



# A Hybrid Feature Representation Framework with Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory for Diabetic Retinopathy Grading Using Retinal Fundus Images

Sharda M. Dhavale<sup>1,2\*</sup>, Pushpa M. Bangare<sup>3</sup>

<sup>1</sup> Sinhgad College of Engineering, Savitribai Phule Pune University, Pune 411041, India

<sup>2</sup> Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University, Pune 411044, India

<sup>3</sup> SKN College of Engineering, Savitribai Phule Pune University, Pune 411041, India

Corresponding Author Email: [sharda.dhvale@pccoe pune.org](mailto:sharda.dhvale@pccoe pune.org)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310921>

## ABSTRACT

**Received:** 18 June 2026

**Revised:** 26 August 2026

**Accepted:** 4 September 2026

**Available online:** 30 September 2026

### Keywords:

*diabetic retinopathy, dynamic multi-neighbor local ternary pattern, deep learning, Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory*

Diabetic retinopathy (DR) grading from retinal fundus images is essential for early intervention and prevention of vision impairment. However, existing deep learning (DL)-based approaches often suffer from limited interpretability and insufficient representation of complex retinal patterns caused by variations in texture, shape, and lesion characteristics. This study proposes a hybrid feature representation framework integrating handcrafted image descriptors and a deep neural network model for automated DR grading. Texture and shape information are extracted using gray-level co-occurrence matrix (GLCM), histogram of oriented gradients (HOG), and a proposed dynamic multi-neighbor local ternary pattern (DMLTP), which introduces adaptive thresholding based on local intensity variation and contrast information. The extracted features are processed by a Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN-BiLSTM) classifier to capture hierarchical representations and sequential dependencies among retinal features. The proposed framework is evaluated on the publicly available Asia Pacific Tele-Ophthalmology Society (APTOS)-2019 dataset containing five DR severity categories. Experimental results demonstrate that the proposed GLCM+HOG+DMLTP feature combination with DCNN-BiLSTM achieves 98.67% accuracy, precision, recall, and F1-score on the augmented dataset. The proposed approach effectively integrates handcrafted descriptors with deep feature learning, providing an accurate solution for automated multi-class DR grading from retinal fundus images.

## 1. INTRODUCTION

Diabetic retinopathy (DR) is the primary cause of blindness and vision impairment across the world, particularly in underdeveloped and developing countries. In DR, high blood sugar damages the retinal blood vessels [1]. If DR is not detected early, it may lead to blindness or severe vision impairment. The human eye is a crucial organ that enables the brain to perceive visual information [2]. Therefore, early DR detection is imperative to provide prompt treatment, prevent disease progression, and reduce severity. Traditional DR detection is performed manually by examining fundus images, which provide high-resolution views of the retina [3]. The fundus images consist of the optic disc, macula, blood vessels, retina, and pathological features such as exudates, microaneurysms (MAs), neovascularization, macular edema, and hemorrhages [4, 5].

Retinal fundus imaging is more popular for DR because of its cost-effectiveness, non-invasiveness, and lower cost. The ophthalmologist reviews fundus images and identifies the characteristics of pathological lesions, such as hard exudates, soft exudates, MAs, and hemorrhages, to assess disease

severity. The exudates are yellowish-white spots near the macula, resulting from the accumulation of proteins and lipids leaked from damaged blood vessels [6, 7]. The soft exudates are white, fluffy patches caused by ischemic swelling of nerve fibers. MAs are tiny red dots caused by the rupture of capillary walls, and hemorrhages are blot-like or dark red spots generated due to retinal blood vessel rupture. Neovascularization is the abnormal growth of blood vessels due to retinal ischemia. Macular edema is swelling of the macula caused by fluid leakage [7]. Manual screening of DR by an ophthalmologist is time-consuming, labor-intensive, and prone to interobserver variability, making it difficult to screen larger populations. To improve the scalability, efficiency, and reliability of the DR detection system, an automated diagnostic scheme is needed [8, 9].

Various machine learning (ML) and deep learning (DL)-based schemes have been employed for accurate DR diagnosis. Convolutional Neural Networks (CNNs) have improved DR diagnosis performance on fundus images by capturing hierarchical representations and fine-grained texture details. The spatial representation capacity of CNNs helps recognize DR severity levels [10-14].

In recent years, Washburn et al. [15] suggested various DL algorithms for depicting the texture features of the DR images. The DenseNet-121 along with gray-level co-occurrence matrix (GLCM) features provided an overall accuracy of 95% for the Asia Pacific Tele-Ophthalmology Society (APTOS) dataset. However, the huge complexity of the algorithms limits their deployment on standalone devices. Mohan et al. [16] explored multi-scale feature extraction using a CNN architecture for real-time DR grading. The CNN-based system achieves an overall accuracy of 96.62% on the public IDRiD dataset. The system provided a lightweight framework with 2.7 M trainable parameters but failed to recognize complex lesions and suffered from poor interpretability of the model. Agarwal et al. [17] presented a transformer model with focal weighting and an ordinal loss function for five-level DR grading. The DR system achieved an overall accuracy of 84.6% and a micro-F1 score of 0.707 on the APTOS 2019 dataset. The higher computational complexity of transformer models limits their deployment flexibility on standalone devices with limited computational resources.

Further, Abraham and Kovoov [18] proposed an ensemble learning classifier that combines ResNet50, DenseNet-121, and EfficientNet-B0, along with multi-head self-attention, to preserve ordinal relations in the features. The DR system resulted in a Quadratic Weighted Kappa (QWK) score of 91.85% on APTOS 2019. Though the ensemble model helps capture a better local-global feature depiction, it requires longer training time due to higher computational complexity and intricacy of the model. Khandakar et al. [19] also suggested an ensemble classifier based on a feedforward neural network for DR, which provided an accuracy between 98.27% and 99.05% for the Bangladeshi local DR datasets. The model shows strong reliability and generalization for the regional dataset, but has higher computational complexity.

Abbasi et al. [20] implemented a CNN model for DR, along with an adaptive gamma enhancement scheme, to improve the quality of low-contrast images and preserve the structural information of fundus images. The CNN-based classifier achieved 95.88% accuracy on the Messidor dataset for DR, but the system's accuracy is highly dependent on image quality and requires extensive image enhancement. Men et al. [21] explored self-supervised vision transformers (ViTs) for achieving cross-domain generalization. The ViT with multi-source domain fine-tuning provides better generalization, and the heatmap-based representation offers explainability in the model. Sushith et al. [22] suggested the CNN-recurrent neural network (RNN) model for capturing the spatial and temporal characteristics of retinal scans for early DR. The CNN-RNN offered overall accuracies of 97.5% for DRIVE, 94.04% for Kaggle, and 96.9% for EyePACS. However, the system suffers from class imbalance due to a disparity in the training dataset. Delavari et al. [23] suggested VGG-15 pretrained on ImageNet, along with a guided Gradient-weighted Class Activation Mapping (Grad-CAM) explainable model for DR, which results in an area under the curve of 0.72 on the EyePACS dataset. However, the limited feature depiction capability limits the accuracy for the complex DR system. Naveen et al. [24] explored EfficientNetV2 for feature representation of fundus images and a support vector machine with radial basis function (SVM-RBF) for DR. The EfficientNetV2-SVM achieved an overall accuracy of 97.26% on the APTOS dataset for binary classification between DR and normal. The results of the DR system are limited for multiclass DR.

From the analysis of existing DL and ML-based DR detection, the following research gaps limit the effectiveness of DR diagnosis systems. The DL-based DR diagnosis methods require larger, diverse datasets; however, limited healthcare infrastructure and ethical issues limit their availability. The system's interpretability and explainability are poor due to the black-box nature of the DL models. Annotation and labeling of DR severity levels are time-consuming and tedious. The uneven and limited dataset creates class imbalance, leading to disparities in the system's qualitative and quantitative results. The DR diagnosis system has limited generalization capability with different demographic characteristics: genders, age, and ethnicity. Existing CNN, transformer, and ensemble methods show strong DR classification performance but may not adequately exploit complementary handcrafted features depicting the texture, shape, and edge of images along with sequential dependencies. Therefore, the proposed Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN–BiLSTM) integrates handcrafted and deep features to improve discriminative representation and classification robustness.

Thus, this paper presents a DL-based DR system that leverages both traditional handcrafted features and a hybrid DCNN–BiLSTM model. The chief contributions of the article are summarized as follows:

- Development of multiple DR features based on GLCM, histogram of oriented gradients (HOG), and novel dynamic multi-neighbor local ternary pattern (DMLTP). The novel DMLTP considers multiple neighbors to capture a better spatial representation and an adaptive threshold to minimize the effects of contrast variation and noise, enhancing the local texture description of fundus images.

- Development of the hybrid DCNN–BiLSTM-based classifier for DR, where DCNN provides a multilevel hierarchical depiction, and BiLSTM offers the long-term correlation and temporal depiction of the fundus image features.

- The effectiveness of the DR system is assessed on the APTOS 2019 dataset for grading No DR, mild, moderate, proliferative, and severe DR.

The remainder of the article is organized as follows: Section 2 provides details on the DR methodology, focusing on feature representation and the classification model. Section 3 presents the augmented discussion on the simulation results of the DR system. Further, Section 4 delivers the conclusions and future scopes of the suggested DR system.

## 2. METHODOLOGY

The proposed DR pipeline consists of preprocessing, feature extraction, feature selection, and DR classification using a DCNN–BiLSTM framework, as shown in Figure 1. The preprocessing stage consists of converting the original color fundus images to grayscale. Various texture and shape features are extracted using GLCM, HOG, and DMLTP to describe the complex patterns in the local and global regions of the fundus image. The features are provided to the lightweight DCNN–BiLSTM to improve feature representation. The DCNN helps provide high-level correlations across the various texture and shape features of fundus images. The BiLSTM captures long-term dependencies in the texture and shape features of fundus images. Further, a fully connected layer improves connectivity

in the deep features, and a Softmax classifier is used to classify fundus images into DR labels.

Texture and shape features of retinal fundus images are extracted to detect changes caused by exudates. The final

feature vector consists of a total of 35508 features, comprising 12 GLCM features, 900 DMLTP features, and 34596 HOG features.

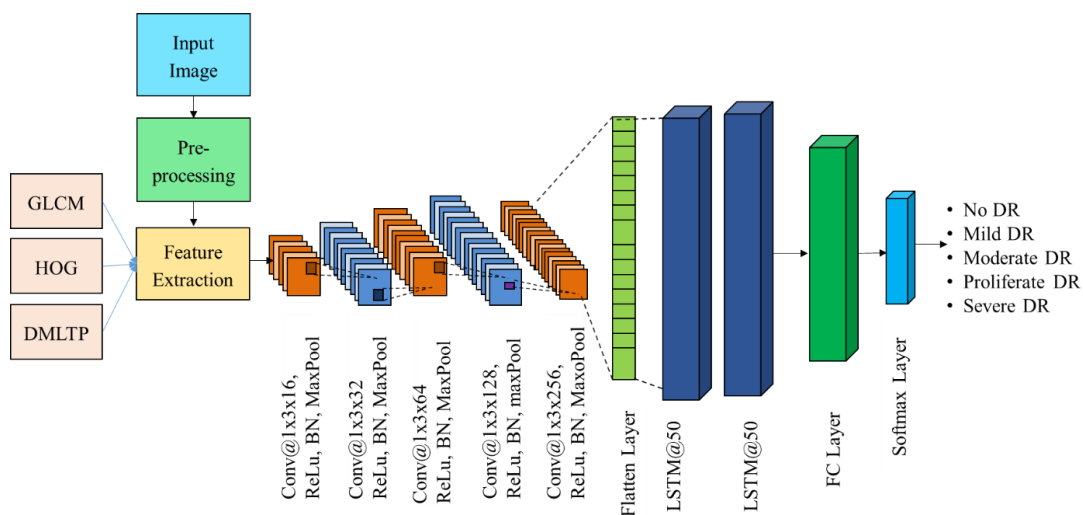
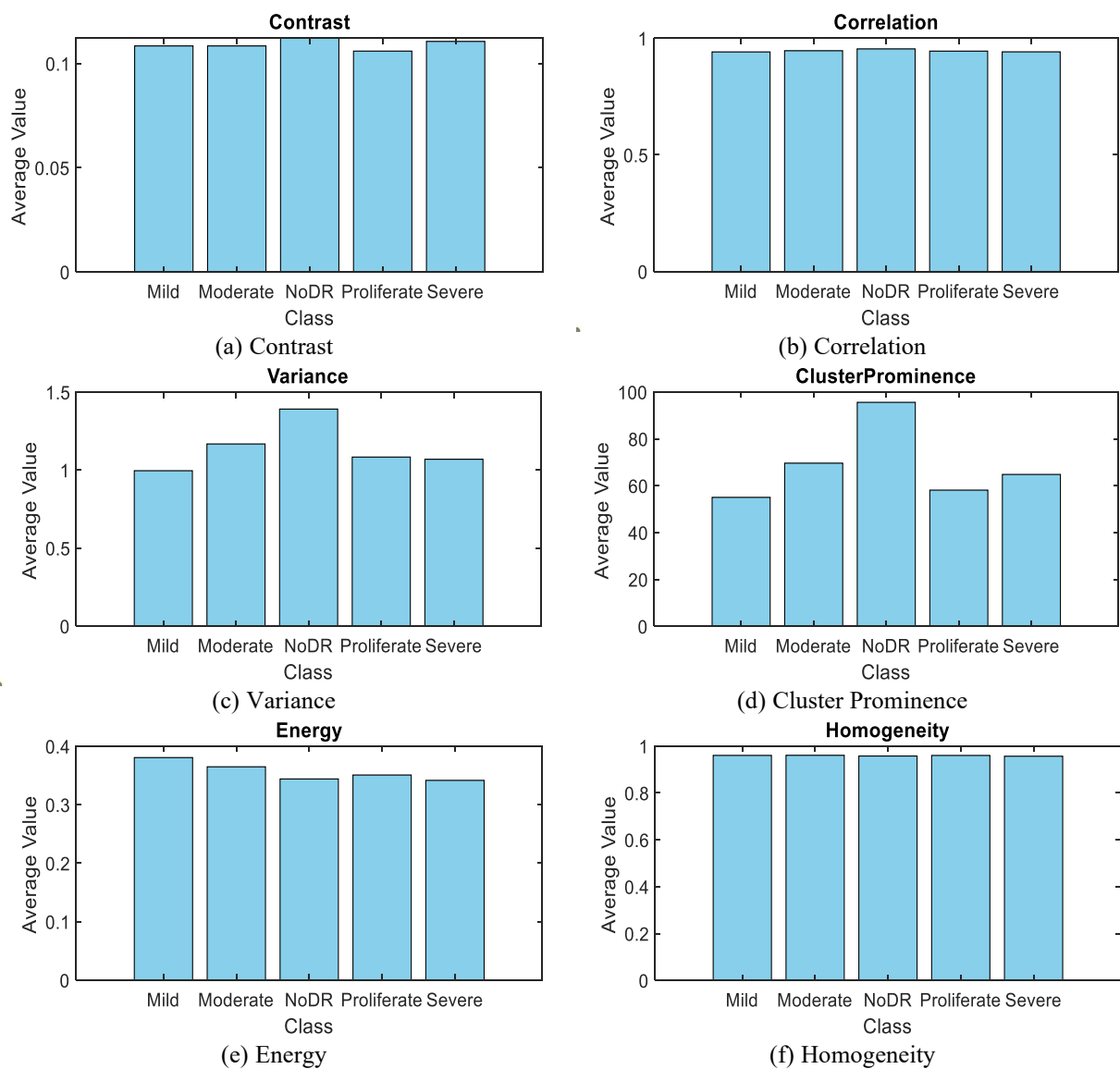


**Figure 1.** Process flow of the proposed diabetic retinopathy (DR) system based on Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN–BiLSTM)



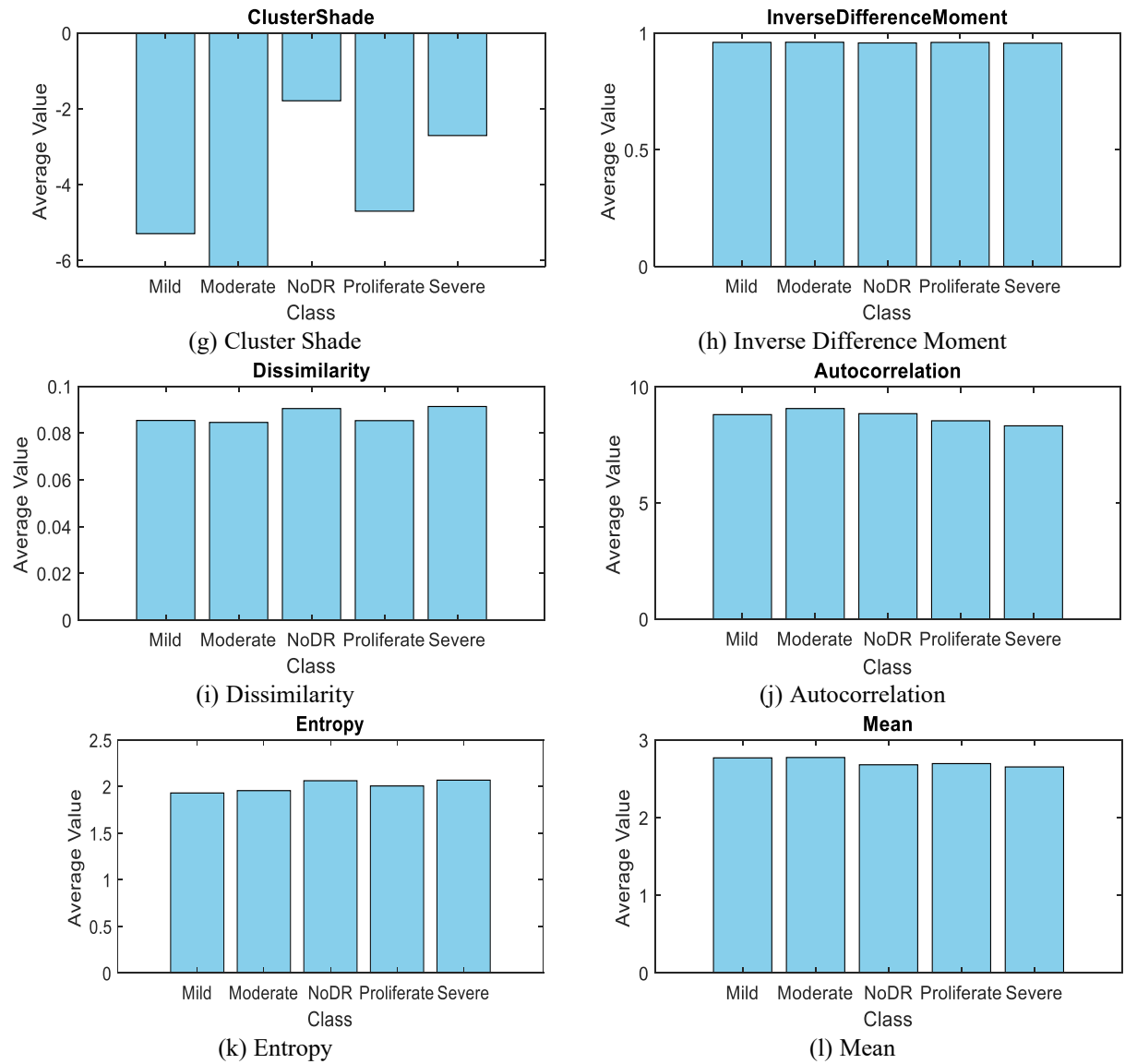


**Figure 2.** Average gray-level co-occurrence matrix (GLCM) feature visualizations for the diabetic retinopathy (DR) images

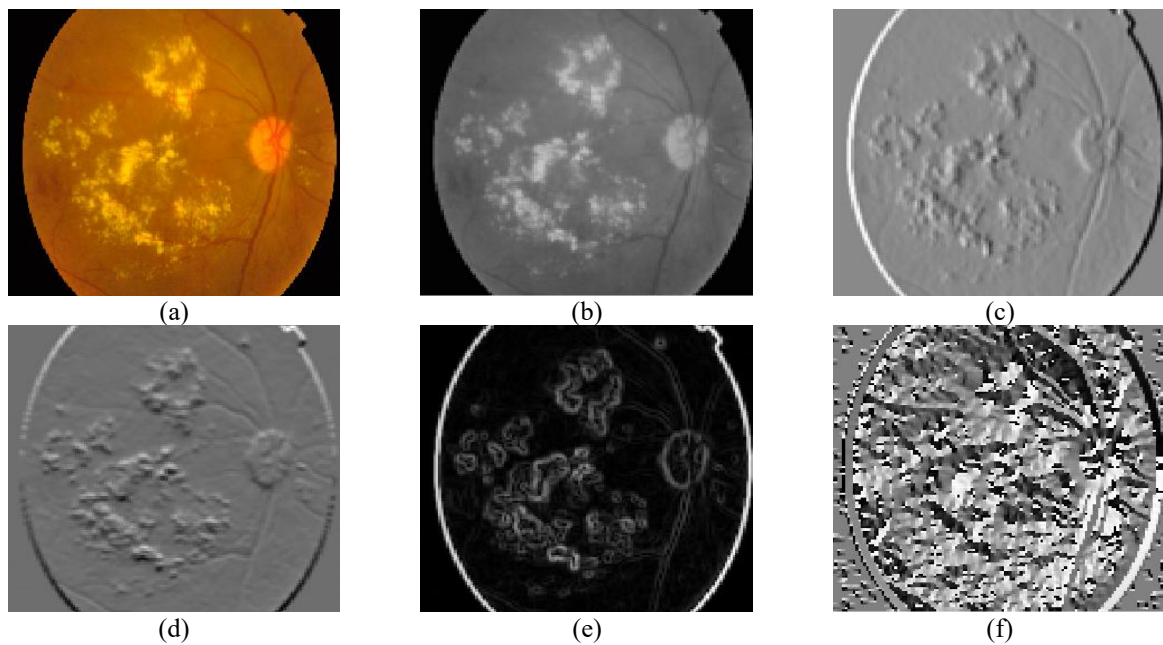


**Figure 3.** Histogram of oriented gradients (HOG) visualization: (a) original fundus image (Severe class-0f96c358a250.png), (b) grayscale image, (c) horizontal gradients of HOG, (d) vertical gradients of HOG, (e) magnitude of gradients of HOG, and (f) gradient orientation of HOG

## 2.1 Gray-level co-occurrence matrix features

The GLCM provides spatial correlation-based global texture features for fundus images based on the cooccurrence of pixel intensities. The GLCM features include energy, entropy, homogeneity, correlation, cluster prominence, variance, cluster shade, inverse difference (ID), contrast, dissimilarity, and autocorrelation of the images. Figure 2 shows visualizations of the average values for different features of DR across the mild, moderate, normal, proliferative, and severe conditions. The variance and cluster prominence show higher values in the normal condition than in the DR condition. The cluster prominence shows a higher value for the normal condition and a lower value for mild and moderate conditions. The various GLCM features exhibit variations across the global texture of fundus images.

## 2.2 Histogram of oriented gradients features

HOG features capture the shape properties of objects in retinal fundus images. The HOG computes the orientation of gradients to characterize the shape disparity of retinal image objects. The HOG features are extracted for a  $2 \times 2$  cell block size, an  $8 \times 8$  pixel cell size, 50% block overlap, and 9-bin

histograms per cell, resulting in a total of 34596 features for an image size of  $256 \times 256$  pixels. The HOG feature visualizations are shown in Figure 3.

## 2.3 Dynamic multi-neighbor local ternary pattern features

The local binary pattern (LBP) is a widely used texture descriptor that provides scale- and rotation-invariant features for images. It helps analyze fine changes in the local regions of images. The LBP is based on a binary pattern obtained by comparing the central pixel with its neighboring pixels in a local  $3 \times 3$  window. The effectiveness of the LBP is challenging due to smaller fluctuations in local regions caused by contrast and noise, which lead to higher pattern changes. The local ternary pattern (LTP) provides ternary pattern features that provide robustness against contrast variations and noise.

However, the effectiveness of the LTP is limited by a manual, fixed threshold and poor spatial correlation arising from a single neighborhood. The proposed DMLTP considers a dynamic threshold based on the local window's standard deviation and contrast to capture fine variations across the image region.

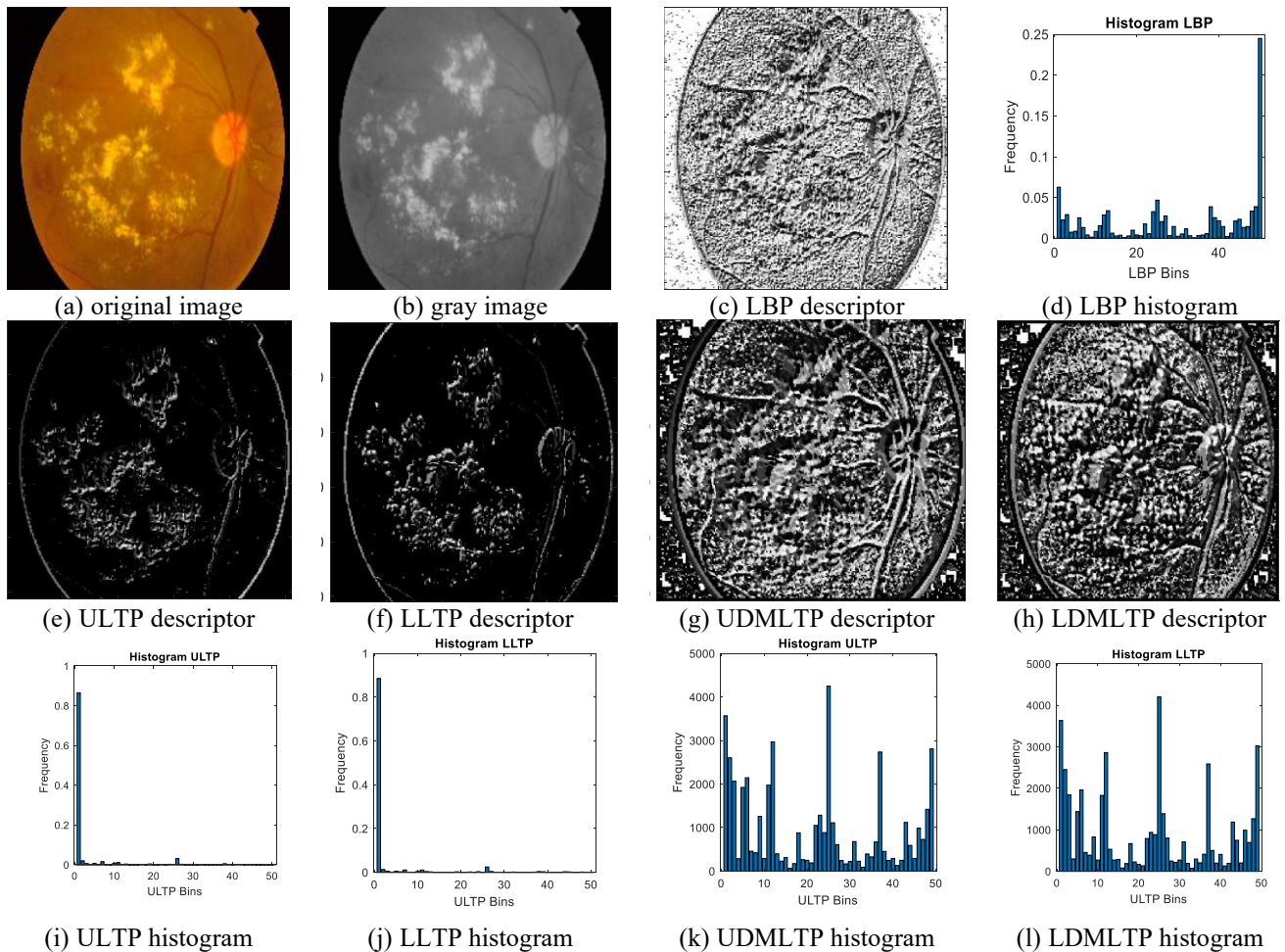


**Figure 4.** Visualizations of LBP, LTP and DMLTP descriptors and histograms for 50 bins for  $N = 1$  block, (a) original image, (b) grayscale image, (c) LBP descriptor, (d) LBP histogram, (e) ULTP descriptor of LTP, (f) LLTP descriptor of LTP, (g) UDMLTP descriptor of MLTP, (h) LDMLTP descriptor of MLTP, (i) ULTP histogram of LTP, (j) LLTP histogram of LTP, (k) UDMLTP histogram of DMLTP and (l) LDMLTP histogram of DMLTP

Note: LBP = local binary pattern; LTP = local ternary pattern; DMLTP = dynamic multi-neighbor local ternary pattern; ULTP = upper local ternary pattern; LLTP = lower local ternary pattern; UDMLTP = upper dynamic multi-neighbor local ternary pattern; LDMLTP = lower dynamic multi-neighbor local ternary pattern; MLTP = multi-neighbor local ternary pattern.

The LTP considers the local region of the  $3 \times 3$  pixels and compares the neighboring pixels with the centered pixel ( $cv$ ) of the local region. It considers the manual threshold ( $\tau$ ) between 3 and 15 for the computation of LTP. If the value of the neighboring pixel ( $x$ ) is greater than the  $cv + \tau$ , then the LTP value is set to +1. If the neighboring pixel value lies between  $cv - \tau$  to  $cv + \tau$ , then the LTP value is set to 0. When the neighboring pixel value is less than  $cv - \tau$ , the LTP is set to -1. The upper LTP and Lower LTP patterns are provided to capture the pattern for brighter and darker regions of the fundus. The upper LTP is constructed by altering all -1 to 0, and the lower LTP is formed by changing -1 to 1 and 1 to 0.

$$ltp(x) = \begin{cases} -1 & \text{if } x < cv - \tau \\ 0 & \text{if } cv - \tau \leq x \leq cv + \tau \\ 1 & \text{if } x > cv + \tau \end{cases} \quad (1)$$

$$ULTP(x) = \begin{cases} 1, & LTP(x) > cv + \tau \\ x, & \text{otherwise} \end{cases} \quad (2)$$

$$LLTP(x) = \begin{cases} 1, & LTP(x) < cv - \tau \\ x, & \text{otherwise} \end{cases} \quad (3)$$

The proposed DMLTP considers three neighbors when computing the neighborhood spatial correlation, as given in Eq. (4). Here,  $A_i$ ,  $B_i$  and  $C_i$  denote the pixels at radius of  $R1 = 1$ ,  $R2 = 2$ , and  $R3 = 3$ . A higher weightage is given to the immediate neighbors to retain the spatial correlation of the local region.

$$x = (A_i - cv) \times R3 + (B_i - cv) \times R2 + (C_i - cv) \times R1 \quad (4)$$

The dynamic threshold is computed using Eq. (5), which considers the standard deviation (SD) and contrast measure (CT) to enable the LTP pattern to be aware of intensity and contrast variations in the local region. The SD for the local region of the retinal fundus image (RF) having  $M$  rows and  $N$  columns is computed using Eqs. (6) and (7), where  $\mu$  denotes the mean of the local patch.

$$\tau a = \alpha_1 * SD + \alpha_2 * CT \quad (5)$$

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (RF(i, j) - \mu)^2} \quad (6)$$

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N RF(i, j) \quad (7)$$

The contrast provides local contrast variations using Eq. (8), which considers the maximum and minimum intensities of the local patch.

$$CT = \max_{(i,j) \in W} RF(i, j) - \min_{(i,j) \in W} RF(i, j) \quad (8)$$

The  $n$ -bin histogram is computed over a block size of  $N$ , which splits the image into local blocks of  $N \times N$  regions, each with  $(R/N)$  rows and  $(C/N)$  columns. The block-based LTP helps capture better local variations to minimize the impact of contrast and illumination variations in the fundus image.

Figure 4 shows the various descriptors for the fundus image

from the severe class, where the LBP descriptor shows local variations but fails to capture fine texture changes and is unable to distinguish the salient region in the fundus image. The LTP provides a texture description of the salient regions of the fundus image. The histogram of LTP is uneven, with an accumulation towards lower intensities. The DMLTP provides superior texture description and offers a better representation of the fine and salient regions in fundus images. The DMLTP histogram shows uniform variability across the bins due to better feature representation. The algorithms for the DMLTP are provided in Algorithm 1.

---

**Algorithm 1.** Dynamic multi-neighbor local ternary pattern (DMLTP) for fundus image texture feature extraction

---

**Input:** Fundus image (im), number of blocks (Nb), number of bins (NBins)

**Output:** DMLTP histogram features

1. Read fundus image
  2. Convert color fundus image to grayscale
  3. Apply zero padding for fundus images
  4. Set neighbor radius (R)
  5. For radius 1: R
  6. Consider a local window of  $w \times w$  pixels
  7. Compute dynamic threshold ( $\tau a$ ) using Eq. (5).
  8. Construct upper LTP as UDMLTP
  9. Construct lower LTP as LDMLTP
  10. Compute  $NBins$  histogram over block of Nb for UDMLTP as  $H_{UDMLTP}$
  11. Compute  $NBins$  histogram over block of Nb for LDMLTP as  $H_{LDMLTP}$
  12. Combine DMLTP =  $\{H_{UDMLTP}, H_{LDMLTP}\}$
- Return DMLTP histogram as final fundus image texture feature vector
- 

## 2.4 Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory

The DCNN-BiLSTM classifier consists of five convolution layers with 16, 32, 64, 128, and 256 filters with a filter size of  $1 \times 3$ . The convolution layers capture the multilevel hierarchical correlation in the handcrafted features. The convolution layers are followed by a rectified linear unit (ReLU) layer, batch normalization (BN), and maximum pooling (MaxPool) layers. The ReLU increases the nonlinearity of the features, BN standardizes the features, and MaxPool reduces the feature dimensions of the deep features. Further, the features are flattened and fed into two BiLSTM layers with 50 hidden units to acquire the long-term connectivity in features. Lastly, a fully connected layer followed by a Softmax classifier is utilized to improve feature connectivity and classification, respectively.

## 3. EXPERIMENTAL RESULTS AND DISCUSSION

The suggested system is implemented using MATLAB R2025b on a personal computer with 20 GB RAM, a Core i7 processor, and 8 GB of graphics memory. The dataset is split in a 70:30 ratio for training and testing for experimental evaluation. Also, the effectiveness of the DR system is validated using 10-fold cross-validation to analyse the statistical significance of the methods. The hyperparameters of the GLCM, DMLTP, HOG, and DCNN-BiLSTM model are provided in Table 1.

**Table 1.** Parameter specification of methods

| Parameter                                         | Specification                |
|---------------------------------------------------|------------------------------|
| Gray-level co-occurrence matrix (GLCM) parameters |                              |
| Distance                                          | 1 pixel                      |
| Angle                                             | 0°, 45°, 90°, 135°           |
| Gray levels                                       | 256                          |
| DMLTP parameters                                  |                              |
| Radius                                            | 3                            |
| $\alpha_1$                                        | 0.5                          |
| $\alpha_2$                                        | 0.5                          |
| Block size                                        | 3                            |
| Number of bins                                    | 50                           |
| Histogram of oriented gradients (HOG) parameters  |                              |
| Cell size                                         | 8 × 8 pixels                 |
| Block size                                        | 2 × 2 cells                  |
| Orientation bins                                  | 9                            |
| Orientation                                       | 0 to 180°                    |
| DCNN–BiLSTM parameters                            |                              |
| Optimization method                               | Adam                         |
| Learning rate                                     | 0.001                        |
| Epoch                                             | 100                          |
| Batch size                                        | 32                           |
| Dropout                                           | 0.5                          |
| Momentum                                          | 0.8                          |
| Convolution filters                               | 16, 32, 64, 128, and 256     |
| Filter size                                       | 1 × 3                        |
| Activation function                               | Rectified linear unit (ReLU) |
| Pooling strategy                                  | Maximum pooling              |
| BiLSTM units                                      | 2 Layers with 50 units       |

Note: DMLTP = dynamic multi-neighbor local ternary pattern, DCNN–BiLSTM = Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory.

**Table 2.** Details about the diabetic retinopathy (DR) fundus image dataset

| Class         | Original Samples | Training Samples (70%) | Testing Samples (30%) | Augmented Samples |
|---------------|------------------|------------------------|-----------------------|-------------------|
| No DR         | 1805             | 1264                   | 541                   | 1264              |
| Mild          | 370              | 259                    | 111                   | 1264              |
| Moderate      | 999              | 699                    | 300                   | 1264              |
| Proliferate   | 295              | 206                    | 89                    | 1264              |
| Severe        | 193              | 135                    | 58                    | 1264              |
| Total samples | 3662             | 2563                   | 1099                  | 6320              |

### 3.1 Dataset

The effectiveness of the DR system is evaluated on the APTOS 2019 DR dataset, which consists of five classes: No DR, Mild, Moderate, Proliferate, and Severe [25, 26]. The images are augmented using cropping, rotation, noise addition, and scaling to minimize the class imbalance problem. The original training data samples are augmented to 1264 samples per class, and the testing samples are kept the same to avoid biased evaluation and information leakage. The original

dataset consists of 2563 samples for training, and the augmented dataset includes 6320 samples for training. The class-wise distribution of the dataset is provided in Table 2.

### 3.2 Discussion on results

The performance of the DR system is evaluated for the original as well as augmented dataset for the DCNN–BiLSTM classifier as given in Table 3 and Figures 5–8, respectively. The combination of the GLCM+HOG+DMLTP features shows improved accuracy of 98.67%, recall of 98.67%, precision of 98.67%, and an F1-score of 98.67%, indicating a good balance between recall and precision due to the augmented dataset. The DCNN–BiLSTM achieves 97.14% overall accuracy and 97.14% recall on the original dataset; however, it shows poor precision (93.02%) and an F1-score of 95.04% due to a disparity in the training dataset. The DCNN–BiLSTM, along with DMLTP, shows improved accuracy of 91.82% compared to traditional texture feature extraction techniques such as LBP (80.24%), LTP (83.33%), multi-neighbor local ternary pattern (MLTP) (87.06%), and HOG (90.59%) on the original dataset. However, DCNN–BiLSTM shows poor F1-scores of 88.21% for DMLTP, 87.50% for HOG, 83.62% for MLTP, 79.22% for LTP, and 76.34% for LBP, due to a large disparity in recall and precision across classes in the original dataset. For the augmented dataset, the DCNN–BiLSTM achieves an improved accuracy of 93.56% for DMLTP, compared with 92.86% for HOG, 89.87% for MLTP, 86.66% for LTP, and 84.49% for LBP.

The confusion matrix for the DCNN–BiLSTM-based DR system is shown in Figure 9. The performance of the DCNN–BiLSTM-based classifier shows a clear improvement when trained on the augmented dataset compared to the original dataset, as given in Table 4 and Figure 10. For the original dataset, the model achieves an average precision of 93.02%, recall of 97.14%, F1-score of 95.04%, and accuracy of 97.14%, indicating strong detection capability but slightly lower precision for certain classes. For the augmented dataset, the DCNN–BiLSTM provides an overall accuracy of 98.5%, recall of 97.93%, precision of 96.63%, and F1-score of 97.78%. However, the Severe class shows relatively lower precision (84.13%) despite a high recall (98.15%), suggesting more false positives in this category. The No DR class achieves 98.59% precision and 99.20% recall, while the Moderate class reaches 99.07% precision and 98.40% recall, maintaining excellent consistency. The Proliferate class improves from 91.11% precision to 98.81%, and overall accuracy across classes rises to around 98.42%–98.89%. The DCNN–BiLSTM provides lower precision of 84.13% and an F1-score of 90.60% for the original data. However, augmentation helps to improve the balance between recall and precision and helps to improve precision to 98.56% and the F1-score to 98.56% for the Severe class.

**Table 3.** Performance comparison of DCNN–BiLSTM-based diabetic retinopathy (DR) for various features for the original and augmented dataset

| Method     | Original Dataset |        |          |          | Augmented Dataset |        |          |          |
|------------|------------------|--------|----------|----------|-------------------|--------|----------|----------|
|            | Precision        | Recall | F1-Score | Accuracy | Precision         | Recall | F1-Score | Accuracy |
| LBP        | 72.82            | 80.24  | 76.34    | 80.24    | 76.19             | 84.49  | 80.13    | 84.49    |
| LTP        | 75.50            | 83.33  | 79.22    | 83.33    | 77.88             | 86.66  | 82.04    | 86.66    |
| MLTP       | 80.43            | 87.06  | 83.62    | 87.06    | 83.35             | 89.87  | 86.49    | 89.87    |
| HOG        | 84.62            | 90.59  | 87.50    | 90.59    | 85.47             | 92.86  | 89.02    | 92.86    |
| DMLTP      | 84.87            | 91.82  | 88.21    | 91.82    | 86.50             | 93.56  | 89.89    | 93.56    |
| GLCM+DMLTP | 84.75            | 92.59  | 88.50    | 92.59    | 88.56             | 93.63  | 91.03    | 93.63    |

|                |       |       |       |       |       |       |       |       |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|
| GLCM+HOG       | 88.56 | 93.63 | 90.89 | 93.63 | 90.53 | 94.44 | 92.45 | 94.44 |
| HOG+DMLTP      | 87.29 | 93.64 | 90.35 | 93.64 | 91.51 | 95.63 | 93.53 | 95.63 |
| GLCM+HOG+DMLTP | 93.02 | 97.14 | 95.04 | 97.14 | 98.67 | 98.67 | 98.67 | 98.67 |

Note: GLCM = gray-level co-occurrence matrix, DMLTP = dynamic multi-neighbor local ternary pattern, HOG = histogram of oriented gradients, DCNN–BiLSTM = Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory.

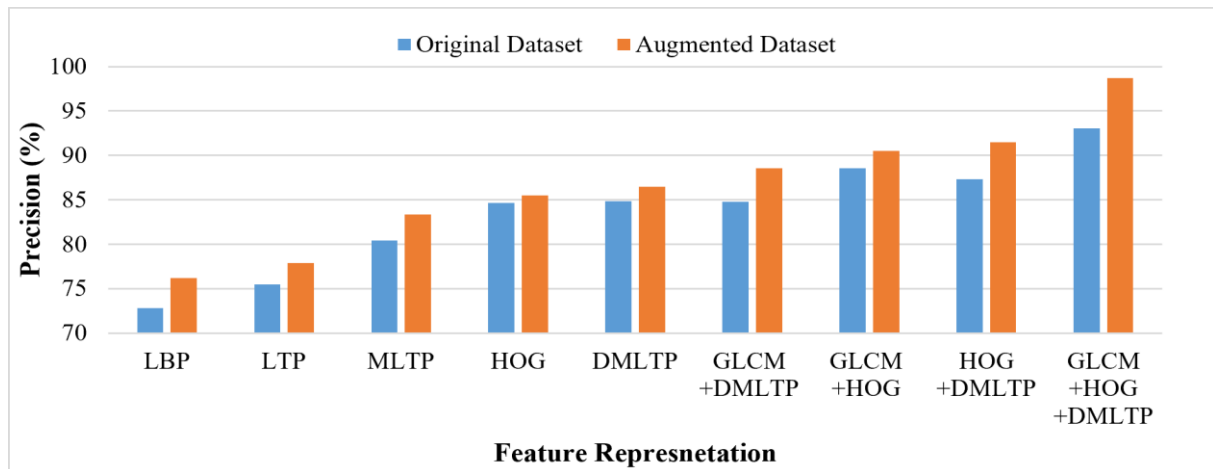


**Figure 5.** Precision comparison of Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN–BiLSTM) based diabetic retinopathy (DR) system for various features

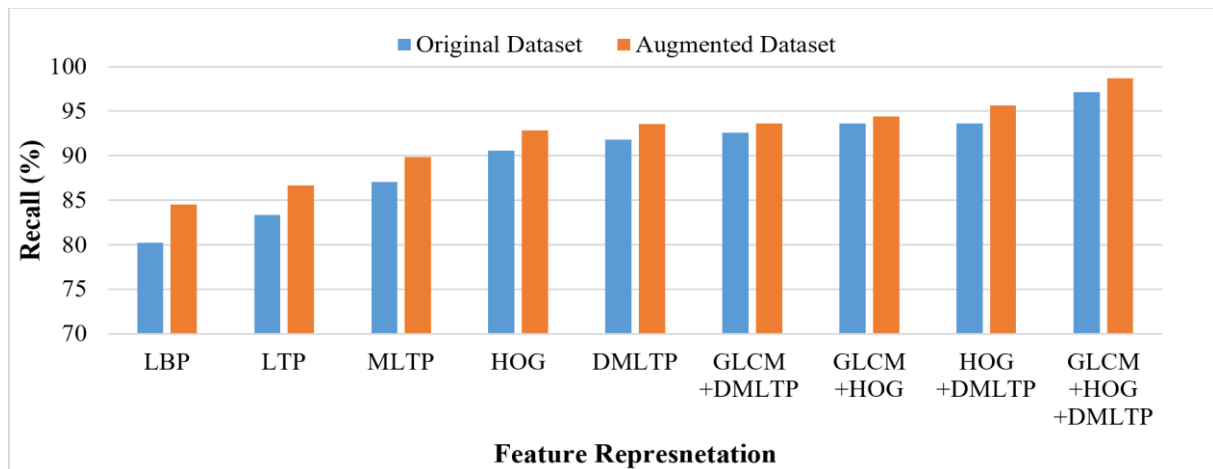


**Figure 6.** Recall comparison of Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN–BiLSTM) based diabetic retinopathy (DR) system for various features

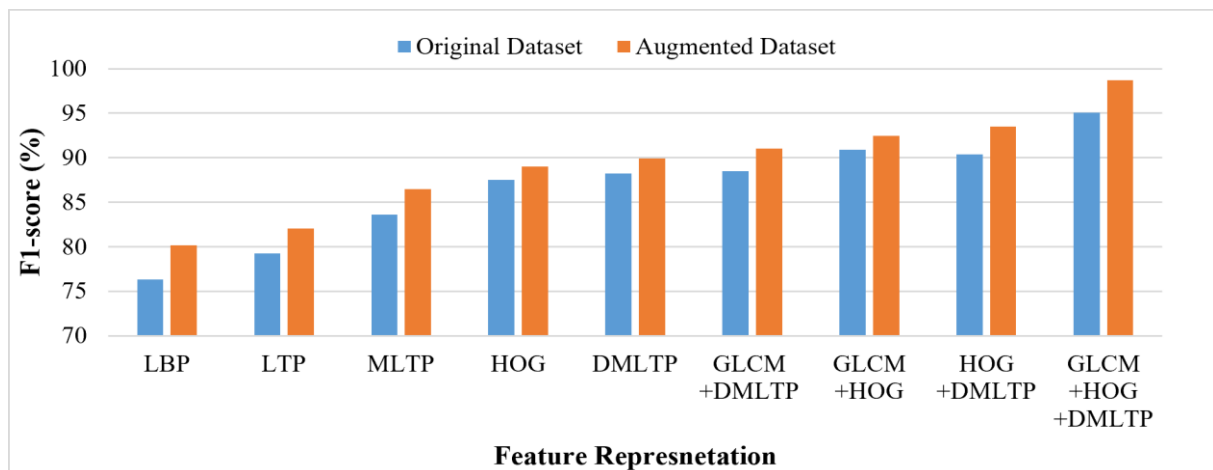


**Figure 7.** F1-score comparison of Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN–BiLSTM) based diabetic retinopathy (DR) system for various features

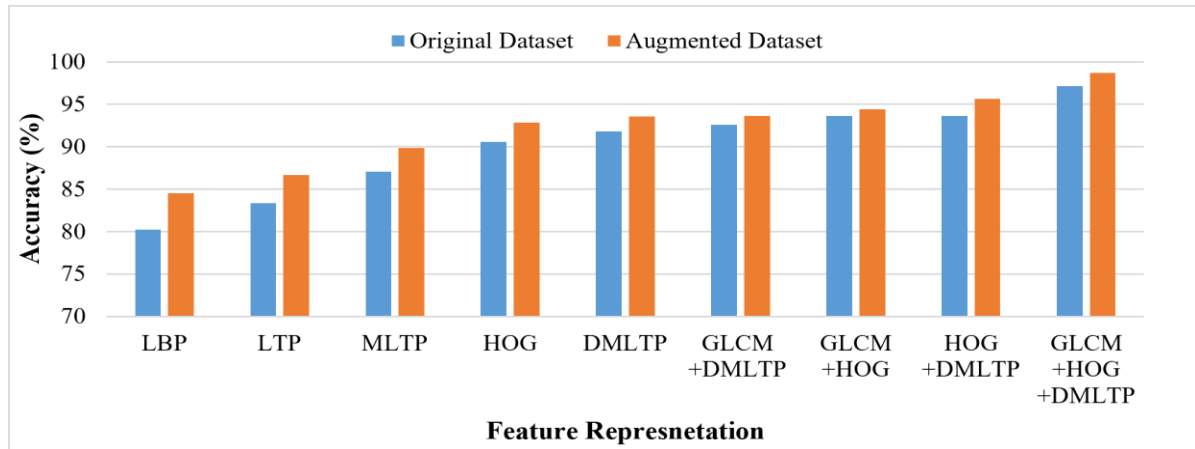


**Figure 8.** Accuracy comparison of Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN–BiLSTM)-based diabetic retinopathy (DR) system for various features

**Table 4.** Performance comparison of DCNN–BiLSTM-based diabetic retinopathy (DR) for GLCM+HOG+DMLTP features for various DR classes for the original and augmented dataset

| Disease Class | Original Dataset |        |          |          | Augmented Dataset |        |          |          |
|---------------|------------------|--------|----------|----------|-------------------|--------|----------|----------|
|               | Precision        | Recall | F1-Score | Accuracy | Precision         | Recall | F1-Score | Accuracy |
| No DR         | 99.23            | 95.71  | 97.44    | 95.71    | 99.26             | 99.07  | 99.17    | 99.17    |
| Mild          | 92.24            | 97.27  | 94.69    | 97.27    | 97.30             | 97.30  | 97.30    | 97.30    |
| Moderate      | 98.40            | 98.09  | 98.25    | 98.09    | 98.33             | 98.33  | 98.33    | 98.33    |
| Proliferate   | 91.11            | 96.47  | 93.71    | 96.47    | 96.70             | 96.70  | 96.70    | 96.70    |
| Severe        | 84.13            | 98.15  | 90.60    | 98.15    | 96.55             | 98.25  | 97.39    | 97.39    |
| Average       | 93.02            | 97.14  | 95.04    | 97.14    | 96.63             | 97.93  | 97.78    | 98.50    |

Note: GLCM = gray-level co-occurrence matrix, DMLTP = dynamic multi-neighbor local ternary pattern, HOG = histogram of oriented gradients, DCNN–BiLSTM = Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory, ReLU = rectified linear unit layer.

|                 |             |               |               |               |               |               |               |
|-----------------|-------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Actual Class    | No DR       | 535<br>48.7%  | 1<br>0.1%     | 2<br>0.2%     | 1<br>0.1%     | 0<br>0.0%     | 99.3%<br>0.7% |
|                 | Mild        | 2<br>0.2%     | 108<br>9.8%   | 1<br>0.1%     | 0<br>0.0%     | 0<br>0.0%     | 97.3%<br>2.7% |
|                 | Moderate    | 2<br>0.2%     | 2<br>0.2%     | 295<br>26.8%  | 1<br>0.1%     | 0<br>0.0%     | 98.3%<br>1.7% |
|                 | Proliferate | 1<br>0.1%     | 0<br>0.0%     | 1<br>0.1%     | 88<br>8.0%    | 1<br>0.1%     | 96.7%<br>3.3% |
|                 | Severe      | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.1%     | 1<br>0.1%     | 56<br>5.1%    | 96.6%<br>3.4% |
|                 |             | 99.1%<br>0.9% | 97.3%<br>2.7% | 98.3%<br>1.7% | 96.7%<br>3.3% | 98.2%<br>1.8% | 98.5%<br>1.5% |
|                 |             | No DR         | Mild          | Moderate      | Proliferate   | Severe        |               |
| Predicted Class |             |               |               |               |               |               |               |

**Figure 9.** Confusion matrix for Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN–BiLSTM)-based diabetic retinopathy (DR) system for augmented dataset

The accuracy of the DMLTP-based DR system is evaluated using the DCNN–BiLSTM for various block sizes, as shown in Table 5. Increasing the number of blocks helps capture more fine details in fundus images. The DCNN–BiLSTM-based DR system achieves 86.66% for LTP, 89.87% for MLTP, and 93.56% for DMLTP on the augmented dataset for  $N = 3$ . However, it resulted in overall accuracies of 83.33% for LTP, 87.06% for MLTP, and 91.82% for DMLTP on the original dataset. Increasing the block size increases the feature size, as

LTP, MLTP, and DMLTP each provide 100 features per block. It offers 100, 400, 900, 1600, and 2500 features for block sizes of 1 to 5, respectively. The DMLTP achieves overall DR accuracy of 84.78% for  $N = 1$ , 86.87% for  $N = 2$ , 93.56% for  $N = 3$ , 92.63% for  $N = 4$ , and 93.56% for  $N = 5$  on the augmented dataset. The proposed implementation uses a block size of  $N = 3$  and 50 bins for histogram-based computation of LTP variants.

The statistical significance of the system is evaluated for 10-

fold cross-validation based on mean accuracy, standard deviation, and 95% confidence interval (CI). It analyzes the impact of features on different DR diagnosis algorithms such as BiLSTM, DCNN, and DCNN-BiLSTM for GLCM+HOG+DMLTP as given in Table 6. The DCNN-BiLSTM achieved a mean accuracy of 98.50% with a relatively low standard deviation of 0.89, resulting in a narrow 95% CI of 97.39%–99.61%. In comparison, the BiLSTM achieved 89.88% accuracy with a wider CI of 86.84%–92.92%, while the DCNN obtained 95.20% accuracy with a CI of 93.62%–96.78%. The narrower CI of the DCNN-BiLSTM indicates lower variation across repeated experiments and suggests that its performance is more stable and dependable than the individual BiLSTM and DCNN models.

The results of the DR system are compared with various existing algorithms, as given in Table 7. The transformer model provided by Agarwal et al. [17] achieved 85.6% accuracy on raw images, whereas the EfficientNet+SVM-based DR model achieved 97.27% accuracy on the APTOS 2019 dataset. The OcuVit [27], ResNet-25 [28], and DenseNet-121 [15] offered overall accuracies of 95%, 91%, and 95% for the APTOS dataset for DR diagnosis. For raw images, the suggested DCNN-BiLSTM provides an average accuracy of 92.35% for original images and 94.80% for augmented samples. The proposed DCNN-BiLSTM-based DR system achieves 98.67% and 97.14% accuracy on the augmented and original datasets, respectively, using GLCM+HOG+DMLTP features. The DMLTP features help improve the texture representation of retinal fundus images, capturing fine changes in image texture due to exudates and MAs.

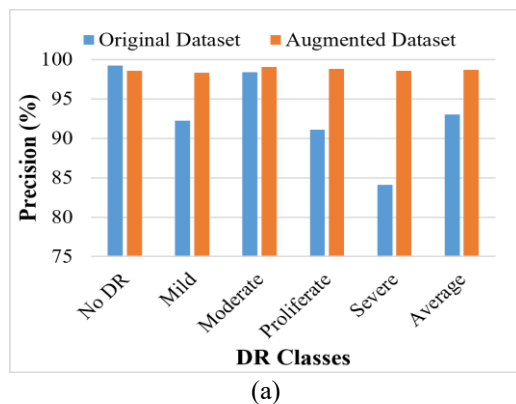


**Figure 10.** Performance visualizations of the Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory (DCNN-BiLSTM)-based diabetic retinopathy (DR) system for various DR classes: (a) precision, (b) recall, (c) F1-score, and (d) accuracy

**Table 5.** Effect of the number of blocks for the texture descriptor for diabetic retinopathy (DR)

| No of Blocks | Number of Features | Original Dataset |       |       | Augmented Dataset |       |       |
|--------------|--------------------|------------------|-------|-------|-------------------|-------|-------|
|              |                    | LTP              | MLTP  | DMLTP | LTP               | MLTP  | DMLTP |
| 1            | 100                | 74.24            | 79.87 | 82.68 | 78.63             | 82.37 | 84.78 |
| 2            | 400                | 77.53            | 82.41 | 85.62 | 81.81             | 85.62 | 86.87 |
| 3            | 900                | 83.33            | 87.06 | 91.82 | 86.66             | 89.87 | 93.56 |
| 4            | 1600               | 82.64            | 86.50 | 90.10 | 85.00             | 88.56 | 92.63 |
| 5            | 2500               | 83.33            | 87.06 | 91.82 | 86.66             | 89.87 | 93.56 |

Note: LTP = local ternary pattern; MLTP = multi-neighbor local ternary pattern; DMLTP = dynamic multi-neighbor local ternary pattern.

**Table 6.** Statistical results comparison for different methods for diabetic retinopathy (DR) diagnosis for GLCM+HOG+DMLTP features

| Authors and Year | Mean Accuracy (%) | Standard Deviation | 95% Confidence Interval |
|------------------|-------------------|--------------------|-------------------------|
| BiLSTM           | 89.88%            | 2.45               | [86.84, 92.92]          |
| DCNN             | 95.20%            | 1.27               | [93.62, 96.78]          |
| DCNN-BiLSTM      | 98.50%            | 0.89               | [97.39, 99.61]          |

Note: GLCM = gray-level co-occurrence matrix, DMLTP = dynamic multi-neighbor local ternary pattern, HOG = histogram of oriented gradients, DCNN-BiLSTM = Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory, ReLU = rectified linear unit layer.

**Table 7.** Accuracy comparison of DCNN–BiLSTM-based diabetic retinopathy (DR) method with existing techniques

| References                | Features       | Classifier                   | Accuracy |
|---------------------------|----------------|------------------------------|----------|
| Washburn et al. [15]      | GLCM           | DenseNet-121                 | 95.00%   |
| Agrawal et al. [17]       | Raw Images     | Transformer Model            | 84.6%    |
| Naveen et al. [24]        | EfficientNetv2 | SVM                          | 97.26%   |
| Ahmed and Uddin [27]      | Raw Images     | OcuViT                       | 95.00%   |
| Vallukappully et al. [28] | Raw Images     | Resnet-25                    | 91.00%   |
|                           | Raw images     | DCNN–BiLSTM (Original Data)  | 92.35%   |
|                           | Raw images     | DCNN–BiLSTM (Augmented Data) | 94.80%   |
| Proposed Method           | GLCM+HOG+DMLTP | DCNN–BiLSTM (Original Data)  | 97.14%   |
|                           | GLCM+HOG+DMLTP | DCNN–BiLSTM (Augmented Data) | 98.67%   |

Note: GLCM = gray-level co-occurrence matrix, DMLTP = dynamic multi-neighbor local ternary pattern, HOG = histogram of oriented gradients, DCNN–BiLSTM = Deep Convolutional Neural Network and Bidirectional Long Short-Term Memory.

#### 4. CONCLUSIONS AND FUTURE SCOPE

This paper presents a DCNN–BiLSTM-based framework for detecting mild, moderate, proliferative, severe, and normal classes of DR. The DCNN–BiLSTM-based DR system considers texture and shape features extracted using GLCM, DMLTP, and HOG techniques. The proposed DMLTP-based texture descriptor provides a better alternative to LBP and LTP, which capture fine and salient regions in fundus images. The DCNN–BiLSTM-based DR system achieves 97.14% and 98.67% overall accuracy with GLCM+HOG+DMLTP features on the original and augmented datasets, respectively. The DCNN–BiLSTM provides improved precision of 98.67% on the augmented dataset compared to 93.02% on the original dataset for GLCM+HOG+DMLTP features.

In the future, the effectiveness of the DR system can be enhanced by improving the model's explainability through a reliable interpretation model. The DL model's performance can be improved by optimizing the hyperparameters of the DCNN–BiLSTM. In the future, the system can be deployed on standalone portable devices and clinically validated with the help of medical practitioners for real-time datasets.

#### DATA AVAILABILITY STATEMENT

The open-access Kaggle Diabetic Retinopathy dataset (APTOS 2019 version) can be found at [https://www.kaggle.com/datasets/sovitrat/diabetic-retinopathy-224x224-2019-data?select=colored\\_image](https://www.kaggle.com/datasets/sovitrat/diabetic-retinopathy-224x224-2019-data?select=colored_image).

#### REFERENCES

- [1] Fung, T.H.M., Patel, B., Wilmot, E.G., Amoaku, W.M.K. (2022). Diabetic retinopathy for the non-ophthalmologist. *Clinical Medicine*, 22(2): 112-116. <https://doi.org/10.7861/clinmed.2021-0792>
- [2] Tan, T.E., Wong, T.Y. (2023). Diabetic retinopathy: Looking forward to 2030. *Frontiers in Endocrinology*, 13: 1077669. <https://doi.org/10.3389/fendo.2022.1077669>
- [3] Ansari, P., Tabasumma, N., Snigdha, N.N., et al. (2022). Diabetic retinopathy: An overview on mechanisms, pathophysiology and pharmacotherapy. *Diabetology*, 3(1): 159-175. <https://doi.org/10.3390/diabetology3010011>
- [4] Vujosevic, S., Aldington, S.J., Silva, P., et al. (2020). Screening for diabetic retinopathy: New perspectives and challenges. *The Lancet Diabetes & Endocrinology*, 8(4): 337-347. [https://doi.org/10.1016/S2213-8587\(19\)30411-5](https://doi.org/10.1016/S2213-8587(19)30411-5)
- [5] Yang, Z., Tan, T.E., Shao, Y., Wong, T.Y., Li, X. (2022). Classification of diabetic retinopathy: Past, present and future. *Frontiers in Endocrinology*, 13: 1079217. <https://doi.org/10.3389/fendo.2022.1079217>
- [6] Atwany, M.Z., Sahyoun, A.H., Yaqub, M. (2022). Deep learning techniques for diabetic retinopathy classification: A survey. *IEEE Access*, 10: 28642-28655. <https://doi.org/10.1109/ACCESS.2022.3157632>
- [7] Sebastian, A., Elharrouss, O., Al-Maadeed, S., Almaadeed, N. (2023). A survey on deep-learning-based diabetic retinopathy classification. *Diagnostics*, 13(3): 345. <https://doi.org/10.3390/diagnostics13030345>
- [8] Gautam, A., Shanker, R. (2025). Diabetic retinopathy detection from fundus images: A wide survey from grading to segmentation of lesions. *Computers in Biology and Medicine*, 196: 110715. <https://doi.org/10.1016/j.compbiomed.2025.110715>
- [9] Saadna, Y., Mezzoudj, S., Khelifa, M. (2025). Efficient Transformer architectures for diabetic retinopathy classification from fundus images: DR-MobileViT, DR-EfficientFormer, and DR-SwinTiny. *Informatica*, 49(29): 1-16. <https://doi.org/10.31449/inf.v49i29.8695>
- [10] Tsiknakis, N., Theodoropoulos, D., Manikis, G., et al. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine*, 135: 104599. <https://doi.org/10.1016/j.compbiomed.2021.104599>
- [11] Bhulakshmi, D., Rajput, D.S. (2024). A systematic review on diabetic retinopathy detection and classification based on deep learning techniques using fundus images. *PeerJ Computer Science*, 10: e1947. <https://doi.org/10.7717/peerj-cs.1947>
- [12] 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>
- [13] Ashraf, M.H., Mehmood, M.N., Ahmed, M., et al. (2025). HIRD-Net: An explainable CNN-based framework with attention mechanism for diabetic retinopathy diagnosis using CLAHE-D-DoG enhanced fundus images. *Life*, 15(9): 1411. <https://doi.org/10.3390/life15091411>
- [14] Lin, Z., He, Z., Wang, X., et al. (2025). Cross-scale fuzzy holistic attention network for diabetic retinopathy grading from fundus images. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(3): 2164-2178.

- <https://doi.org/10.1109/TETCI.2025.3543361>
- [15] Washburn, S.P.S., Mahendran, G. (2026). Investigations of diabetic retinopathy using deep learning techniques. *PeerJ Computer Science*, 12: e3605. <https://doi.org/10.7717/peerj-cs.3605>
- [16] Mohan, N.J., Murugan, R., Goel, T., Roy, P. (2026). A real-time screening system for diabetic retinopathy grading using a novel multi-scale feature extraction through retinal fundus images. *Biomedical Signal Processing and Control*, 112: 108548. <https://doi.org/10.1016/j.bspc.2025.108548>
- [17] Agarwal, A.K., Hamid, A.B.B.A., Ather, D., Tiwari, R.G., De, I., Krishna, K.R. (2026). Lesion-aware ordinal transformer for diabetic retinopathy classification from fundus images. *Biomedical and Pharmacology Journal*, 19(1): 233-245. <https://doi.org/10.13005/bpj/3349>
- [18] Abraham, S.E., Kovoov, B.C. (2026). MHSA-enhanced CNNs with TOPSIS-driven ensemble learning for automated diabetic retinopathy grading. *Biomedical Signal Processing and Control*, 112: 108614. <https://doi.org/10.1016/j.bspc.2025.108614>
- [19] Khandakar, M.S., Samsuddoha, M., Jahan, S., Faisal, R.H. (2026). An automated analytics approach for diabetic retinopathy detection with ensemble deep learning models in healthcare. *Healthcare Analytics*, 9: 100449. <https://doi.org/10.1016/j.health.2026.100449>
- [20] Abbasi, R., Amin, F., Alabrah, A., et al. (2025). Diabetic retinopathy detection using adaptive deep convolutional neural networks on fundus images. *Scientific Reports*, 15(1): 24647. <https://doi.org/10.1038/s41598-025-09394-0>
- [21] Men, Y., Fhima, J., Celi, L.A., Ribeiro, L.Z., Nakayama, L.F., Behar, J.A. (2025). Deep learning generalization for diabetic retinopathy staging from fundus images. *Physiological Measurement*, 46(1): 015001. <https://doi.org/10.1088/1361-6579/ada86a>
- [22] Sushith, M., Sathiya, A., Kalaipoonguzhali, V., Sathya, V. (2025). A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images. *Scientific Reports*, 15(1): 15166. <https://doi.org/10.1038/s41598-025-99309-w>
- [23] Delavari, P., Ozturan, G., Navajas, E.V., Yilmaz, O., Oruc, I. (2025). AI-assisted identification of sex-specific patterns in diabetic retinopathy using retinal fundus images. *PLOS ONE*, 20(8): e0327305. <https://doi.org/10.1371/journal.pone.0327305>
- [24] Naveen, K.V., Anoop, B.N., Siju, K.S., Kar, M.K., Venugopal, V. (2025). EffNet-SVM: A hybrid model for diabetic retinopathy classification using retinal fundus images. *IEEE Access*, 13: 79793-79804. <https://doi.org/10.1109/ACCESS.2025.3566073>
- [25] Rath, S.R. (2019). Diabetic retinopathy 224x224 (2019 data). Kaggle Dataset.
- [26] Karthik, M., Dane, S. (2019). APTOS 2019 blindness detection. Kaggle Competition Dataset.
- [27] Ahmed, F., Uddin, M.J. (2025). Ocuvit: A vision transformer-based approach for automated diabetic retinopathy and AMD classification. *Journal of Imaging Informatics in Medicine*, 39: 2511-2521. <https://doi.org/10.1007/s10278-025-01676-3>
- [28] Vallukappully, S., van der Linde, I., Chakraborty, A. (2025). Early detection and classification of diabetic retinopathy by transfer learning of NASNet-large and ResNet-50 convolutional neural networks. *Informatics in Medicine Unlocked*, 58: 101688. <https://doi.org/10.1016/j.imu.2025.101688>