.

DR in retinal images. The proposed method in the research aims to build an automated system, i.e., OptiCNN, that performs the analysis and classifies DR using retinal fundus images. Figure 2 represents the method that involves the given key steps:

- **Data Loading:** We have used the retinal fundus images as input data from the traditional reference database, EyePACS available via Kaggle, which are marked with the extent of DR.

- **Image Pre-processing:** Image preprocessing is performed by gray scale conversion, followed by resizing the images, which brings the images to the fixed exact size dimensions.

- **Segmentation Approaches:** Segmentation is done to obtain the areas of interest in the images of the retina. A local thresholding method is used to enhance the segmentation of the regions of the retina that are concerned with DR.

- **Training and Test Data Creation:** Here in the study, the dataset is categorised into train and test subsets according to the severity levels of DR in the proportion of 80% and 20% for training and testing purposes. This separation makes it possible to evaluate the effectiveness of the models in estimating the severity of DR.

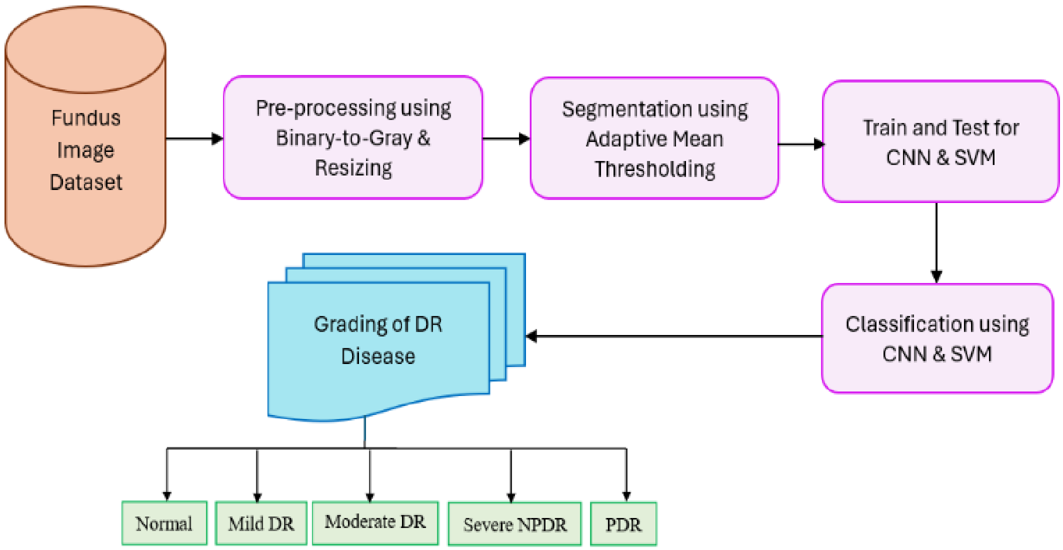

Fig. 2: Functional Pattern of the OptiCNN

- **CNN with SVM Classifier:** We have used the convolutional neural network (CNN) in the feature extraction process. They describe their model as a CNN that utilises four layers, and the final layer operates using the SVM classifier with a one-vs-rest approach.
- **Grading Computation:** The detected DR in retinal fundus images is graded based on the obtained classification results. This grading system helps make quantitative estimates of DR and helps clinicians decide on the cure and management of pathological conditions.

DataSet Used

Fundus images of the retina have been publicized and shared on the internet in several datasets. These datasets make DR detection possible, includi ng fundus images of patients diagnosed with DR. The retinal image sets are essential for testing and preparation for computer-aided disease diagnosis. These diagnostic systems involve sophisticated digital screening programs and a computerized algorithm.

In this study, we used an open-source dataset, EyePACS, which is publicly available on Kaggle. The dataset has a total number of 35126 retinal fundus images, which were utilised to train the proposed model. It comprises 5 Classes: No DR, Mild DR, Moderate DR, Severe NPDR and PDR. The bifurcation of classes in the dataset is listed out in Table 2.

Due to the imbalanced distribution of DR severity levels in the dataset, data resampling was applied. Specifically, SMOTE was used to synthetically generate additional samples for the minority classes synthetically, ensuring a more balanced class representation. This helped improve classification performance and reduce model bias toward the 'No DR' class.

Table 2: Details of Fundus Image Dataset

Class Label	Stage of DR	Image Count
0	No DR	25376
1	Mild DR	2495
2	Moderate DR	5476
3	Severe NPDR	1013
4	PDR	765

Out of these images, 80% (i.e., 28100 images) were used for training, while 20% (i.e., 7026) were used for testing purposes.

Image Preprocessing

This section aims to implement functions on the fundus images collected for the model. These functions ensure the images are in better condition so the system can read them correctly. Two image preprocessing techniques have been applied to make the input images suitable for subsequent processing.

Gray Scale conversion: This process converts an image from other colour spaces (e.g., RGB, CMYK, HSV) into different shades of gray, which range between completely white and black, using the equation (1):

$$\text{Gray Scale} = 0.299\ R + 0.587\ G + 0.114 \qquad 1)$$

Resize the image: This process reduces the size of fundus images. Each image with a high length and width can be reduced to a fixed size. The available fundus image dataset is of size 1024 X 1024. The resize function reduces all the images to a fixed size (32 X 32), so that the whole dataset reaches the same condition (Vincet L. 1993). Figure 3 presents the resulting images obtained after applying the above-mentioned preprocessing steps.

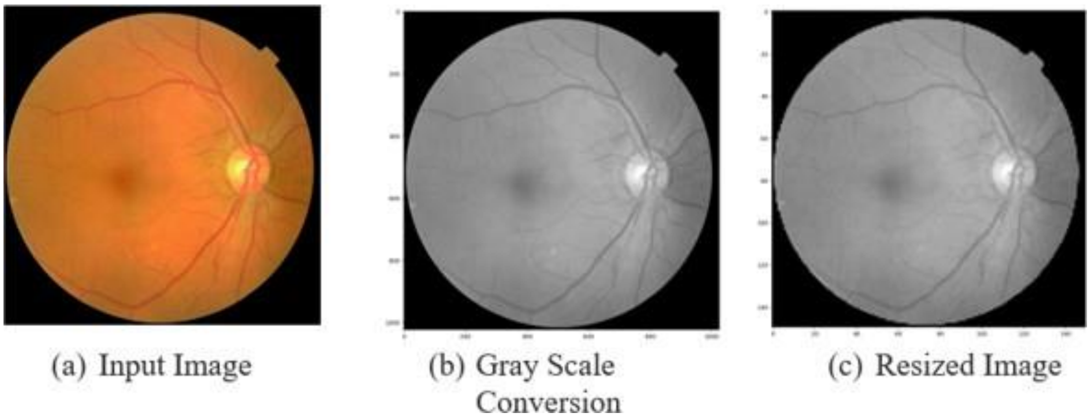

Fig. 3: Images after applying Preprocessing

Segmentation

The primary objective of image segmentation in computer vision is to partition an input image into different homogeneous regions, separating foreground and background objects. Here, Local thresholding has been applied to filter out the noise from images to extract the Region of Interest(RoI) (Jaffery ZA, 2013). Local thresholding determines the threshold value in the vicinity of each pixel based on the features of the adjacent pixels (Eckhardt U.,2013). Local thresholding is basically of two types: (a) Adaptive Mean Thresholding (AMT), (b) Adaptive Gaussian Thresholding (AGT). In this study, we used the former method to segment the image, as shown in Figure 4.

The AMT technique was chosen owing to its simplicity and low computational cost compared to the AGT technique, which gives it an edge over large datasets. In AMT, the threshold value is calculated as the mean of the pixel values in the local neighbourhood (Mandal AK, 2019).

The final threshold value is calculated by subtracting the constant value from the mean as given in equation (2):

$$T(x, y) = \mu(x, y) - C \quad 2)$$

Where, (x,y) are the pixel coordinates in the image, T(x,y) is the threshold value, μ(x,y) is the mean of the pixel intensities calculated within the local neighbourhood.

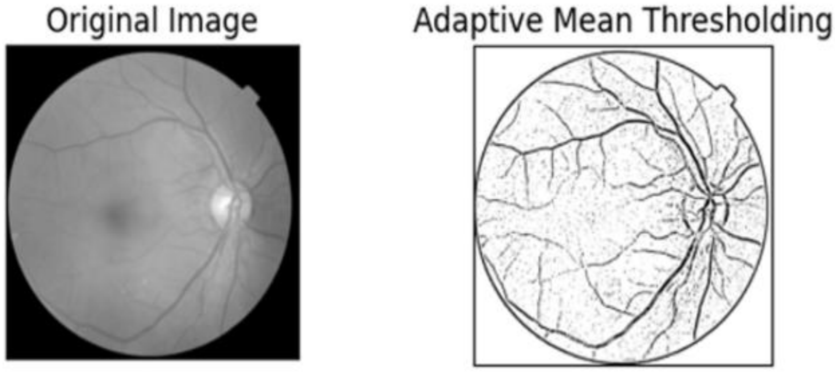

Fig. 4: Image after segmentation technique

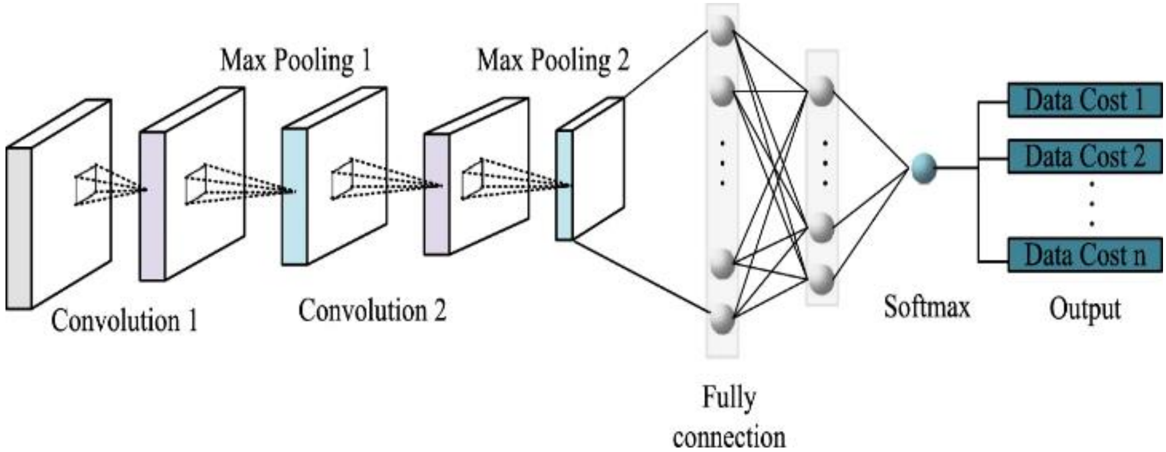

Fig. 5: Structure of CNN

OptiCNN: CNN integration with SVM

In OptiCNN, the CNN architecture is used along with SVM to detect DR stages. It utilises the benefits of CNN and SVM for classifying distinct cases from 0 to 4. Figure 5. represents the stages of the OptiCNN model. It showcases convolutional layers, subsampling layers and a fully connected layer that integrates with SVM as the output layer. Before explaining the proposed methodology in detail, a basic CNN model is discussed: Be aware that they are part of the ensemble. This allows for seamless integration of various models into a unified framework.

Convolution Neural Network (CNN)

A Convolutional Neural Network (CNN) is an Artificial Neural Network (ANN) type that incorporates multiple hidden layers between the input and output layers, specifically designed to extract spatial hierarchies of features from input data. CNN has various hidden layers such as convolutional layers, max pooling layer, dense layer, flatten layer, fully connected layer and other unique layers as shown in Figure 5.

Each convolutional layer consists of multiple filters, or kernels, which act as small sliding windows traverse the input data to extract local features (Duan KB, 2005).

The first layer is the convolutional layer, the weights of these filters are learned, function as feature extractors, and extract features from the image. During this process, the property of each pixel and the relation of that pixel with neighbour**ing** pixels **are** calculated with the help of existing mathematical operations (Bazgir, 2020). After extracting features, the necessary information is stored by neglecting useless data through pooling layers (Max Pooling, Avg Pooling, Sum Pooling). Pooling layers are incorporated following convolutional layers to downsample the spatial dimensions of feature maps, thereby reducing computational complexity and enhancing feature abstraction (Angulo C, 2000). This process is called subsampling, in which all significant features are mapped. Sometimes, the feature mapping process leads to an overfitting state. The flatten layer converts 2D to 1D arrays before the fully connected layer. The CNN model uses a fully connected layer as its final layer. Combined with activation functions, these fully connected layers perform tasks such as classification or regression on the features extracted by earlier layers.

Activation functions are essential in CNNs because they introduce nonlinearity by being applied to each neuron's output, allowing the network to learn complex relationships between input data and their feature0. Within CNNs, several commonly used activation functions include Rectified Linear Units (ReLU), Leaky ReLU, Parametric ReLU, and softmax (Hsu C W, 2002).

GPU-based architecture is best suited to the convolutional neural networks as it takes less time to train than classical CPUs.

Support Vector Machine (SVM)

Support Vector Machine (SVM), proposed by Vapnik, is a kernel-based machine learning model employed for Support Vector Machine (SVM), proposed by Vapnik, is a kernel-based machine learning model used for classification and regression tasks. SVMs are especially effective in binary classification problems. The model constructs its decision function directly from the training data by maximising the margin between decision boundaries in a high-dimensional feature space. This margin maximisation minimises classification errors on the training set and enhances the model's generalisation ability, improving its predictive performance (Shyngyskhan, 2024).

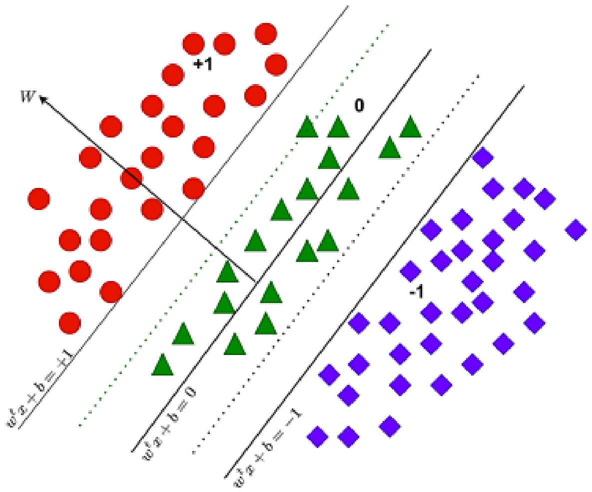

Fig. 6: K-class SVM

SVMs and their variants are inherently designed for binary classification problems, whereas real-world applications often involve multiclass solutions. A widely adopted method to address this limitation is the divide-and-combine strategy, wherein the outputs of independently trained binary classifiers are aggregated to form the final decision. Two of the most popular strategies within this approach are "One-versus-Rest" (OvR), also known as "One-versus-All" (OvA), and "One-versus-One" (OvO), which is also referred to as "All-versus-All." (Ahmed I, 2021) (Jiang Z, 2021). This concept leads to a novel approach known as the K-class Support Vector Classification (SVC) method for addressing K-class classification challenges represented in Figure 6 (Younesi A, 2024).

This study uses SVM for multiclass classification via the one-vs-rest approach. The SVM classifier is placed at the end of a fully connected layer of CNN to use the advantages of both methods, improve the model's efficiency, and increase the model's performance.

(One-vs-All)

For each class i (where i = 1, 2,……, C), SVM classifier has been trained:

$$\min_{w_i,b_i} \frac{1}{2}\|w_i\|^2 + C\sum_{j=1}^{n}\xi_j \tag{7}$$

Subject to:

$$y_j(w_i^T x_j + b_i) \geq 1 - \xi_j, \quad \forall j \tag{8}$$

Where:

- x_j = Input feature vector
- y_j = Class label (mapped as +1 for class i, -1 for others)
- w_i, b_i = Parameters for class i
- ξ_j = Slack variable for misclassification
- C = Regularization parameter

For a new data point x, the predicted class is:

$$\hat{y} = \arg\max_i (w_i^T x + b_i) \tag{10}$$

Multiclass SVM (Direct Formulation), another approach solves multi-class SVM directly using the following optimisation:

$$\min_{W,b} \frac{1}{2}\sum_{i=1}^{C}\|w_i\|^2 + C\sum_{j=1}^{n}\sum_{i\neq y_j}\max(0, 1 + w_i^T x_j - w_{y_j}^T x_j) \tag{11}$$

Where:

- W = Set of weight vectors $w_1, w_2, \ldots, w_C$ for all classes.

The Hinge Loss term ensures that for each training example, its correct class has a higher score than other classes by at least a margin of 1.

OptiCNN: CNN with SVM

The framework of the OptiCNN model is designed to enhance accuracy for the betterment of the healthcare system. An SVM model restores the last layer of the CNN. Therefore, the output of the fully connected layer of CNN is provided to the SVM model as an input for classification. The primary goal of combining CNN and SVM is to add both models' advantages and enhance the architecture's accuracy and performance. This combination has exceptional feature extraction characteristics, providing superior efficacy amongst the existing algorithms. In this OptiCNN: CNN + SVM architecture, different convolutional and subsampling layers are used for feature mapping. These mapped features are then provided to an SVM as input, and the feature vectors are trained, classified, and predicted for the class/stage of DR.

Proposed model: CNN with SVM

This section demonstrates the proposed model for classifying DR disease into five stages, wherein SVM is used as the final classification layer. The theoretical foundation of the basic CNN model is provided in the above section. This study experimented with Python 3.6 using Keras libraries to build the suggested CNN model. In addition, Tensorflow and Theano were used as a backend in the Keras API (Cervantes J, 2020). The proposed model's architecture, OptiCNN, is given in Figure 7, in which CNN is provided with segmented fundus images as input to the input layer. After feeding the data into the input layers, the data was processed for feature extraction by two layers. Subsequently, the processed data is input to the final SVM classification layer for final prediction. The image provided to the input layer was resized to 32 X 32. In the input layer, a kernel size 5 X 5 filter is applied for feature mapping and extraction, resulting in a feature map of 6 @ 14 X 14. Second convolutional layer (CL 2) uses a 5 X 5 kernel size filter, which produces a 10 X 10 spatial dimension in each layer, resulting in 16 @ 10 X 10. Next step is to apply the max pooling function of 2 X 2, which maps the feature in terms of 16 @ 5 X 5. These mapped features are then provided to the Fully Connected layer (400) through a flattened one-dimensional channel, eventually feeding the obtained data into SVM for final classification into five classes using the one vs rest technique. The one-vs-rest technique allows the classifier to employ SVM to yield the desired result in five classes (Hussain SF, 2023). The tuning of hyperparameters for OptiCNN given in Table 2.

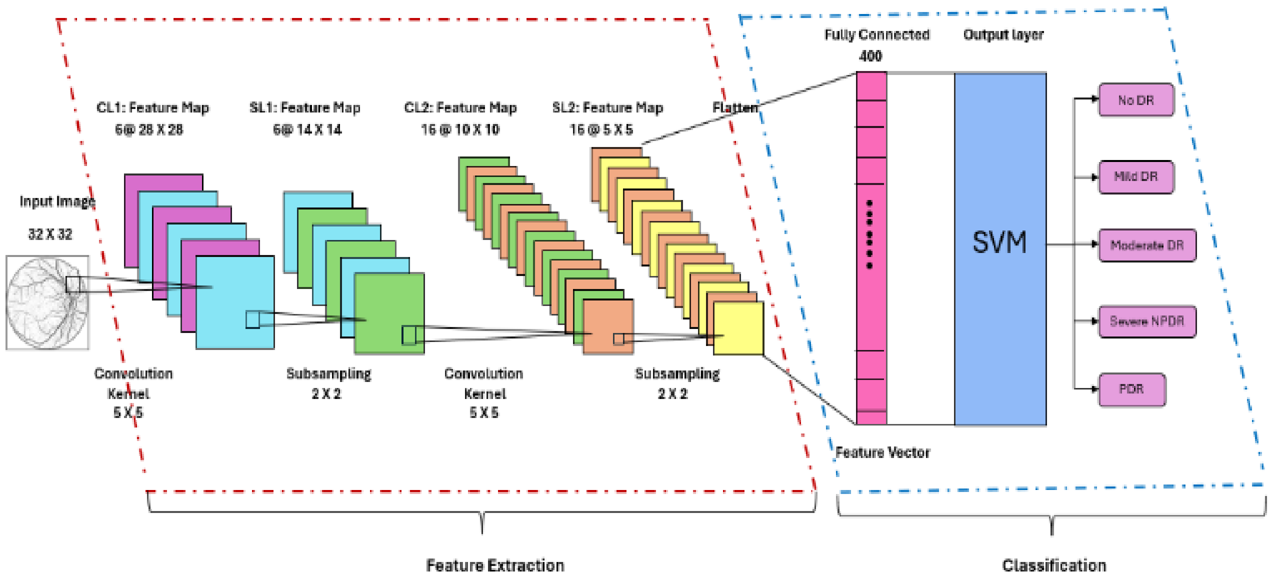

Fig. 7: Structure of OptiCNN

Table 3: Hyperparameter Tuning

Hyperparameter	Description	Values
Optimizer	The optimisation algorithm used during training	RMS Prop
Learning Rate	The step size for revising the model parameters during training	0.03
Batch Size	Count of samples used in each training iteration	10
Weight Decay	Avoid overfitting by regularising the parameters	0.01
Dropout Probability	The probability of dropping out a neuron	0.5, 0.5, 0.2

To evaluate the model's generalisation performance and ensure reliable results, 10-fold cross-validation was utilised during the training process. The dataset was divided into 10 equal parts, where each part. Iteration, 9 parts were used for training and 1 part for validation (Chakraborty S, 2024). This process was repeated 10 times, and the average performance metrics across all folds were considered the final evaluation (Chengfeng C, 2023).

Results and Discussions

OptiCNN is performed on the system with (9th Generation Core i7 CPU, 16 GB RAM, Nvidia GTX 1650 GPU). Retinal images used in OptiCNN to simulate the severity of DR estimation simulate the grading process in DR. The fine details of tuning of hyperparameters for OptiCNN are guided in Table 3, exclusively devised for DR. RMS Prop optimiser is used in this model with 0.03 step size. Each training iteration is provided with a batch of 10 samples. The neurons' dropout probability is considered 0.5 and 0.2 during training. Activation function, ReLU, is used in the neural network layers, and the one-vs-rest method is used for multiclass classification of SVM on the output layer.

Evaluation Parameters

This study uses the following parameters to measure the performance of OptiCNN (Leopold HA, 2017).

(a) Accuracy:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (3)$$

In a binary classification problem with only two classes (such as positive and negative), accuracy is computed using the following formula:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}} \quad .(4)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

(b) Precision:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \ldots\ldots..(7)$$

(c) Recall:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \quad \ldots\ldots.(8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \ldots\ldots(9)$$

(d) F1- Score:

The F1-Score, harmonic mean, is a type of average calculated by taking the reciprocal of the average of the individual values, making it particularly suitable when balancing two competing metrics (Lin G, 2018).

$$\text{F1-Score} = \frac{\text{2X Precision X Recall}}{\text{Precision+Recall}} \quad \ldots...(10)$$

Evaluated Results

The results presented in Table 4 demonstrate that OptiCNN gives better results in terms of accuracy than the existing state-of-the-art methods.

Owing to use of one-vs-rest approach, underlying problem has become the binary classification problem [40], where each class was assigned a level i.e, Case 0-> No DR, Case 1-> Mild DR, Case 2-> Moderate DR, Case 3-> Severe NPDR, Case 4-> Proliferative DR. Details of the assigned levels are given as follows:

i. Case 0 vs (Rest of the Cases).
ii. Case 1 vs (Rest of the Cases).
iii. Case 2 vs (Rest of the Cases).
iv. Case 3 vs (Rest of the Cases).
v. Case 4 vs (Rest of the Cases).

The overall result in terms of accuracy, precision, recall, and F1 score is listed in Table 4. Figure 8 represents the confusion smatrix of the model.

Table 4: Evaluated Results of OptiCNN

Class Label	Class Name	Precision	Recall	F1- score	Accuracy
0	No DR	0.99	0.92	0.96	0.95
1	Mild DR	0.91	0.90	0.91	0.92
2	Moderate DR	0.99	0.91	0.95	0.97
3	Severe Non-Proliferative DR	0.95	0.89	0.92	0.93
4	PDR	0.94	0.90	0.92	0.93
Average		**0.96**	**0.91**	**0.93**	**0.94**

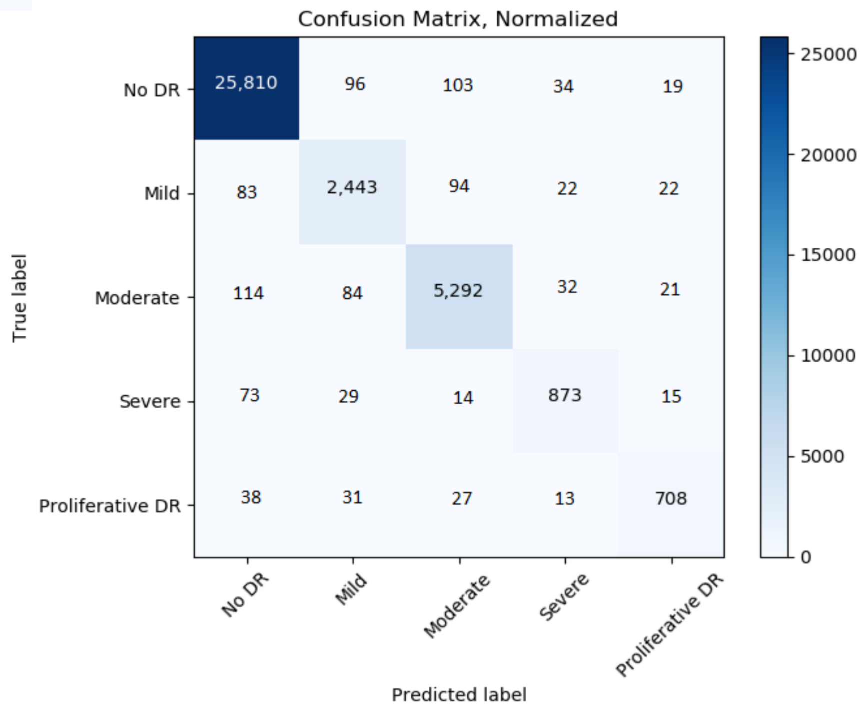

Fig. 8: Confusion Matrix

In general, the performance metrics, viz., precision, recall, F1-score, and accuracy, are preferred to assess the performance of ML and DL models. Hence, we also used them as given in Table 4. The results of other state-of-the-art methods for the same problem are provided in Table 5. All the research work has been performed on the EyePACS dataset.

The obtained result reveals that the proposed method, OptiCNN, achieved promising results across all metrics. It exhibits a sensitivity of 0.96, a recall of 0.91, an F1-score of 0.93, an AUC of 0.92 and an accuracy of 0.94. These high values indicate that OptiCNN is capable of correctly identifying positive instances (high precision) while avoiding false positives. The overall accuracy of 0.94 suggests that the method performs exceptionally well in classifying both positive and negative instances.

Table 5: Comparative Analysis of OptiCNN with existing research

Ref. No.	Precision	Recall	F1-score	Accuracy
Proposed	**0.96**	**0.91**	**0.93**	**0.94**
[24]	0.87	0.87	0.87	0.85
[25]	0.67	0.56	0.61	0.85
[29]	0.67	0.68	0.68	0.93
[30]	0.76	0.73	0.75	0.86

Table 6 shows the comparative analysis of SVM, CNN and OptiCNN accuracy. OptiCNN has achieved an accuracy of 94.2 % for the classification of DR, which is high compared to SVM and CNN. Separately, SVM and CNN obtained 76.8% and 90.3 %. Figure 9 represents comparative graphs for SVM, CNN and OptiCNN accuracy. On observing the graphs closely, it is clearly visible that SVM and CNN are not as progressed as the suggested model, OptiCNN is. Graphical comparison of SVM, CNN and OptiCNN precision is shown in Figure 10, demonstrating the superiority of OptiCNN over the other methods. In the closer view of SVM and CNN, the results are not satisfactory and should be worked on to improve. The combination of CNN and SVM, i.e., OptiCNN, adds advantages of both SVM and CNN. Hence, the proposed model, OptiCNN, gives better classification outputs. Figure 11 reveals that OptiCNN gives a high rate of accuracy, which means better output can be provided in comparison to other existing algorithms.

Table 6: Accuracy comparison of SVM, CNN and OptiCNN

Label	Class Name	SVM	CNN	OptiCNN
0	No DR	82.2	93.1	95.1
1	Mild DR	79.2	90.1	92.2
2	Moderate DR	75.1	91.2	97.3
3	Severe NonProliferative DR	72.7	88.2	93.2
4	Proliferative DR	74.6	89.1	93.4
	Average	**76.8**	**90.3**	**94.2**

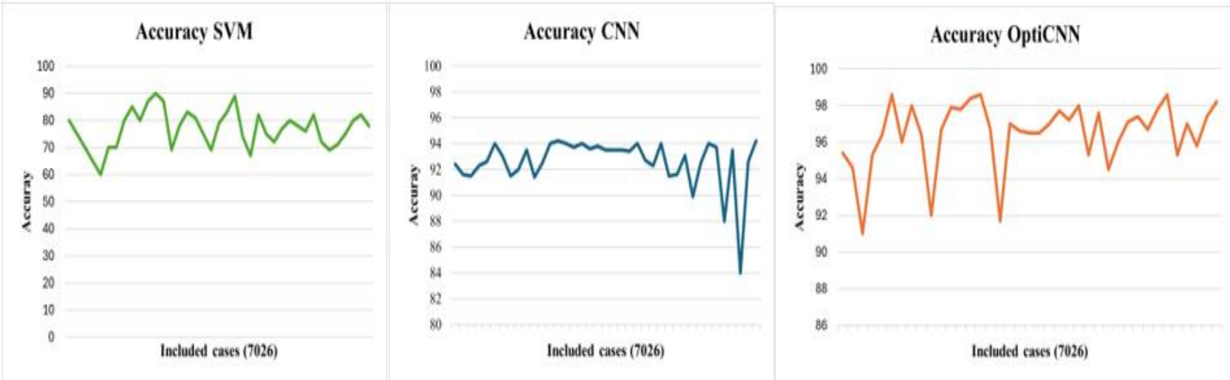

Fig. 9: Accuracy comparison of SVM, CNN and OptiCNN

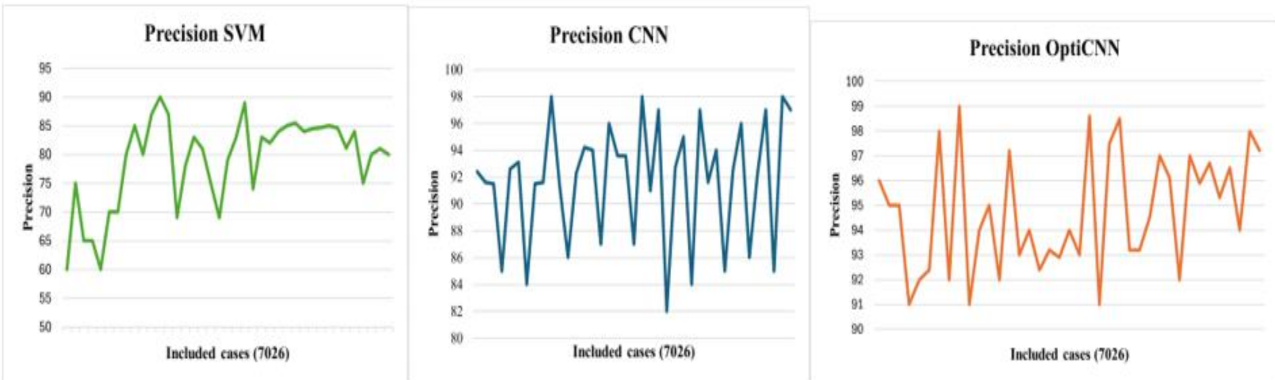

Fig. 10: Precision Comparison of SVM, CNN and OptiCNN

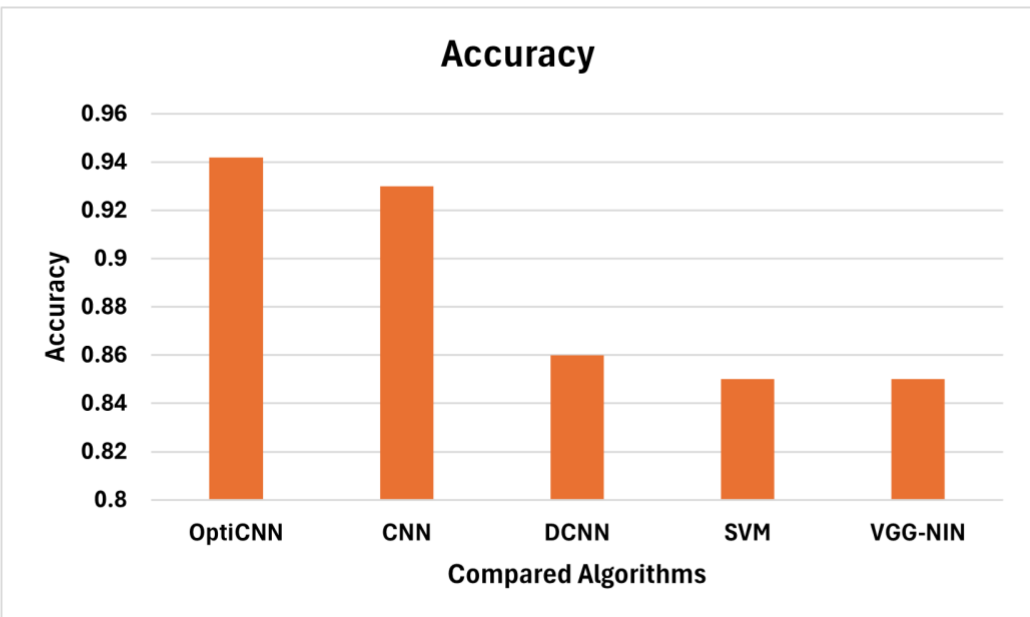

Fig. 11: Comparison of OptiCNN with existing models

Conclusion

In this research paper, the author proposes OptiCNN for DR classification and grading based on image preprocessing, segmentation, and a deep neural network using an SVM approach. This study shows that all elements of the OptiCNN yielded high classification performance for DR. The results revealed high precision, recall, F1-score and accuracy values in the context of OptiCNN, which determined its efficient performance in identifying both positive and negative cases of DR sufferers. The performance assessment for each stage of DR also re-established that it is a practical approach in enhancing the identification of positive and negative cases. The study also dealt with DR grading based on the classification output with the conventional grading system. This helps clinicians to have proper severity evaluation for proper management, helping develop diagnosis/management in this field.

In conclusion, the research paper provides future directions and methods to enhance the classification and grading of DR. OptiCNN achieved an accuracy of 94.2 %, whereas SVM and CNN obtained 76.8% and 90.3% respectively. These results show that OptiCNN improves the immunity and versatility of the models, offers precise localisation and segmentation of the damaged area, refines the choice of the discriminative features, and increases classification efficacy using the combination of a convolutional neural network and an SVM model. About these findings, the current work provides significant enhancements to diagnosing and managing DR that will aid patients and other professionals in the health sector.

Future Work

With the rapid advancements in deep learning, diabetic retinopathy (DR) detection has significantly improved in recent years. However, there remains considerable scope for enhancement. Emerging architectures such as Vision Transformers (ViT), EfficientNet, and hybrid CNN-transformer models offer improved representation learning capabilities. Their integration could lead to more accurate and robust DR detection systems. Implementing federated learning can enable collaborative model training across multiple healthcare institutions without sharing sensitive patient data, thereby enhancing model generalisation while ensuring data privacy. Incorporating explainability techniques such as Grad-CAM, SHAP, or attention visualisation will help clinicians understand the reasoning behind model predictions. This fosters trust and promotes adoption in clinical practice. Combining fundus images with patient metadata (e.g., age, HbA1c levels, blood pressure) or other imaging modalities like OCT can lead to a more comprehensive and accurate diagnosis of DR. Developing lightweight, real-time models optimised for edge devices or mobile platforms can facilitate affordable DR screening in rural and under-resourced regions. Future models should undergo rigorous validation in real-world healthcare environments to ensure clinical applicability. Collaboration with medical professionals and clinical trials will be critical in assessing reliability and bias reduction.

Acknowledgement

Authors express sincere gratitude to Dr. Mrinal Pandey for their invaluable guidance, support, and encouragement throughout this research. We are also thankful to Manav Rachna University, Faridabad, for providing the necessary resources and facilities to carry out this work.

References

Abini, M. A., & Sathya Priya, S. S. (2025). Detection and classification of diabetic retinopathy using modified Inception V3. *International Journal of Bioautomation, 29*(1), 77–92. https://doi.org/10.7546/ijba.2025.29.1.001004

Ahmed, I., Hammad, B. T., & Jamil, N. (2021). A comparative analysis of image copy-move forgery detection algorithms based on hand and machine-crafted features. *Indonesian Journal of Electrical Engineering and Computer Science, 22*(2), 1177–1190. https://doi.org/10.11591/ijeecs.v22.i2.pp1177-1190

Alyoubi, W. L., Abulkhair, M. F., & Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors, 21*(11), Article 3704. https://doi.org/10.3390/S21113704

Angulo, C., & Català, A. (2000). K-SVCR: A multi-class support vector machine. In *Machine learning: ECML 2000* (pp. 31–38). Springer.

Aslam, T., Fleck, B., & Patton, N. (2021). Multidisciplinary approach for diabetic retinopathy care. *The Lancet Diabetes & Endocrinology, 9*(8), 453–454. https://doi.org/10.1016/S2213-8587(21)00166-7

Bazgir, O., et al. (2020). Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. *Nature Communications, 11*(1), Article 4391. https://doi.org/10.1038/s41467-020-18157-4

Biswas, A., & Banik, R. (2025). Deep learning-based image segmentation for early detection of diabetic retinopathy and other retinal disorders. In *Deep learning applications in medical image segmentation* (pp. 133–148).

Cai, C., et al. (2023). Improved deep convolutional neural networks using chimp optimization algorithm for COVID-19 diagnosis from X-ray images. *Expert Systems with Applications, 213*, 119206.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing, 408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Chakraborty, S., & Dey, L. (2024). Multi-class classification. In *Multi-objective, multi-class and multi-label data classification with class imbalance* (pp. 51–76). Springer Nature.

Duan, K. B., & Keerthi, S. S. (2005). Which is the best multiclass SVM method? An empirical study. In *Multiple classifier systems* (pp. 278–285). Springer. https://doi.org/10.1007/11494683_28

Eckhardt, U. (2013). Root images of median filters. *Journal of Mathematical Imaging and Vision, 19*, 63–70.

Gundluru, N., Singh Rajput, D., Lakshmanna, K., Kaluri, R., Shorfuzzaman, M., Uddin, M., & Khan, M. A. R. (2022). Enhancement of detection of diabetic retinopathy using Harris Hawks optimisation with deep learning model. *Computational Intelligence and Neuroscience, 2022*, Article 8512469. https://doi.org/10.1155/2022/8512469

Hsu, C.-W., & Lin, C.-J. (2002). A simple decomposition method for support vector machines. *Machine Learning, 46*, 291–314. https://doi.org/10.1023/A:1012427100071

Ikram, A., & Imran, A. (2025). ResViT FusionNet model: An explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images. *Computers in Biology and Medicine, 186*, 109656. https://doi.org/10.1016/j.compbiomed.2025.109656

Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z., & Jabbar, A. (2022). Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sciences, 12*(5), Article 535. https://doi.org/10.3390/brainsci12050535

Jaffery, Z. A., Zaheeruddin, & Singh, L. (2013). Performance analysis of image segmentation methods for the detection of masses in mammograms. *International Journal of Computer Applications, 82*(2), 44–50.

Kaushik, K., Bhardwaj, A., Cheng, X., Dahiya, S., Shankar, A., Kumar, M., & Mehrotra, T. (2025). Residual network-based deep learning framework for diabetic retinopathy detection. *Journal of Database Management, 36*(1), 1–21. https://doi.org/10.4018/JDM.368006

Kumar, S. A., Kumar, J. S., & Mahabaleswara, S. C. B. (2025). Efficient diabetic retinopathy detection using deep learning approaches and Raspberry Pi 4. *Bulletin of Electrical Engineering and Informatics, 14*(2), 1063–1072. https://doi.org/10.11591/eei.v14i2.8248

Kuryati, K. (2018). *Morphological and Otsu's thresholding-based retinal blood vessel segmentation for detection of retinopathy*. Science Publishing Corporation.

Leopold, H. A., Orchard, J., Zelek, J., & Lakshminarayanan, V. (2017). Segmentation and feature extraction of retinal vascular morphology. In *Proceedings of SPIE Medical Imaging* (Vol. 10133).

Lin, G., Chen, M., Yeh, C., Lin, Y., Kuo, H., Lin, M., et al. (2018). Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy. *Journal of Ophthalmology*.

Liu, Z., Gao, A., Sheng, H., & Wang, X. (2025). Identification of diabetic retinopathy lesions in fundus images by integrating CNN and vision mamba models. *PLoS ONE, 20*(1), e0318264. https://doi.org/10.1371/journal.pone.0318264

Mandal, A. K. (2019). Study the behavior of Y, Cb and Cr of YCbCr on blurred image segmentation using local thresholding. *International Journal for Research in Applied Science and Engineering Technology*.

Mishra, A., Singh, L., Pandey, M., & Lakra, S. (2022). **RETRACTED:** Image-based early detection of diabetic retinopathy: A systematic review on artificial intelligence-based recent trends and approaches. *Journal of Intelligent & Fuzzy Systems, 43*(5), 6709–6741. https://doi.org/10.3233/JIFS-220772

Mishra, A., Singh, L., & Pandey, M. (2021). Short survey on machine learning techniques used for diabetic retinopathy detection. In *Proceedings of the International Conference on Computing, Communication, and Intelligent Systems* (pp. 601–606). IEEE. https://doi.org/10.1109/ICCCIS51004.2021.9397142

Mohanty, C., et al. (2023). Using deep learning architectures for detection and classification of diabetic retinopathy. *Sensors, 23*(12).

Nazir, A., Hussain, A., Singh, M., & Assad, A. (2024). Deep learning in medicine: Advancing healthcare with intelligent solutions and the future of holography imaging in early diagnosis. *Multimedia Tools and Applications*, 1–64.

Shyngyskhan, S., & Zhadra, Z. (2024). Recognizing different images in Python using TensorFlow and Keras. *Universum: Technical Sciences, 10*(11), 61–66.

Silva, P. S., Cavallerano, J. D., Sun, J. K., & Aiello, L. M. (2021). Effectiveness of artificial intelligence–based diabetic retinopathy screening in a primary care setting: A pilot study. *JAMA Ophthalmology, 139*(10), 1076–1082. https://doi.org/10.1001/jamaophthalmol.2021.2924

Swapna, T., Shivani, T., Srija, P., Akhila, D., & Srinidhi, P. (2022). Feature extraction to detect and classify diabetic retinopathy using fundus images. *International Research Journal of Engineering and Technology, 9*(11).

Tilahun, M., Gobena, T., Dereje, D., Welde, M., & Yideg, G. (2020). Prevalence of diabetic retinopathy and its associated factors among diabetic patients at Debre Markos Referral Hospital, Northwest Ethiopia. *Diabetes, Metabolic Syndrome and Obesity, 13*, 2179–2187. https://doi.org/10.2147/DMSO.S260694

Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Transactions on Image Processing, 2*(2), 176–201.

Yaqoob, M. K., Ali, S. F., Bilal, M., Hanif, M. S., & Al-Saggaf, U. M. (2021). ResNet-based deep features and random forest classifier for diabetic retinopathy detection. *Sensors, 21*(11).

Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. *IEEE Access, 12*, 41180–41218. https://doi.org/10.1109/ACCESS.2024.3376441

Zhang, G., Sun, B., Zhang, Z., Pan, J., Yang, W., & Liu, Y. (2022). Multi-model domain adaptation for diabetic retinopathy classification. *Frontiers in Physiology, 13*, 918929. https://doi.org/10.3389/fphys.2022.918929