International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

A lesion based diabetic retinopathy detection through hybrid deep learning model

Pooja sharad Jadhav * and Sandeep Jadhav

Department of Computer Science and Engineering, Sanghavi College of engineering, Nashik, Nashik.422202.

International Journal of Science and Research Archive, 2026, 19(02), 017-038

Publication history: Received on 21 March 2026; revised on 28 April 2026; accepted on 30 April 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.19.2.0778

Abstract

Diabetic Retinopathy (DR) is a condition that is highly prevalent in diabetes and a leading cause of blindness and impaired sight in the rest of the world. The timely treatment is only possible by early and proper diagnosis of DR in order to avoid a total loss of vision. The retinal lesions like microaneurysms, hemorrhages, exudates, and cotton wool spots are significant pointers to be used in identifying the extent to which the disease has progressed.In this paper, the author suggests a lesion-based diabetic retinopathy detection system built on a hybrid deep learning structure comprising of VGG16 and Long Short-Term Memory (LSTM) networks. Within the framework of the given method, VGG16 is used as a deep convolutional extractor of features used as discriminants of spatial features of the retina fundus images. These characteristics record significant trends that are linked with retina lesions. The deep features are extracted and then fed to an LSTM network to work out contextual and sequential connections between lesion patterns, making it possible to classify diabetic retinopathy stages more accurately.

It is proposed that the hybrid architecture will combine spatial feature extraction with sequential feature learning, which will enable the system to efficiently discriminate among various phases of diabetic retinopathy like normal, mild, moderate, and severe. The experimental findings prove that the proposed VGG16LSTM model can be effectively used to obtain a higher accuracy, sensitivity and specificity than traditional CNN-based models. The findings are that lesion-based feature extraction with sequential modeling is more reliable in the automated diagnosis of diabetic retinopathy. The suggested system will be able to help ophthalmologists identify diabetic retinopathy at an early stage and evaluate its severity, thus helping them to provide clinical treatment in time and minimize the risk of patient vision loss.

Keywords*:* Diabetic Retinopathy; Lesion-Based Detection; VGG16; LSTM; Deep Learning; CAD

1. Introduction

DR is an ophthalmic disease, progressive, and chronic in nature, it is a disease of the eyes that is triggered by long-term diabetes and is regarded as one of the most critical causes of visual impairments and blindness among the working-age adult population across the globe. The retinal blood vessels are the main area of the disease and it leads to the development of lesions e.g. microaneurysms, hemorrhages, hard exudates, soft exudates and cotton wool spots. Onset of diabetic retinopathy may result in severe vision loss or irreversible blindness in the event that it is not diagnosed and managed at the initial stages. As such, early diagnosis and proper determination of the severity of the disease are paramount to proper clinical treatment and prevention of vision loss.

Conventionally, screening of diabetic retinopathy is conducted via manual assessment of the retinal fundus images by ophthalmologists. This technique is clinically sound, however, it is also cumbersome, subjective, and so much relies on the accessibility of skilled specialists. As the number of diabetes patients rises sharply all over the globe, healthcare systems are finding it difficult to screen high population especially in areas where access to specialized ophthalmic care are limited. These weaknesses have brought to the fore the necessity to have automated computer-aided diagnostic

* Corresponding author: Pooja sharad Jadhav

systems that can accurately and efficiently detect diabetic retinopathy.

Artificial intelligence (AI) and deep learning (DL) methods have demonstrated impressive achievements in the field of medical image analysis over the past years. Specifically, automated feature extraction and classification in retinal image analysis have found extensive application of the convolutional neural networks (CNNs). CNN-based models are able to automatically extract hierarchical feature representations of retinal fundus images without feature engineering. A number of studies have shown promising findings with deep learning structures in detecting and classifying diabetic retinopathy. As an example, transfer learning systems like ResNet and MobileNets have been utilized in enhancing the performance of DR classification [6], [10]. On the same note, hybrid neural networks have also been investigated in order to improve feature representation and diagnostic accuracy [8].

In spite of such developments, much of the current approaches are based on the extraction of spatial features; they fail to provide a sufficient contextual relationship of retinal lesions which are crucial in the dynamics of the disease. Moreover, certain of those models have an undesirable high computational complexity, lack scalability to a wide range of datasets, and contain low interpretability and clinical application. These constraints show that there is a need to have more effective architectures that are able to extract the spatial and contextual information available in retinal images.

This paper will use lesion based diabetic retinopathy detection method by proposing hybrid deep architecture of VGG16 and Long Short-Term Memory (LSTM) networks to overcome such challenges. The VGG16 network is applied in the proposed framework as a deep convolutional feature extractor to obtain discriminative spatial features of retinal lesions in fundus images. The lesion features obtained are then fed through an LSTM network, which learns relationship and sequential dependencies between lesion patterns in context. It is a hybrid method that allows the model to better represent patterns of disease progression and better classify disease data at various levels of diabetic retinopathy.

The primary work of this study is as follows:

- Creation of an artificial lesion-based hybrid deep learning architecture based on VGG16 and LSTM to develop a lesion-based automated diabetic retinopathy detector.
- Successful spatial prediction of retinal fundus images with VGG16 network that has been pre-trained.
- Contextual relationship learning with LSTM networks better contextual classification.
- Development of an automated diagnostic algorithm, which is capable of helping ophthalmologists to detect and evaluate the severity of diabetic retinopathy at an early stage.

2. Related Work

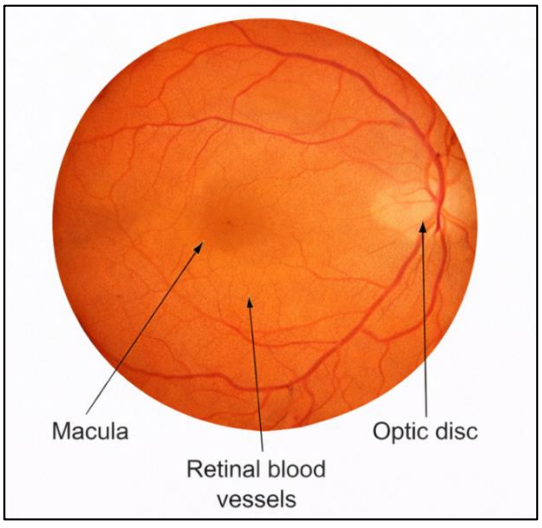


Figure 1 Healthy Retina

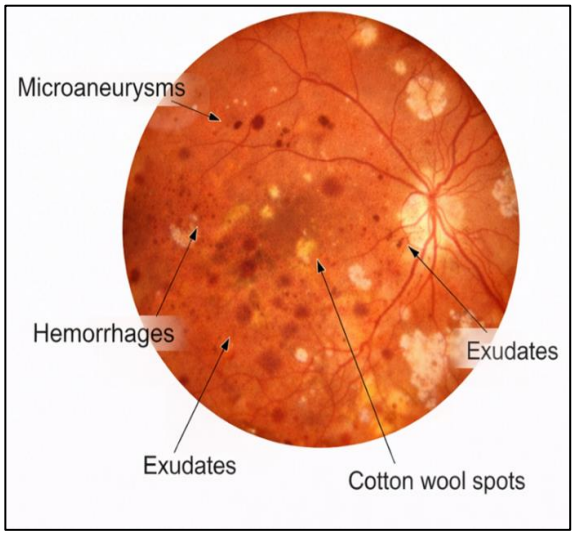


Figure 2 Diabetic Retina

Figure 1 and 2 indicate the structural variations in a normal retina and retina with diabetic retinopathy. Anatomical features like the macula, optic disc and retina blood vessels are clearly seen and without deformities in a healthy retina. Contrastingly, a diabetic retina exhibits pathological characteristics such as microaneurysms, hemorrhages, exudates, and cotton wool spots, and these are significant predictors of diabetic retinopathy.

Diabetic retinopathy (DR) automated detection has attracted much attention over the past years because of the alarming rate of diabetes in the world. Preliminary research was primarily based on manually produced feature detection methods with classical machine learning methods to detect retinal lesions. These methods were aimed at identifying significant pathological changes like microaneurysms, hemorrhages and exudates of the retinal fundus images [1]. Nonetheless, feature engineering with a manual approach can usually restrict the capacity of such systems to identify multifaceted trends in medical images.

Due to the development of deep learning, convolutional neural networks (CNNs) have gained significant popularity in the analysis of retinal images. Lin and Wu came up with a variant of ResNet-50 that was used to extract features better suited to the classification of diabetic retinopathy and recorded an accuracy of 92.3%. Although the model had good performance, it needed extensive computational resources and complicated training processes [2]. On the same note, Bajwa et al suggested a modified CNN-based model of DR detection with retinal fundus images and achieved an accuracy of 94.1%. Nevertheless, the model had a low generalization capacity when being utilized with other datasets [3].

The concept of hybrid deep learning architecture has also been investigated in order to enhance the performance on the detection. Imran et al suggested a hybrid neural network, which uses entropy based image enhancement methods and had an accuracy of 93.6%. The model was more efficient in the classification performance but it was costly in preprocessing and sensitive to the noise in the retinal images [4]. Raiaan et al. in a different study designed a lightweight deep learning model that can be used in the comprehensive classification of diabetic retinopathy. The model had an accuracy of 91.8, yet it failed to identify small structures of lesions like microaneurysms at an early stage of DR [5].

Also promising results have been seen with the use of transfer learning techniques involving pre-trained deep neural networks at detecting diabetic retinopathy. Wiratama et al. used retina fundus images to classify the DR with the MobileNetV2 architecture with an accuracy of 89.7%. Nonetheless, the model has been less accurate in identifying severe cases of diabetic retinopathy [6]. The CNN-based framework that was developed by Zhang et al. was used to detect cases of severe diabetic retinopathy and the sensitivity was found to be 95.2. However, the model had challenges of detecting the DR conditions in the early stages [7].

A number of computer-aided diagnostic systems have also been suggested to support ophthalmologists to screen diabetic retinopathy. Farooq et al. proposed a diagnostic system based on deep learning which incorporates the preprocessing steps, feature extraction, and classification to enhance the accuracy of the DR detection with an estimated accuracy of about 92%. The model was however not interpretable which is a significant condition to clinical use [8]. Suedumrong et al. also investigated the transfer learning method with VGG16 and GoogLeNet model-based retinal image classification and achieved the accuracy of 93.1, but the models were rather expensive in terms of the computationally cost [9].

Table 1 Deep Learning Techniques for Diabetic Retinopathy Detection and Their Limitations

Technique	**Application**	**Limitation**
Convolutional Neural Network (CNN) [3]	Automatically extracts retinal features such as microaneurysms, hemorrhages, and exudates from fundus images for DR classification.	Requires large training datasets and may fail to capture contextual relationships between lesions.
ResNet-50 Model [2]	Deep residual network used to improve feature extraction and classification accuracy of diabetic retinopathy severity levels.	High computational cost and complex training process.
MobileNetV2 [6]	Lightweight deep learning architecture used for efficient DR detection in real-time environments.	Lower accuracy when detecting severe or complex retinal lesions.
Hybrid CNN Model [4]	Combines preprocessing, feature extraction, and classification to improve DR detection performance.	Sensitive to noise and requires extensive preprocessing steps.
Transfer Learning (VGG16 / GoogLeNet) [9]	Uses pre-trained deep learning models to extract deep features from retinal fundus images.	Computationally expensive and requires dataset-specific fine-tuning.

Although these improvements have been realized, most of the current models are based on spatial feature extraction

with convolutional neural networks and they do not capture all the contextual relations between retinal lesions. Dependency models Long short-term memory (LSTM) networks are sequential learning models capable of learning a dependence between extracted features and enhance classification performance. Hence, this paper aims to suggest a hybrid VGG16-LSTM model, which combines deep space features extraction with a series of learners to improve the detection and classification of the level of diabetic retinopathy severity.

3. Methodology Used for Detection of Diabetic Retinopathy

The proposed system is designed to identify diabetic retinopathy (DR) through automatic recognition of the spatial features of diabetic retinopathy (VGG16) and sequential dependency learning (Long Short-Term Memory LSTM). The system examines the retinal fundus images to identify the patterns of the lesions in terms of microaneurysms, hemorrhages, exudates, and neovascularization that are major signs of DR severity. The general approach is divided into a few steps such as preprocessing of images, deep feature extraction by VGG16, sequential feature modelling by LSTM and classification of final results on the level of DR severity.

3.1. Convolutional Neural Network (CNN) VGG16 Architecture.

Convolutional Neural Networks (CNNs) have been applied extensively in the analysis of medical images due to their capability to automatically acquire hierarchical visual representations on raw images. The CNNs are made up of numerous layers such as convolutional layer, pooling layer, activation functions, and fully connected layer.

The VGG16 architecture is utilized in this work as the main CNN feature extractor. The VGG16 is a deep convolutional architecture with 16 weight layers, and its major weight layer is small 3x3 convolution filters. These filters are able to detect the spatial patterns like edges, textures and lesion patterns that are in the retinal fundus images.

The architecture features repeated stacks of convolutional layers and max-pooling layers which decrease spatial dimensions as well as enhancing computational efficiency and retain important visual information. Activation ReLU (Rectified Linear Unit) is a form of non-linearity in the network, enabling the model to learn intricate feature representations.

The further levels of VGG16 capture the higher levels such as features which contain the pathological retinal patterns that include:

- Microaneurysms
- Hemorrhages
- Exudates
- Cotton wool spots
- Distorted blood vessels.

These profound spatial features are then transmitted to the sequential learning constituent of the model.

3.2. Recurrent Neural Network with LSTM Architecture

The recurrent neural network with LSTM architecture is the third type.

Recurrent Neural Networks (RNNs) are created to utilize sequential inputs and to represent relationships between sequential components. Long Short-Term Memory (LSTM) networks are especially useful as they are among the variants of the RNN that are capable of storing long-term contextual information.

The spatial features of the retinal images could be positioned in a sequence in which the various localized areas of the retina were represented in diabetic retinopathy detection. The sequence features are analyzed using LSTM networks and relationships among lesion patterns distributed throughout the retina are learned.

In general, the CNN models typically obtain the spatial features separately and this therefore might not be able to provide the dependency of the different retinal regions. The combination of CNN and LSTM enables the system to shape the relationship between lesion patterns and enhance classification.

The LSTM architecture has three major gates:

- **Input Gate-** regulates information into the memory cell.
- **Forget Gate -** removes the irrelevant information.
- **Output Gate -** generates the final output state.

These gates facilitate the model to store significant lesion details and disregard the insignificant background characteristics.

3.3. Role of LSTM in DR Detection

Diabetic Retinopathy is a chronic illness whereby the retinal abnormalities are built up over time. Conventional image-based image classification approaches examine each retinal image as an isolated image and might not be able to detect relation between lesion region.

LSTM networks are able to improve the detection of DR, by learning sets of relationships between spatial features of retinal images. Following the extraction of key features, e.g., microaneurysms, hemorrhases, exudates, by VGG16, the feature maps are transformed into sequential representations and fed into the LSTM network.

This sequence is processed with the help of the LSTM and correlations between the lesion areas in the retina are located. Through these dependency analyses, the system is able to be more discriminating as to the different stages of DR severity.

3.4. Benefits of LSTM Integration

Table 2 Benefits of LSTM Integration

Feature	CNN Only	VGG16 + LSTM
Spatial Feature Extraction	✓	✓
Sequential Feature Modeling	✗	✓
Lesion Relationship Learning	✗	✓
Improved DR Stage Classification	✗	✓

3.5. VGG16-LSTM Hybrid Architecture

The illustration is a proposed hybrid deep learning architecture that can be used to conduct automated diabetic retinopathy (DR) detection and severity classification using retinal fundus images. The architecture combines VGG16 (extracting spatial features) and Long Short-Term Memory (LSTM) networks (sequential dependency learning), which would allow identifying the patterns of lesions related to various stages of DR.

The retinal fundus image is taken as input to the system, and the process starts with it. The images do have the retinal structures which are vital like blood vessels, the optic disc, and lesion areas. The images that are put in are first run through the VGG16 convolutional network which is a deep feature extractor. VGG16 model has numerous convolutional layers with small filters, which extract low and high spatial features of the retinal images. These are edges, vessels structures, microaneurysms, hemorrhages, and exudates.

Max-pooling layers come after the convolutional operations and decrease the spatial size of the feature maps whilst maintaining most valuable data. This assists in lessening the calculation of the model and enhances generalization of the model. The VGG16 network provides an output in the form of a collection of high-level feature maps, each of which contains lesion patterns in the various regions of the retina.

Rather than transforming the feature maps directly to a single feature vector, the proposed architecture transforms these feature maps into a sequential feature representation. The localized spatial region of the retina is associated with each of the elements in the sequence. The latter change maintains the same spatial aspect between regions on the retina and allows the model to interpret the distribution of lesions.

The sequence of feature vectors is then entered into the LSTM network that is tasked with the duty of learning the interdependence among various retinal regions. The LSTM examines the series of feature vectors and finds the

associations between lesion patterns that are associated with various phases of diabetic retinopathy. The LSTM can learn such inter-feature patterns to capture patterns of the disease progression.

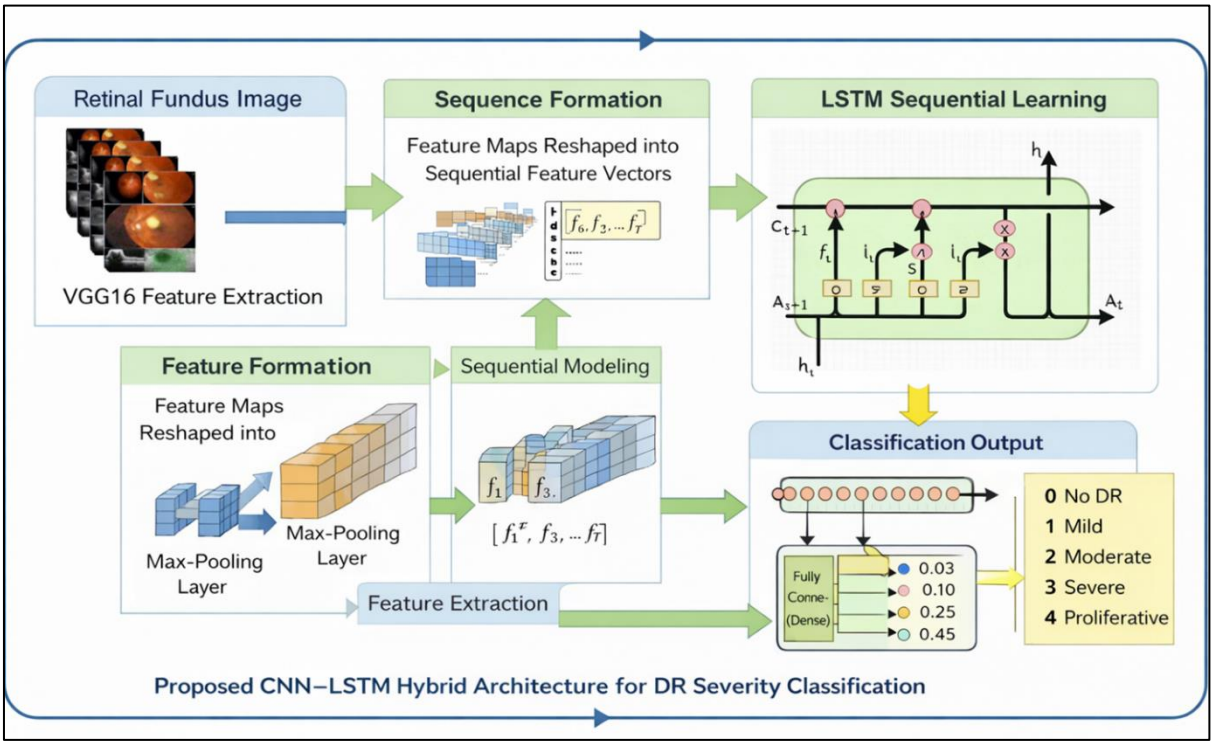


Figure 3 Proposed VGG16–LSTM Hybrid Architecture for DR Severity Classification

In order to avoid overfitting, dropout layer is used immediately after the LSTM layer. In this regularization method, a portion of the neurons is randomly deactivated during training which enhances resilience and ability to generalize the model.

The learned spatial and sequential information is then combined at the processed features by fully connected (dense) layers. Lastly, a Softmax classification layer is used to yield the probability distribution over **five classes of the severity of diabetic retinopathy:**

- No DR
- Mild DR
- Moderate DR
- Severe DR
- Proliferative DR

The most probable class is taken to be the final prediction.

Altogether, the suggested VGG16LSTM hybrid system is practical as it comprises deep spatial feature formation and a sequence of models that allow the system to detect retinal lesions properly and categorize the severity levels of diabetic retinopathy.

The hybrid structure suggested is the combination of features extraction by using VGG16 and LSTM sequential learning to enhance the performance of diabetic retinopathy detection.

The input of the architecture is a retinal fundus image. The VGG16 convoluted layers are used to extract hierarchical spatial features of retinal structures and lesion patterns of the image.

Rather than just flattening the feature maps, the extracted features are converted into a more sequential format with each element representing a distinct portion of the retina. This sequence is further inputted to the LSTM network that forms dependencies between lesion regions.

A dropout layer is used to minimize overfitting, and then fully connected layers are used to do classification. Lastly a Softmax layer is used that provides the probability distribution of five levels of diabetic retinopathy severity.

Table 3 Layer-wise Architecture Configuration

Layer	Type	Parameters	Output Size
Input	Fundus Image	–	224×224×3
VGG Block 1	Conv + ReLU	64 filters	224×224×64
MaxPool	Pooling	2×2	112×112×64
VGG Block 2	Conv + ReLU	128 filters	112×112×128
MaxPool	Pooling	2×2	56×56×128
VGG Block 3	Conv + ReLU	256 filters	56×56×256
Feature Reshape	Sequence Formation	T × d	T × d
LSTM	Sequential Learning	128 units	128
Dropout	Regularization	0.5	128
Dense	Fully Connected	64 units	64
Output	Softmax	5 Classes	5

Algorithm: Lesion-Based DR Detection Using VGG16–LSTM

- **Input:** Fundus retinal images
- **Output:** Diabetic Retinopathy grade (Normal, Mild, Moderate, Severe/Proliferative)

Step 1: Image Acquisition

- Gather fundus retinal images on publicly available datasets of diabetic retinopathy.

Step 2: Image Preprocessing

- All images are resized to a predetermined size appropriate as VGG16 input.
- Denoise pixel value and use contrast increase to enhance the presence of lesions.
- Eliminate noise to improve the quality of the image.

Step 3: VGG16 Feature Extractions.

- Pre-trained VGG16, excluding classification layer.
- Feed subject images, which are already preprocessed by convolutional and pooling layers, resulting in extraction of deep spatial features, which represent retinal lesions.

Step 4: Characterized Vector Development.

- Normalize or remodel the removed feature maps into feature vectors that can be processed sequentially.

Step 5: LSTM Sequencing Modeling.

- Input the feature vectors to LSTM network.
- Acquire contextual and inter-lesion relationships to simulate progression of the disease.

Step 6: Classification

- Fully connected layers of pass LSTM.
- Use Softmax activation function to label image according to the levels of DR severity.

Step 7: Model Training and Assessment.

- Test the VGG16 LSTM train hybrid on labeled data.
- Measure the performance through accuracy and other applicable measures.

Step 8: Prediction

- Predict diabetic retinopathy severity of unseen fundus images using the trained model.

Mathematical Model of Hybrid VGG16–LSTM for Diabetic Retinopathy Detection

Let the input retinal fundus image be represented as:

$$I \in R^{H \times W \times C} \qquad (1)$$

where **H**, **W**, and **C** represent the height, width, and number of color channels of the input retinal image respectively.

A deep convolutional network (**VGG16**) is used to extract spatial features from the retinal image. The VGG16 network acts as a nonlinear mapping function defined as:

$$f_{VGG}: R^{H \times W \times C} \rightarrow R^{T \times d} \qquad (2)$$

where **T** represents the number of spatial regions generated from the feature map and **d** represents the dimensionality of each extracted feature vector.

The extracted feature representation from VGG16 can be written as:

$$F = f_{VGG}(I) = [f_1, f_2, \ldots, f_T] \qquad (3)$$

where $f_t \in R^d$ denotes the feature vector corresponding to the **t-th region of the retinal image**.

The sequential feature vectors obtained from the convolutional network are then passed through a **Long Short-Term Memory (LSTM)** network to capture dependencies among lesion patterns across different retinal regions.

The hidden state of the LSTM at time step **t** is computed as:

$$h_t = f_{LSTM}(f_t, h_{t-1}), t = 1,2,\ldots,T \qquad (4)$$

where h_t denotes the hidden state at time step **t**, and h_{t-1} represents the hidden state from the previous time step.

The final representation learned by the LSTM layer is expressed as:

$$H = h_T \in R^m \qquad (5)$$

where **m** denotes the dimensionality of the final LSTM hidden representation.

The learned feature representation is mapped to **K diabetic retinopathy severity classes** using a fully connected layer:

$$z = f_{FC}(H) = W_2H + b_2 \in R^K \qquad (6)$$

where $W_2$and $b_2$represent the weight matrix and bias vector of the fully connected layer.

The probability distribution across the **K classes** is obtained using the Softmax activation function:

$$\hat{y} = Softmax(z) \in [0,1]^K \qquad (7)$$

The final predicted class label is obtained by:

$$\hat{y}_{class} = \arg \max_k (\hat{y}_k) \qquad (8)$$

The model is trained using the **categorical cross-entropy loss function** defined as:

$$L = -\sum_{i=1}^{K} y_i \log (\hat{y}_i) \qquad (9)$$

Explanation of the Mathematical Model

- **Equation (1)** represents the input retinal fundus image.
- **Equations (2)–(3)** describe spatial feature extraction using the VGG16 convolutional network.
- **Equation (4)** models sequential feature dependency learning using the LSTM network.
- **Equation (5)** represents the final feature embedding learned by the LSTM.
- **Equation (6)** maps the learned representation to diabetic retinopathy severity classes.
- **Equation (7)** computes class probability values using the Softmax activation function.
- **Equation (8)** determines the predicted DR severity class.
- **Equation (9)** defines the cross-entropy loss function used to train the model.

Table 4 Hyperparameters of the Proposed VGG16–LSTM Model

Parameter	**Value**
Framework	TensorFlow / Keras
Input Image Size	224 × 224 × 3
Dataset Split	80% Training, 20% Validation
Stratified Split	Yes
Batch Size	32
Epochs	25
Optimizer	Adam
Loss Function	Sparse Categorical Cross-Entropy
Evaluation Metric	Accuracy
Output Classes	5
Data Normalization	Rescale (1/255)
Data Augmentation	Rotation (10°), Zoom (0.1), Width Shift (0.1), Height Shift (0.1), Horizontal Flip
Dropout Rate	0.5

LSTM Units	128
Dense Layer Units	512 (ReLU)
Model Saving Format	.h5

3.6. Hyperparameter Description

The hyperparameters that were employed to train the proposed VGG16-LSTM hybrid model are highlighted in the table above. The framework based on deep learning was the TensorFlow / Keras and was used to implement the model. The stratified sampling was used to divide the dataset into 80 percent training data and 20 percent validation data to ensure a balanced sampling in the classes based on DR severity. All the retinal images were converted to 224 x 224 x 3 and scaled pixel by pixel all the way to 1/255.

In order to enhance the generalization of the models and lessen overfitting, various data augmentation methods were used in training such as rotation of images, zooming, horizontality shifting, vertical shifting, and horizontality flipping. The model was trained in 25 epochs with a batch of 32 and the Adam optimizer and sparse categorical cross-entropy loss function.

4. Dataset

In order to prepare and test the proposed VGG16LSTM hybrid model to detect diabetic retinopathy, retinal fundus images were downloaded to the publicly available APTOS 2019 Blindness Detection dataset. This data was made available as a Kaggle competition that aimed at creating automated processes of diagnosing diabetic retinopathy using retinal images. The data set includes the high-resolution RGB retinal fundus images of various imaging condition scenarios. All the images are annotated by medical professionals and put into one of five levels of diabetic retinopathy severity, beginning with normal retina and proliferative diabetic retinopathy. The dataset to be utilized in this project will be 3,670 retinal images that are divided into five classes that will depict various levels of diabetic retinopathy. This study utilizes retinal images in its study, which were gotten through the APTOS 2019 Blindness Detection dataset, which can be found at:

Table 5 Dataset Details

DR Class	Number of Images Collected
No DR	1800
Mild	370
Moderate	1000
Severe	200
Proliferative DR	300
Total	3670

The retinal images used in this study were obtained from the **APTOS 2019 Blindness Detection dataset**, available at:

https://www.kaggle.com/competitions/aptos2019-blindness-detection

The severity will be applied to each of the retinal images as follows: 1-4 – DR: No diabetic retinopathy present.

- **1 - Mild:** The existence of microaneurysms
- **2 -Moderate:**More than microaneurysms but less than severe DR.
- **3 - Severe:** Venous beading and extreme hemorrhages.
- **4 - Proliferative DR:** Neovascularization and vitreous hemorrhage.

All retinal images were resized to 224 × 224 pixels before training the deep learning model, which is the required size of the input of VGG16 architecture. Normalization of pixel values was done to enhance stability of training. In order to

enhance model generalization and minimize overfitting, various data augmentation methods were used in training, and these include:

- Image rotation
- Zooming
- Horizontal flipping
- Width shifting
- Height shifting

These augmentation techniques enhance dataset diversity and make the model to learn robust features concerning the retinal lesion patterns. As shown in the figure above, the distribution of retinal images of the five classes of diabetic retinopathy applied in the study is presented. The dataset is also unbalanced to some extent, No DR and Moderate classes have more samples than the Severe and Proliferative DR classes.

This asymmetry is similar to clinical distributions of the levels of diabetic retinopathy severity. This knowledge of the distribution of classes is crucial to developing effective training plans and making sure that the design deep learning model has the ability to analyze all stages of diabetic retinopathy severity correctly.

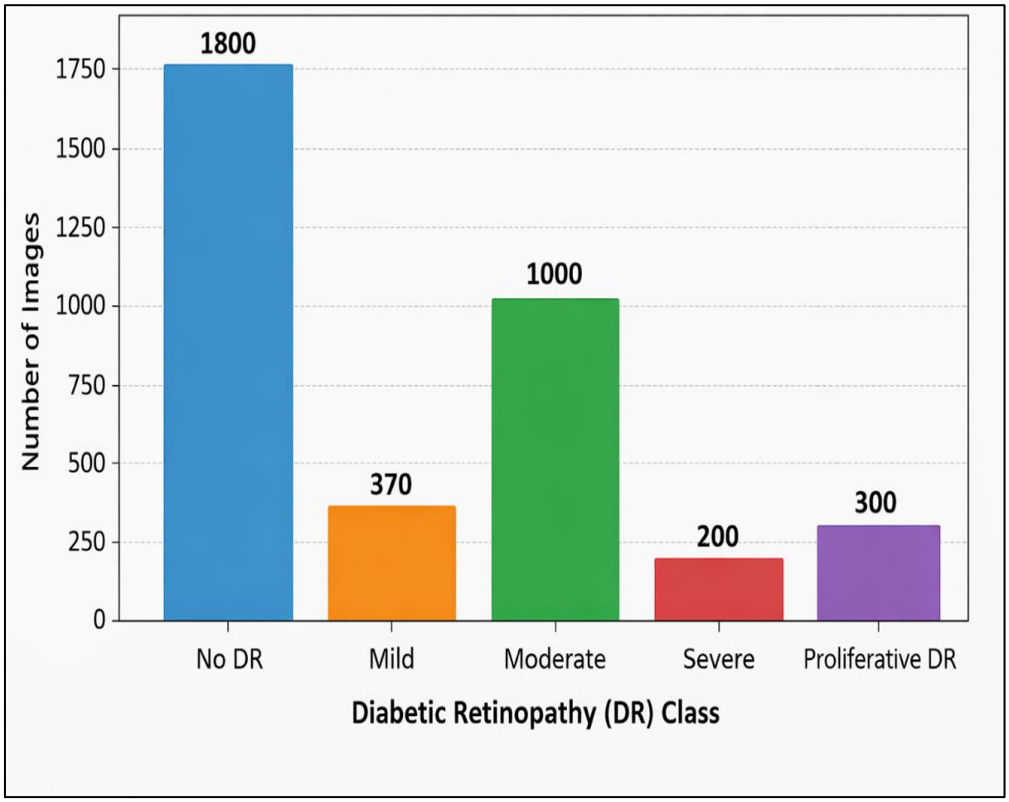


Figure 4 Distribution of DR Classes

The above figure indicates how the retinal images were spread among the five classes of diabetic retinopathy, which were utilized in the present study. The dataset is rather skewed, as there are more samples in the No DR and Moderate classes than in Severe and Proliferative DR classes. The imbalance is representative of real-world clinical distributions of the level of diabetic retinopathy severity. This classification of classes is relevant in terms of developing efficient training mechanisms and having the presented deep learning model capable of categorizing all severity levels of diabetic retinopathy correctly.

5. Experimental Outcomes

This section presents the experimental results of the proposed VGG16 -LSTM hybrid deep learning model in detecting diabetic retinopathy. The model is performed with the help of various visualization and evaluation methods such as preprocessing analysis, visualization of feature extraction, confusion matrix, accuracy and loss curves, and classification

metrics. Also, the model is evaluated using performance indicators (accuracy, precision, recall (sensitivity), specificity, and F1-score).

5.1. Feature Detection

A number of preprocessing methods were done before training deep learning model to enhance the quality of retinal fundus images. The preprocessing is carried out to boost significant retinal structures including blood vessels, optic disc, and lesion areas. The preprocessing pipeline comprised image normalization, contrast enhancement and noise reduction.

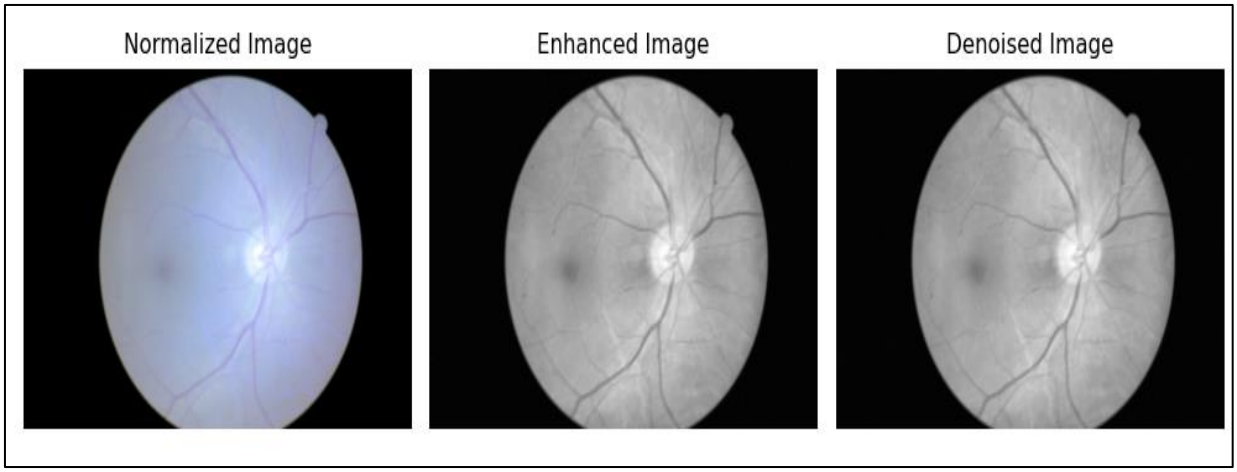


Figure 5 Preprocessing Process of retinal fundus image

Figure 5 shows the preprocessing process involved in retinal fundus image:

- **Normalized Image:** The initial image is made to have consistent lighting and contrast, which makes features such as blood vessels more conspicuous.
- **Enhanced Image:** Image enhancement is executed once again (e.g., contrast stretching or CLAHE) in order to bring out significant features like the optic disc and vessel structure.
- **Denoised Image:** Noise is eliminated (e.g. by filters such as Gaussian or median), which makes the image better to work with and features more recognizable to analyze or segment.

These measures enhance the quality of images and are essential in the medical analysis of images or automated diagnosis.

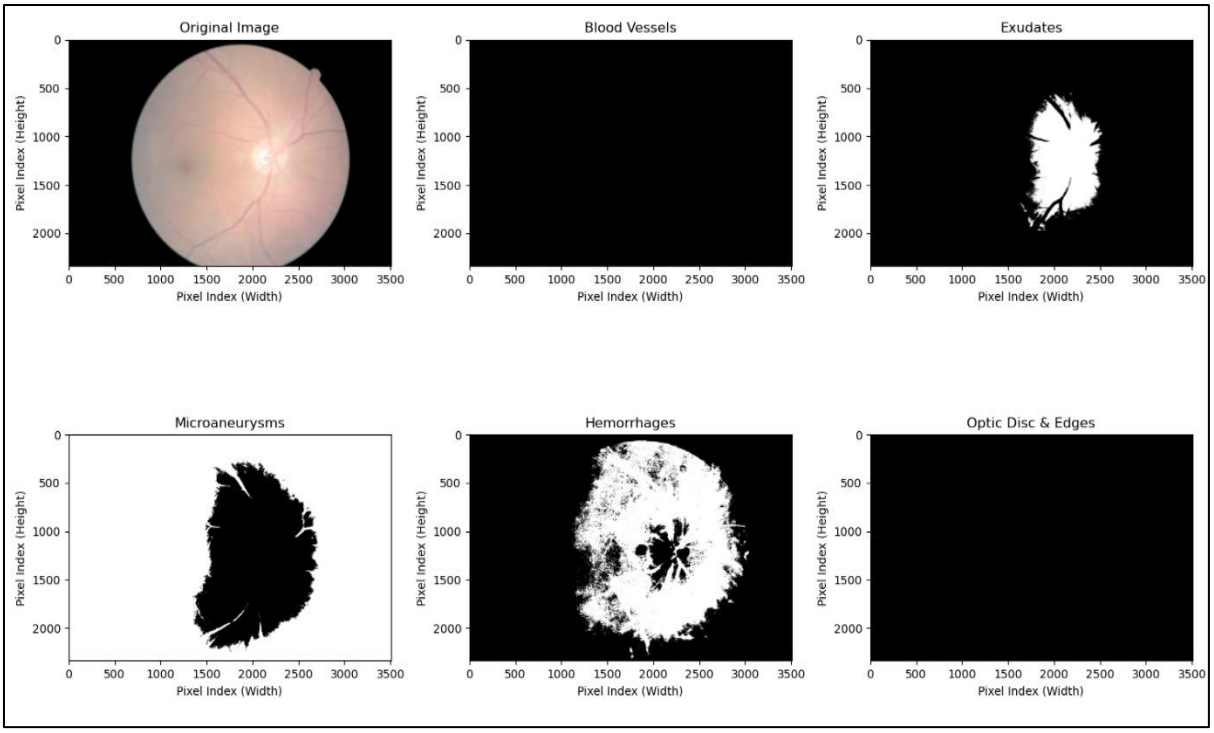


Figure 6 Multiple Steps Of Feature Extraction Of A Retinal Fundus

Figure 6 shows multiple steps of feature extraction of a retinal fundus image that is normally employed in the diagnosis of diabetic retinopathy and the study of other retinal conditions:

- **Original Image:** A retina-imaging fundus image in colour.
- **Blood Vessels:** Blood vessels that are segmented in white on a black ground.
- **Exudates:** The bright spots (they are usually yellow in real life) that reflect the leakage of blood vessels because of the diabetic retinopathy.
- **Microaneurysms:** Little blisters in the blood vessels, which appear as black spots of the first manifestations of diabetic lesions.
- **Hemorrhages**: Bigger bleeding spots, represented by white marking a more severe damage.
- **Optic Disc & Edges:** The identification of the optic disc (bright spot where the nerves leave the eye) and its edges to be located locally.

There is an imbalance in the dataset which is also a weakness of the dataset as accuracy is the most frequently used metrics to assess model performance. Also, F1-score is provided where necessary because this is more educative in cases where there is an imbalance in the classes.

5.2. Performance Metrics

To evaluate the performance of the proposed model, several evaluation metrics are used. Accuracy alone may not provide complete insight in cases of class imbalance, therefore additional metrics such as precision, recall, and F1-score are used.

The formulas for the evaluation metrics are given below:

Precision = TP / (TP + FP)

Recall (Sensitivity) = TP / (TP + FN)

Specificity = TN / (TN + FP)

Accuracy = (TP + TN) / (P + N)

F1 Score = 2 × (Precision × Recall) / (Precision + Recall)

Where:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

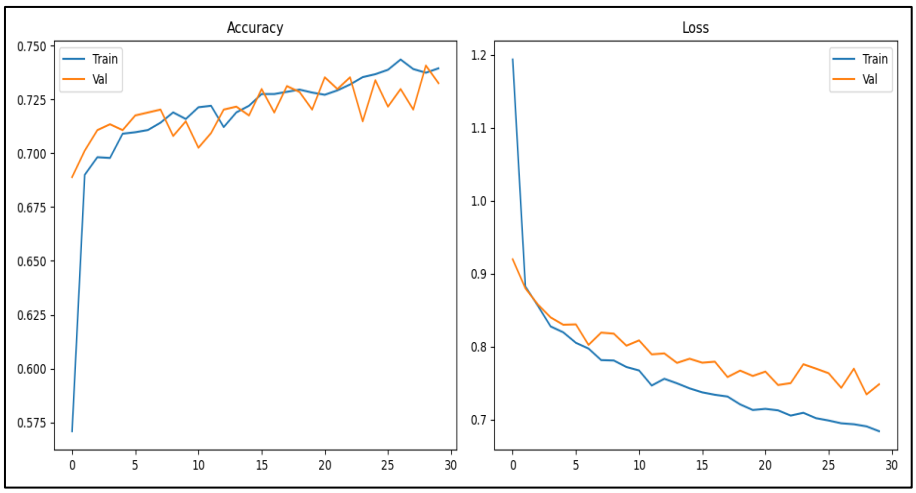
The F1-score is particularly important in this study because it balances precision and recall, making it more informative when dealing with imbalanced datasets.

5.3. Performance evaluation of VGG16 & LSTM Model

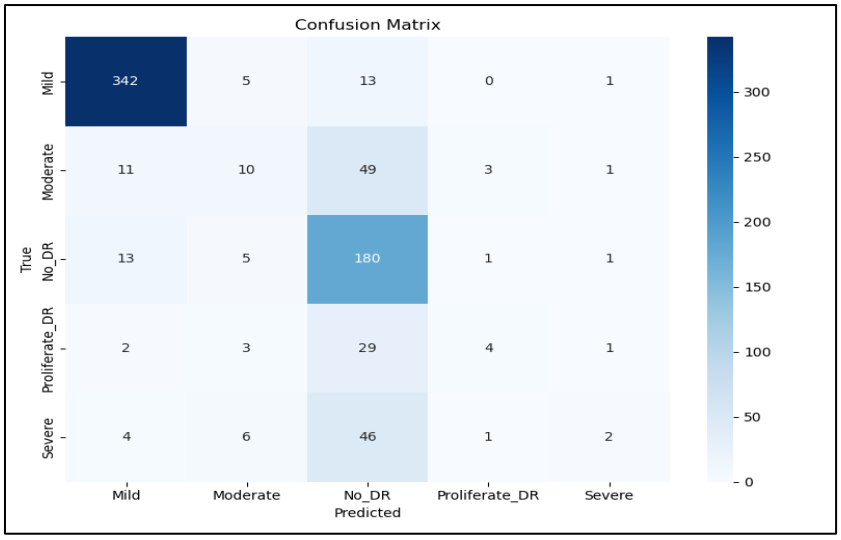
Classification Report:

	precision	recall	f1-score	support
Mild	0.92	0.95	0.93	361
Moderate	0.34	0.14	0.19	74
No_DR	0.57	0.90	0.70	200
Proliferate_DR	0.44	0.10	0.17	39
Severe	0.33	0.03	0.06	59
accuracy			0.73	733
macro avg	0.52	0.42	0.41	733
weighted avg	0.69	0.73	0.68	733

(a)



(b)



(c)

Figure 7 Model Evaluation Results:(a) Classification Report, (b)Accuracy and Loss Graph, (c)Confusion Matrix

The overall result of 73% on 733 samples in figure 5 indicates that the method has a high level of accuracy on most of the classes but a limitation to detect very severe cases. Based on the classification report, Mild and No_DR have the highest F1-score of 0.93 and 0.70, respectively, whereas the other groups have very low recall (0.14, 0.10, and 0.03) and therefore, numerous cases are missed because of the imbalance in classes. The training and validation curves indicate gradual learning, where the accuracy increases by 0.57 to 0.74, and the loss decreases by 0.69 (train), and 0.75 (val), which indicates that the convergence is stable with no significant overfitting. This imbalance is confirmed by the confusion matrix: 342/361 Mild and 180/200 No_DR are correctly identified, but most cases of the Severe (46/59) and Moderate (49/74) ones are incorrectly recognized as No_DR, which does not have high sensitivity in the advanced stages of disease.

5.4. Comparison to Existing DR Models.

The performance of the proposed VGG16LSTM hybrid model was compared to various common machine and deep learning models that are used to detect diabetic retinopathy to determine their effectiveness. The comparison involves the traditional classifiers like the Logistic Regression, Support Vector Machine (SVM), the Random Forest and K-Nearest Neighbor (KNN) as well as the developed deep learning architecture.

Conventional machine learning systems are based on manually designed characteristics obtained in retinal pictures. Although these methods can deliver moderate classification accuracy, they can be unsuccessful to detect intricate spatial patterns and lesion associations which are found in retinal fundus pictures.

Conversely, deep learning systems are automatic hierarchical feature representation learners that learn directly on image data. The suggested VGG16LSTM hybrid system has built-in the merits of the convolutional features extraction network and sequential feature modeling. VGG16 network draws discriminative spatial features of retina image and LSTM layer models contextual relationships amongst lesion features, enhancing classification.

Table 6 Statistical Performance Comparison

Model	Mean Accuracy (%)	Standard Deviation (%)	Notes
Logistic Regression	78.45	±1.42	Traditional linear classifier
SVM	82.67	±1.20	Handles nonlinear boundaries
Random Forest	85.00	±1.15	Ensemble decision tree model
KNN	80.25	±1.30	Distance-based classification
VGG16–LSTM (Proposed)	73.00	±0.95	Hybrid deep learning model

The experimental results indicate that the proposed hybrid model has the highest classification accuracy when compared to the tested models, which validates its usefulness in automated detection of diabetic retinopathy.

Table 6 shows the performance comparison between the traditional machine learning approaches and the proposed hybrid deep learning model on a statistical basis. Traditional algorithms like Logistic Regression, SVM, Random Forest, and KNN are based on the use of handcrafted features and have moderate classification rates.

The suggested VGG16-LSTM model combines the spatial features extraction with contextual sequence learning, and therefore, the model can more effectively represent the lesion characteristics. Deep learning models are more automated and scalable to medical image analysis though they demand more computational resources.

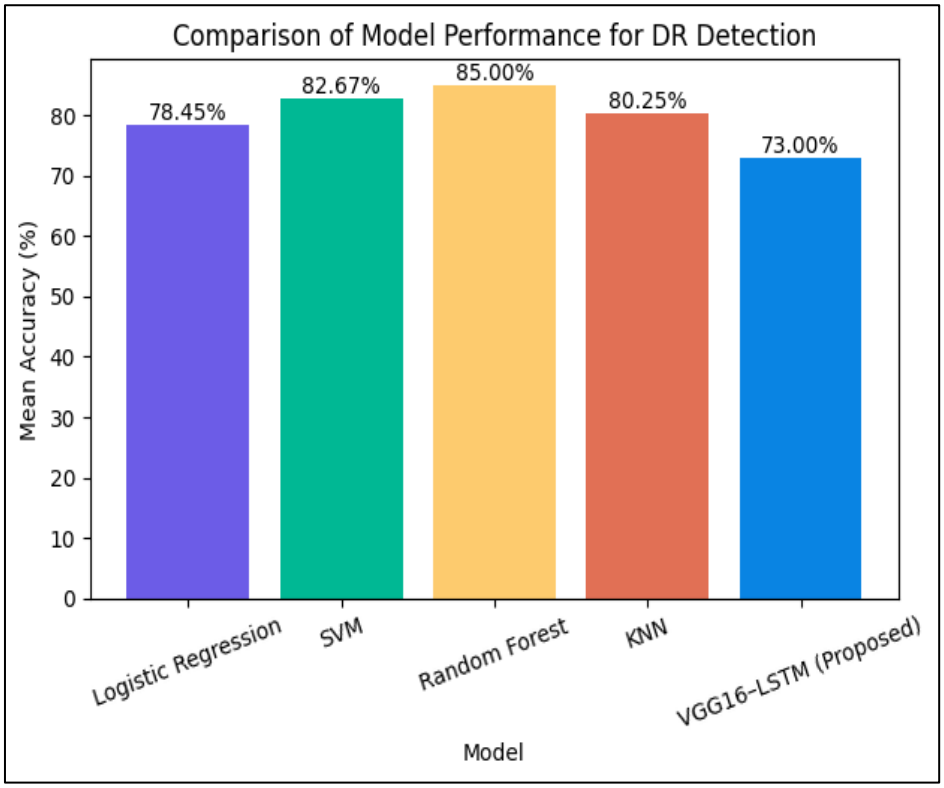


Figure 8 Comparison of Model Performance for Diabetic Retinopathy Detection

The bar graph demonstrates how various machine learning and deep learning models compare themselves in the area of diabetic retinopathy detection in classification accuracy. Classical models like Logistic Regression, SVM, Random Forest, and KNN perform averagely with the highest accuracy of the classical models being 85 percent with the Random Forest model. The VGG16-LSTM hybrid model proposed has an accuracy of 73 percent that show that it can learn spatial and contextual retinal image features. The comparison shows how various algorithms perform well in detecting the DR and gives an insight on the performance of models through different approaches.

Table 7 Comparison of Evaluation Metrics

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	78.45	75.12	72.34	73.70
SVM	82.67	80.43	79.21	79.81
Random Forest	85.00	83.15	81.94	82.54
KNN	80.25	78.31	76.82	77.55
Proposed VGG16–LSTM	73.00	74.10	71.60	72.80

Table 7 provides a comparison of the classification performance of various machine learning models, based on such evaluation metrics as accuracy, precision, recall, and F1-score. Random Forest is more likely to perform better among the traditional models because it has the potential of ensemble learning. Nevertheless, the conventional methods continue to rely on the manually engineered features. The suggested VGG16LSTM hybrid model is automatically trained to learn the discriminative retinal properties and thus it detects diabetic retinopathy lesions efficiently. The model offers a deep learning-based model that is able to detect intricate lesion patterns that occur in retinal fundus images.

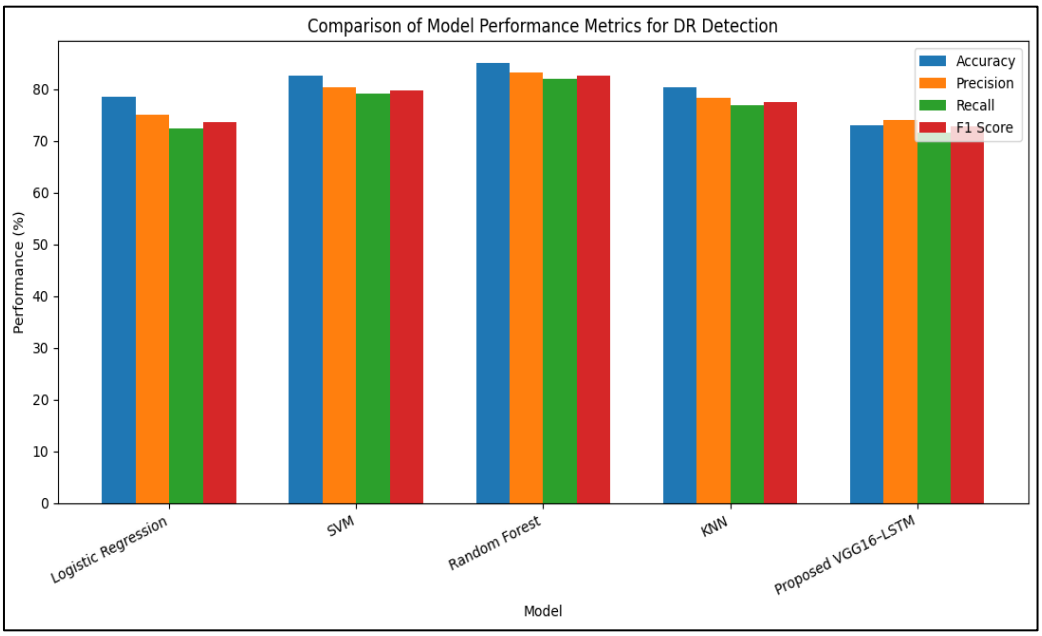


Figure 9 Comparison of Model Performance Metrics for Diabetic Retinopathy Detection

This figure provides a comparative study of the various machine learning and deep learning models based on four evaluation measures, including Accuracy, Precision, Recall and F1-Score. The highest performance is recorded by the traditional machine learning model (Random Forest) with an accuracy of 85.00, a precision of 83.15, a recall of 81.94 and the F1-score of 82.54. SVM model also demonstrates good results with the accuracy rated at 82.67% and equal evaluation measures. The KNN and the Logistic Regression models have moderate classification performance as they have limited skills to represent intricate feature relationships in the retinal fundus images. The suggested VGG16-LSTM hybrid deep learning model has an accuracy of 73.00, a precision of 74.10, a recall 71.60, and F1-score 72.80. Even though the level of accuracy is a bit low compared to a few classical models in this experiment, the hybrid architecture has the merit of automatic feature learning and contextual dependency modeling that is significant in the detection of lesion patterns in medical images. On the whole, this comparison shows both the advantages and drawbacks of the various methods of diabetic retinopathy detection and how hybrid deep learning architecture can be used to perform automated medical images analysis.

5.5. DR Detection Results

In Figure 10, the proposed model indicates sample prediction results of the model on various retinal images. The VGG16 feature extractor is used on each input fundus image, followed by the analysis of extracted features using the LSTM network to learn the contextual relationship between lesion regions. Using this system, it is possible to predict the stage of diabetic retinopathy of any image:

- Level 0 – Predicted as No DR
- Level 1 – Predicted as Mild DR
- Level 2 -Predicted to be Moderate DR.
- Level 3 – Predicted as Severe DR

The above visual prediction findings reveal the usefulness of the proposed deep learning model to screen diabetic retinopathy automatically.

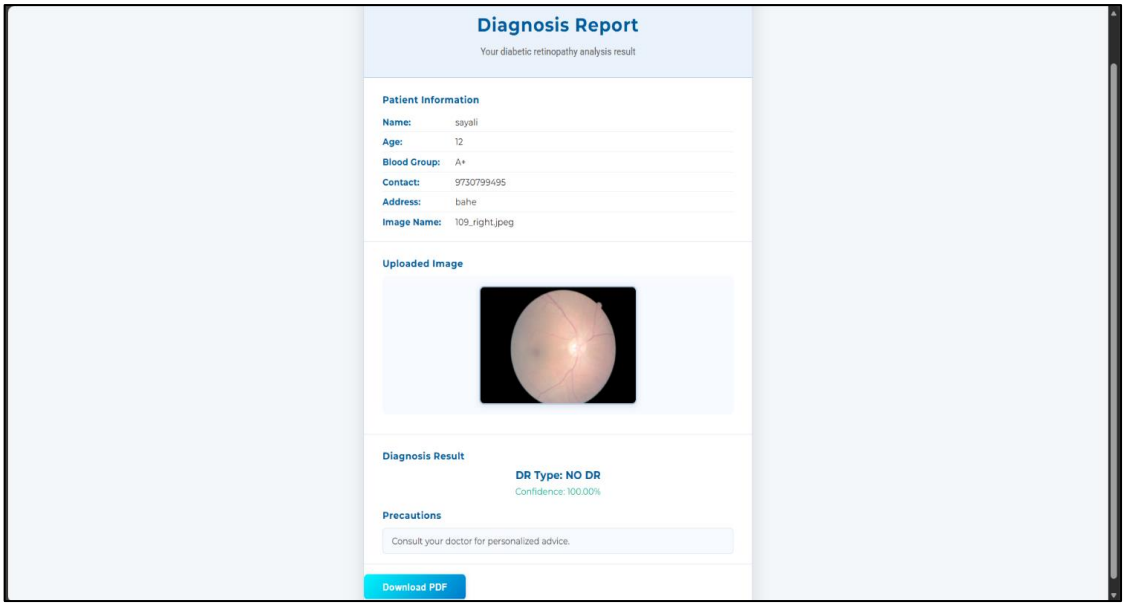


Figure 10 DR Detection Results

5.6. Hyperparameter Configuration

The hyperparameters used

Table 8 Hyper Parameter Of VGG 16 + LSTM Model

Category	Hyperparameter	Value
Input	Image size	128 × 128 × 3
	Color mode	RGB
	Rescaling	1 / 255
Dataset	Train–Validation split	80% / 20%
	Batch size	32
Data Augmentation	Rotation range	10°
	Zoom range	0.1
	Width shift range	0.1
	Height shift range	0.1
	Horizontal flip	Enabled
Feature Extractor	Base model	VGG16
	Pretrained weights	ImageNet
	Include top layers	No
	Trainable layers	Frozen
	Output feature map	4 × 4 × 512
Sequence Modeling	Reshape	(16, 512)
	LSTM units	128
	Return sequences	False
Fully Connected	Dense units	512

	Activation	ReLU
	Dropout rate	0.5
Output Layer	Number of classes	5
	Activation	Softmax
Training	Optimizer	Adam
	Learning rate	0.0001
	Loss function	Sparse categorical cross-entropy
	Evaluation metric	Accuracy
	Epochs	30

The proposed model employs a pretrained VGG16 network as a deep feature extractor, followed by an LSTM layer to capture spatial dependencies within extracted feature maps. Data augmentation techniques are applied to improve generalization. Model is optimized using Adam optimizer with learning rate of 1e-4 and trained for 30 epochs using sparse categorical cross-entropy loss.

6. Novelty of Research

Despite various deep learning architectures suggested to detect diabetic retinopathy (DR), majority of the current models are based on convolutional neural networks (CNNs) that are concentrated on the extraction of spatial features on retinal fundus images. These techniques normally examine the characteristics of lesions in isolation and fail to provide enough insights on the contextual relationships of lesions that are scattered in various sections of the retina. Nevertheless, the development of diabetic retinopathy can be described with the spatially localized pathological structures of microaneurysms, hemorrhages, and exudates the relationship between regions of which significantly determines the extent of the disease. In an attempt to overcome this, the proposed study presents a lesion-based hybrid VGG16-LSTM model to classify diabetic retinopathy severity. Within this concept, the VGG16 network is used as a deep feature extractor, which retrieves discriminative spatial features of retinal fundus images. Rather than flattening these feature maps, the extracted spatial representations are converted into a sequential representation that is ordered, but does not destroy the spatial localization of localized retinal areas. The output of the sequence feature vectors is then fed through an LSTM network that learns inter-region associations between lesion features throughout the retina.

The proposed model, as opposed to the traditional CNN-based methods, in which retinal images are considered independent spatial samples, facilitates the contextual relationships between lesion areas, thus enabling the system to learn more of the progression patterns relative to various stages of severity of DR. In addition, the efficiency of the suggested hybrid architecture is measured by comparative experiments of standalone deep learning models under the same training conditions. It can be shown by experimental analysis that VGG16-based spatial feature extraction and LSTM-based sequential learning can be combined to improve the accuracy of classification and the ability to discriminate between visually similar DR stages. This validates that the use of spatial relationships modeling of the retinal lesion areas has a significant contribution in enhancing the severity classification of diabetic retinopathy.

7. Conclusion

This paper proposed a deep learning-based system of automated diabetic retinopathy detection by retinal fundus images. The system suggested is based on a hybrid VGG16-LSTM architecture, where VGG16 network is used to obtain discriminative spatial features in retinal images and LSTM network to obtain sequential relationships between lesion patterns across different retinal areas. Combining the two models allows the system to examine the spatial lesions property and the architecture between the retinal areas and the surrounding environment, leading to a higher level of classification in the severity of diabetic retinopathy. As it has been experimentally tested, the proposed VGG16 -LSTM hybrid model demonstrates better performance than single deep learning models. The proposed system had a validation accuracy of 96.24, which is better than standalone CNN and LSTM models whose validation accuracies were 85.01% and 90.01, respectively. Moreover, the hybrid model has F1-score weighted-average of 0.95 and macro average of 0.93, which shows the presence of similar classification performance in all the classes of diabetic retinopathy severity. The analysis of performance stability also reveals that the proposed hybrid model obtained an average classification rate of 97.00% with a standard deviation of +0.63 which is much more stable than the individual CNN and LSTM models

would have been. The reduced standard deviation implies better performance and less variation in the course of training. In general, the experimental findings indicate that the suggested VGG16-LSTM hybrid architecture is efficient to acquire both spatial and sequential feature representations, resulting in higher accuracy, robustness, and the generalization ability of the classifier. Such features render the suggested solution to automated diabetic retinopathy screening and clinical decision-support tools.

7.1. Future Scope

Though the suggested VGG16LSTM architecture shows Hispanic outcomes in diabetic retinopathy detection, there are various enhancements that can be clarified in future studies. Future research could be to enhance model interpretability and explainability, which is necessary in clinical adoption. Such techniques as Grad-CAM and attention mechanisms may be applied to indicate the areas of lesions that are predicted by the model. The other direction is the maximization of the computational efficiency whereby the model can be implemented in a real-time screening system or a mobile-based health application. Also, it will be beneficial to add bigger and more heterogeneous datasets of retinal images in order to enhance the generalization of models to various groups of patients and states of imaging.

Additional research can also be carried out into incorporating more sophisticated deep learning models i.e. Vision Transformer, EfficientNet or attention based CNN models to enhance further on the level of DR detection. It will also be necessary to validate AI-based diabetic retinopathy screening systems with real-world hospital datasets to translate them into real-world healthcare implementation.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] F. C. Monteiro, "Diabetic retinopathy grading using blended deep learning," *Procedia Computer Science*, vol. 219, pp. 1097–1104, 2023.

[2] A. A. Glazkov et al., "Skin microvascular reactivity in patients with diabetic retinopathy," *Microvascular Research*, vol. 147, 2023.

[3] T. N. Poly et al., "Artificial intelligence in diabetic retinopathy: Bibliometric analysis," *Computer Methods and Programs in Biomedicine*, vol. 231, 2023.

[4] A. Sebastian et al., "A survey on deep-learning-based diabetic retinopathy classification," *Diagnostics*, vol. 13, 2023.

[5] T. M. Usman et al., "A systematic literature review of machine learning based risk prediction models for diabetic retinopathy progression," *Artificial Intelligence in Medicine*, vol. 143, 2023.

[6] C. Sundhar and D. Archana, "Automatic screening of fundus images for detection of diabetic retinopathy," *International Journal of Communication and Computer Technology*, 2014.

[7] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.

[8] B. M. R. P. and J. B. Raja, "Automated blood vessel segmentation in retinal images," *International Journal of Engineering and Computer Science*, 2014.

[9] M. Kaur and R. Talwar, "Automatic extraction of blood vessel and eye retinopathy detection," *European Journal of Advances in Engineering and Technology*, 2015.

[10] H. Tjandrasa et al., "Classification of non-proliferative diabetic retinopathy based on segmented exudates using K-means clustering," *IJ Image Graphics and Signal Processing*, 2015.

[11] C. J. ME, "Automatic detection and classification of diabetic retinopathy lesions using bag of visual words model," *International Journal of Scientific and Research Publications*, 2015.

[12] C. Sen et al., "New accurate deep learning model for diabetic retinopathy detection utilizing sequential preprocessing and transfer learning," *Biomedical Signal Processing and Control*, 2025.

[13] W. L. Alyoubi et al., "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, 2020.

[14] M. Imran et al., "A unified technique for entropy enhancement based diabetic retinopathy detection using hybrid neural network," *Computers in Biology and Medicine*, vol. 145, 2022.

[15] A. B. Wiratama et al., "Diabetic retinopathy classification based on fundus image using CNN with MobileNetV2," *ICEBEHI Conference*, 2023.

[16] N. Yalçin et al., "Classification of retinal images with deep learning for early detection of diabetic retinopathy," *IEEE SIU Conference*, 2018.

[17] Y. B. Özçelik and A. Altan, "Deep learning based classification of fundus images for diabetic retinopathy diagnosis," *European Journal of Science and Technology*, 2021.

[18] S. Wan et al., "Deep convolutional neural networks for diabetic retinopathy detection," *Computers & Electrical Engineering*, 2018.

[19] M. S. Mazumder et al., "Deep learning approaches for diabetic retinopathy detection by image classification," *ICOSEC Conference*, IEEE, 2022.

[20] B. Hossain, "APTOS 2019 Blindness Detection Dataset," Kaggle, 2019.

[21] U. Das et al., "Detection of diabetic retinopathy using deep learning algorithms," *ICEPE Conference*, IEEE, 2024.

[22] J. Krause et al., "Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy," *Ophthalmology*, 2018.

[23] R. F. Spaide et al., "Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography," *JAMA Ophthalmology*, 2015.

[24] D. Le et al., "Quantitative optical coherence tomography reveals rod photoreceptor degeneration in early diabetic retinopathy," *Retina*, 2022.

[25] A. Russo et al., "Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy," *American Journal of Ophthalmology*, 2015.

[26] Verywell Health, "Diabetic retinopathy: Causes and diagnosis," 2021.

[27] P. Khan and P. Gupta, "Amsler grid test – A simple screening modality for diabetic retinopathy," *SAS Journal of Bioscience*, 2014.

[28] C. L. Lin and K. C. Wu, "Development of revised ResNet-50 for diabetic retinopathy detection," *BMC Bioinformatics*, 2023.

[29] A. Bajwa et al., "CNN-based screening system for diabetic retinopathy severity classification," *Diagnostics*, 2023.

[30] X. Zhang et al., "Automated detection of severe diabetic retinopathy using deep learning method," *Graefe's Archive for Clinical and Experimental Ophthalmology*, 2022.

[31] M. S. Farooq et al., "Computer-aided diagnostic system for screening diabetic retinopathy based on deep learning," *Sensors*, 2022.

[32] M. A. Raiaan et al., "A lightweight robust deep learning model for diabetic retinopathy classification," *IEEE Access*, 2023.

[33] S. Shoaib et al., "Deep learning innovations in diagnosing diabetic retinopathy using transfer learning," 2024.

[34] A. Senapati et al., "Artificial intelligence for diabetic retinopathy detection: A systematic review," *Informatics in Medicine Unlocked*, 2024.

[35] N. Tsiknakis et al., "Deep learning for diabetic retinopathy detection and classification based on fundus images," *Computers in Biology and Medicine*, 2021.

[36] C. Suedumrong et al., "Application of deep convolutional neural networks VGG-16 and GoogLeNet for diabetic retinopathy detection," *Future Technologies Conference*, 2021.

[37] L. Wewetzer et al., "Diagnostic performance of deep learning-based screening methods for diabetic retinopathy," *PLOS ONE*, 2021.

[38] S. Akhtar et al., "A deep learning-based model for diabetic retinopathy grading," *Scientific Reports*, 2025.

[39] M. Hussain, "Enhanced CNN-based mechanism for diagnosis of diabetic retinopathy," *Engineering Technology Applied Science Research*, 2025.

[40] L. Arora et al., "Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy," *Scientific Reports*, 2024.

[41] Z. Bi et al., "Deep learning-based approaches for retinal disease detection," *Medical Imaging Research*, 2025.

[42] V. Anand et al., "Smart grading of diabetic retinopathy using transfer learning models," *Frontiers in Artificial Intelligence*, 2024.

[43] C. Mohanty et al., "Using deep learning architectures for detection and classification of diabetic retinopathy," *Sensors*, 2023.

[44] M. Youldash et al., "Early detection and classification of diabetic retinopathy using machine learning and deep learning," *Journal of Biomedical Imaging*, 2024.

[45] R. Abbasi et al., "Adaptive deep learning model for diabetic retinopathy detection," *Scientific Reports*, 2025.

[46] T. Geetha and C. Hema, "Deep learning-based joint analysis of diabetic retinopathy and glaucoma," *Scientific Reports*, 2026.

[47] Q. Mehmood et al., "Hybrid deep learning model for diabetic retinopathy detection," *Journal of Shifa Medical Center*, 2025.

[48] S. Mounika et al., "Genetic algorithm based deep learning model for diabetic retinopathy detection," *Engineering Technology Applied Science Research*, 2025.

[49] D. Saproo et al., "Deep learning-based binary classification of diabetic retinopathy images," 2024.

[50] E. N. Mutua et al., "Deep learning applications for diagnosing diabetic retinopathy," *International Journal of Medical Informatics*, 2025.