**Research Article**

Yaakoub Boualleg, Kheir Eddine Daouadi, Oussama Guehairia, Chawki Djeddi, Abbas Cheddad*, Imran Siddiqi, and Brahim Bouderah

# Deep multi-view feature fusion with data augmentation for improved diabetic retinopathy classification

https://doi.org/10.1515/jisys-2024-0374
received August 11, 2024; accepted November 17, 2024

**Abstract:** Diabetic retinopathy (DR) is a leading cause of blindness worldwide, necessitating early detection to prevent severe visual impairment. Despite numerous proposed classification techniques, challenges persist due to the high parameter count of deep learning algorithms, imbalanced datasets, and limited performance. This study introduces a novel framework for DR classification that leverages multi-view deep features, multilinear whitened principal component analysis, tensor exponential discriminant analysis, synthetic minority oversampling technique, and deep random forest. We evaluated this architecture using the APTOS blindness dataset under a standard protocol. The results demonstrate that our architecture significantly improves classification accuracy, surpassing existing methods. Our contributions highlight a promising approach for enhancing DR classification performance.

**Keywords:** diabetic retinopathy classification, deep random forest, multi-view deep feature, multilinear whitened principal component analysis, synthetic minority oversampling technique

**MSC 2020:** 68T01, 68T45

## 1 Introduction

Diabetic retinopathy (DR) is a significant cause of blindness worldwide. It arises from the impact of diabetes on the retinal blood vessels. Early detection of DR is crucial to prevent or slow the progression of severe visual impairment. Annual screening for DR is recommended for individuals with diabetes. However, the diagnosis process is challenging due to limited early-stage symptoms. Consequently, manual DR diagnosing is time-consuming and prone to variability between clinicians. As a result, there has been a growing interest in automated DR classification systems that promise faster and more accurate results.

*** Corresponding author: Abbas Cheddad,** Department of Computer Science, Blekinge Institute of Technology, Karlskrona, SE-371 79, Sweden, e-mail: abbas.cheddad@bth.se
**Yaakoub Boualleg:** Laboratory of Vision and Artificial Intelligence (LAVIA), Echahid Cheikh Larbi Tebessi University, Tebessa, 12002, Algeria, e-mail: yaakoub.boualleg@univ-tebessa.dz
**Kheir Eddine Daouadi:** Laboratory of Vision and Artificial Intelligence (LAVIA), Echahid Cheikh Larbi Tebessi University, Tebessa, 12002, Algeria, e-mail: kheireddine.daouadi@univ-tebessa.dz
**Oussama Guehairia:** Laboratory of LESIA, Mohamed Khider University of Biskra, Biskra, 07000, Algeria, e-mail: oussama.guehairia@univ-biskra.dz
**Chawki Djeddi:** Laboratory of Vision and Artificial Intelligence (LAVIA), Echahid Cheikh Larbi Tebessi University, Tebessa, 12002, Algeria, e-mail: c.djeddi@univ-tebessa.dz
**Imran Siddiqi:** Xynoptik Pty Limited, Melbourne, SA 5081, Australia, e-mail: imran.siddiqi@xynoptik.com.au
**Brahim Bouderah:** Department of Computer Science, University of Abdelhamid Ibn Badis, Mostaganem, 27000, Algeria, e-mail: brahim.bouderah@univ-mosta.dz

In recent years, several studies have addressed the challenges of automated DR classification through innovative techniques, such as deep learning and feature fusion. For example, multi-modal approaches are discussed in the study of Madarapu et al. [1], which combine multiple features, providing a broader understanding of DR detection. Furthermore, Wong et al. [2] highlighted advancements in transfer learning for medical image classification, offering insights into how pre-trained models can be fine-tuned for specialized tasks. Another key area of focus has been the challenge of class imbalance in DR datasets, where severe cases of the disease are often underrepresented. Techniques such as ensemble learning and data augmentation have been integrated into DR classification pipelines to mitigate this issue, thereby improving model generalization and robustness [3].

In this study, our proposition builds on these advancements by integrating multi-view feature fusion, data augmentation, and ensemble learning into a unified framework. Our proposed classification framework consists of five main steps. The first step aims to extract deep multi-view features from different pre-trained models. The second step relies on feature dimensionality reduction using multi-linear whitened principal component (MWPCA). This technique reduces the dimensionality of the extracted features while retaining the most significant information. To further enhance the performance of the classification model, we introduce feature fusion through tensor exponential discriminant analysis (TEDA). TEDA leverages the multi-modal characteristics of the reduced-dimensional features to combine the strengths of discriminant analysis and tensor analysis. In the next step, we target the class imbalance that is often overlooked by most of the related studies. More specifically, we employ the synthetic minority oversampling technique (SMOTE) to prevent the classification model from becoming biased toward the majority class. Finally, we use deep random forests (DRF) to overcome the challenges of limited labelled datasets, an ensemble learning technique that combines the strengths of random forests and deep learning. The research questions investigated in this study are as follows:

- How can we balance between low-dimensionality and high discriminative power of retinal image representation to enhance DR classification efficacy and efficiency?
- To what extent does data augmentation contribute to the success of DR classification models?
- Can DRF improve the accuracy in DR classification?

To answer these questions, we compare our proposed framework with most recent related works on APTOS 2019 retinal image dataset. Extensive experiments showed that (1) incorporating multi-view deep feature fusion is effective for DR classification, (2) using the SMOTE technique enhances accuracy while reducing bias toward the majority class, (3) employing DRF improves the accuracy of DR classification, and (4) our proposed method surpasses baseline methods and recent related works in terms of DR classification accuracy. The primary contribution of this study lies in achieving promising accuracy results in the task of DR classification.

The remainder of this manuscript is organized as follows: Section 2 provides a comprehensive review of state-of-the-art methodologies. Section 3 delineates our proposed novel approach. Section 4 presents a detailed analysis of our experimental results. In Section 5, we discuss the key contributions and implications of our findings. Finally, Section 6 offers concluding remarks and suggestions for future research directions.

# 2 Related works

Over recent years, various methods have been proposed to classify DR into two, three, four, or five severity levels. This section provides a comprehensive overview of the most relevant methods, categorized into machine learning, deep learning, and ensemble learning methods.

## 2.1 Traditional machine learning methods

Machine learning algorithms have been widely employed to assist in the detection, diagnosis, and risk stratification of DR. These algorithms can analyze retinal images, extract specific features, or lesions associated

with DR, and classify the condition's severity. By automating these tasks, machine learning models enhance the screening process, mitigate human error, and increase overall efficiency.

A variety of machine learning methods have been explored in the literature to address this problem. For instance, Biran et al. [4] utilized the Star database, comprising 33 retinal images, to classify the images into three categories: normal, non-proliferative diabetic retinopathy, and proliferative diabetic retinopathy. They employed Gabor filters for the detection of exudates and circular Hough transform for hemorrhage region identification, followed by classification using a support vector machine (SVM) with a non-linear kernel function. Similarly, Roy et al. [5] used fundus images from the DIARET, DRIVE, and MESSIDOR datasets, employing Fuzzy C-means and convex hull techniques for feature extraction, with SVM as the classifier.

In another study [6], microaneurysm (MA) detectability was analyzed using small 25 × 25 pixel patches extracted from DIARETDB1 fundus images. The raw pixel intensities of these patches were used as direct inputs to various classifiers, including RF, neural networks (NNs), and SVM. The SVM model, employing a radial basis function kernel, achieved a high accuracy of 98.5% in terms of the receiver-operating characteristic and 92.6% for the area under the curve. Furthermore, in the study by Emon et al. [7], a comprehensive set of machine learning classifiers including naïve Bayes, sequential minimal optimization, logistic regression, stochastic gradient descent, bagging classifier, J48 classifier, decision tree classifier, and RF was applied to the Diabetes Retinopathy Debrecen dataset from the University of California, Irvine (UCI). The retinal image features are extracted using principal component analysis. Moreover, Safitri and Juniati [8] proposed a method to identify DR using fractal analysis and the k-nearest neighbor (kNN) algorithm, where the authors segmented images, calculated the fractal dimension of the segmented regions, and classified the images using kNN. This method was evaluated using cross-validation, with the optimal value of $k$ determined to achieve the highest accuracy.

Despite the effectiveness of machine learning in DR classification, several challenges persist. These include the need for extensive feature engineering, the potential for model overfitting, and the difficulty in capturing complex patterns inherent in retinal images. As a consequence, deep learning has emerged as a compelling alternative, offering superior performance in handling intricate tasks and learning hierarchical representations directly from raw data. Unlike traditional machine learning approaches, deep learning models eliminate the need for manual feature extraction by automatically learning relevant features, which reduces human effort and mitigates potential biases. Additionally, advancements in hardware and the availability of large-scale datasets have further facilitated the adoption of deep learning models, allowing for end-to-end training and deployment at scale. While deep learning has demonstrated remarkable success in fields such as image recognition, speech processing, and natural language understanding, traditional machine learning remains valuable in scenarios with limited data availability or where model interpretability is a critical requirement.

## 2.2 Deep learning-based methods

Deep learning has dramatically transformed computer vision. Its capacity to autonomously learn intricate features from data, adapt to diverse tasks, scale effectively, and generalize across different domains has led to the development of highly accurate, efficient, and adaptable models. These models have set new benchmarks in the medical image analysis field, offering substantial improvements over traditional methods.

The classification of DR using deep learning approaches can be broadly categorized into two primary strategies: learning from scratch and transfer learning. Transfer learning, in turn, can be implemented in two distinct modes: utilizing a pre-trained model as a feature extractor and fine-tuning a pre-trained model on the specific dataset of interest. The following sections delve into the relevant literature within each of these categories.

### 2.2.1 Learning from scratch

Learning from scratch involves designing and training a deep neural network (DNN) model from an initial state with randomly initialized weights. The model is trained on the target task, optimizing it to extract

relevant features directly from the input data without any prior knowledge or guidance. Although this approach has the potential to achieve high performance, it requires a large amount of annotated data and extensive computational resources.

Among the well-known studies in this category, Riaz et al. [9] developed a densely connected convolutional neural network (CNN) and trained it from scratch. They demonstrated that the dense connections between convolutional layers, coupled with a carefully chosen growth rate, significantly enhanced the model's performance. Experimental evaluations on the Messidor-2 and EyePACS datasets yielded promising results, showcasing the model's effectiveness in DR classification. In another study [10], a CNN-based architecture named DMENet was introduced for the detection and grading of diabetic macular edema (DME). This method utilizes a two-stage pipeline: the first stage identifies the presence of DME, while the second stage grades the severity. The use of a hierarchical ensemble of CNNs in both stages resulted in a robust and accurate classification system.

Sunkari et al. [11] and Raiaan et al. [12] proposed new, lightweight ResNet variants tailored for small datasets. These architectures exhibit superior generalization capabilities and reduced learning curve fluctuations compared to the traditional ResNet model. For the same reason, Luo et al. [13] enhanced the Inception architecture by incorporating long-range global dependency units, which link feature maps from various convolutional layers, preventing overfitting and improving performance on small datasets.

Further advancements in learning from scratch have been proposed by Sunkari et al. [11] and Raiaan et al. [12], who introduced lightweight ResNet variants specifically tailored for small datasets. These architectures demonstrated superior generalization capabilities and reduced fluctuations in the learning curve compared to traditional ResNet models. To address the challenges of overfitting and limited performance on small datasets, Luo et al. [13] enhanced the Inception architecture by integrating long-range global dependency units. These units effectively link feature maps from various convolutional layers, thereby improving the model's robustness.

Additionally, Ashwini and Dash [14] and Kommaraju and Anbarasi [15] explored innovative approaches by combining residual blocks, entropy enhancement, and discrete wavelet transforms with CNNs. This combination allowed the models to be trained from scratch, enhancing their ability to learn complex features directly from DR datasets. A novel DR detection system proposed by Khan et al. [16] incorporated skip connections into dilated convolutional blocks, enabling hierarchical feature extraction and significantly improving DR detection accuracy.

In another significant contribution, Li et al. [17] presented a method for DR detection using a deep CNN with fractional pooling layers, which replaced traditional max-pooling layers. The features extracted by this network were subsequently classified using an SVM to categorize DR severity into five levels (0–4). The authors also developed an application, "Deep Retina," designed for use with handheld ophthalmoscopes to provide instant DR classification. Murugappan et al. [18] introduced a novel few-shot learning classification network, DRNet, specifically designed for DR detection and grading. By leveraging attention mechanisms, DRNet achieved promising results on the APTOS 2019 dataset.

Islam et al. [19] proposed a deep CNN model for the early detection of DR, focusing on the identification of MAs and accurate labeling of retinal fundus images. Their model, evaluated on the extensive Kaggle dataset, demonstrated the effectiveness of CNNs in early DR detection. Similarly, the Triple-DRNet model, introduced by Nguyen et al. [20], employed a three-stage classification pipeline to distinguish between different DR severity levels. This model achieved an accuracy of 92.08% and a quadratic weighted kappa metric of 93.62% on the APTOS 2019 Blindness Detection dataset, underscoring its robustness and clinical relevance.

### 2.2.2 Transfer learning

Transfer learning is a powerful machine learning technique that leverages the knowledge acquired by a pre-trained model, originally designed for a specific task, and uses it as a starting point for a new task. Instead of training a model from scratch, transfer learning allows transferring knowledge of a pre-trained model to a new task with relatively smaller datasets and limited computational resources. This approach can lead to

substantial time and resource savings while often delivering superior performance compared to models trained from scratch. Transfer learning has become increasingly prevalent in various fields, particularly in medical image analysis, where it is employed primarily in two strategies: fine-tuning pre-trained models and using them as feature extractors.

**Fine-tuning:** Fine-tuning is the process of taking a pre-trained model and adjusting its parameters on a new dataset or task. The pre-trained model has already learned general patterns or features from large training data. Then, fine-tuning allows the model to further optimize, adapt, and specialize to a specific task or dataset by continuing backpropagation on either all or a subset of layers.

One prominent study in this category by Jian et al. [21] employed pre-trained CNN models for classifying fundus images to detect DR severity grade. The research evaluated several well-known models, including AlexNet, VGGNet, GoogleNet, and ResNet. The findings demonstrated that fine-tuning these models on the target dataset led to enhanced performance in DR classification. Similarly, Wan et al. [22] fine-tuned the MobileNetv2 model by training only the top layers while keeping the remaining layers' weights fixed. This approach was validated on the Kaggle DR dataset, yielding promising results. Further exploration of fine-tuning is evident in the work by Patel and Chaware [23], where the pre-trained EfficientNet-B3 model was utilized for DR severity grading. The study was conducted on the publicly available Kaggle APTOS 2019 dataset, and the results highlighted the model's efficacy in handling this complex classification task.

Additional research by Sugeno et al. [24] explored the application of two DNNs, Inceptionv3 and EfficientNet, for classifying retinal fundus images into five DR severity levels. Besides fine-tuning these CNN models, the study enhanced performance through techniques such as dropout, data preprocessing, data augmentation, and learning rate adjustments. Experimental results on the Kaggle APTOS dataset indicated that the EfficientNetB0 model outperformed the others, achieving the best classification accuracy. In a related study [25], a comparative analysis of different deep learning models, including Inception-v3, GoogLeNet, AlexNet, and ResNet, was performed across three different datasets: Messidor-2, EyePACS-1, and DIARETDB0.

Fine-tuning of VGG model for DR grading is investigated in the study of Guehairia et al. [26] where various data augmentation techniques are explored to balance the classes in the training data. Bidwai et al. [27] proposed an ensemble approach using ResNet and a modified DenseNet architecture. The model is trained on the APTOS dataset (five classes) and DIARET DB1 dataset (two classes), and the reported results show that the ensemble method outperforms the transfer learning of a single pre-trained model. Employing a different strategy, Mondal et al. [28] proposed a hybrid deep learning architecture for DR detection. The study relies on a pre-trained Inception-ResNet-v2 model with an added custom CNN block on top of it. The entire network was fine-tuned for the specific task and evaluated on the ATPOS Kaggle dataset. In another known study [29], Gangwar ad Ravi modified VGG, ResNet, and DenseNet models using global average pooling and dropout layers to reduce over-fitting and reported promising results on the APTOS dataset.

**CNNs as feature extractors:** Using CNNs as feature extractors is a transfer learning method based on pre-trained CNN models. Unlike fine-tuning, the pre-trained model is used to generate a fixed set of embedding features that can be used as inputs for a separate machine-learning classifier.

Among methods in this category, Zia et al. [30] employed two pre-trained CNN models, VGG-19 and Inception-V3, to extract features from fundus images. These extracted features are then combined, and a cubic SVM is trained for classification. Likewise, Butt et al. [31] proposed a hybrid approach for the classification of fundus images using two pre-trained models, GoogleNet and ResNet-18, to extract features. Features are then merged to form a hybrid feature vector used as input to various classifiers, including naï Bayes, RM, and SVM. The proposed approach was evaluated on the publicly available Kaggle APTOS dataset for binary and multi-class classification, showing that SVM outperformed other classifiers for this task. Similarly, multiple deep learning methods are compared in the study of Khojasteh et al. [32] on two publicly available databases, DIARETDB1 and e-Ophtha. The study concluded that ResNet-50 with SVM outperformed other classifiers, including KNNs and optimum-path forest.

Taufiqurrahman et al. [33] employed the pre-trained MobileNet-v2 to carry out DR classification task on the APTOS dataset. Features extracted from the pre-trained model are fed to train an SVM classifier. Similarly, Dhir et al. [34] extracted features using Neural Architecture Search Network and projected them into a lower-dimensional space using the (t-distributed stochastic neighbor embedding) t-SNE method, while the variational

SVM is used as classifier. In a different strategy, Jabbar et al. [35] proposed a modified Xception architecture for DR severity classification. The study employed deep-layer aggregation to fuse features from different convolutional layers of four pre-trained networks (Inception V3, MobileNet, ResNet50, and Xception). These features are then fed to a multi-layer perceptron for training and classification. The proposed approach outperformed the original Xception architecture on the Kaggle APTOS 2019 dataset [36].

### 2.3 Ensemble learning

Ensemble learning is the training of multiple classifiers for a single task. Recently, ensemble learning methods based on stacking, bagging, or boosting have proved that learning multiple classifiers outperforms a single classifier in almost all machine learning tasks. Motivated by the effectiveness of ensemble learning, Dondeti et al. [37] developed an ensemble model using the Adaboost classifier. Three individual models based on Inception-v3, ResNet152, and Inception-Resnet-v2 architectures were trained on a private dataset in collaboration with the Beijing Tongren Eye Centre. The obtained results showed that the ensemble model outperformed the individual models. Likewise, Jiang et al. [38] trained an ensemble of five deep CNN models (Resnet50, Inceptionv3, Xception, Dense121, and Dense169) using the Kaggle dataset to improve the classification of different stages of DR. The final output is obtained by averaging the class prediction obtained by the CNN models.

### 2.4 Gaps and contributions

The extensive review of existing literature highlights the substantial progress made in applying DNN architectures to DR classification. Despite these advancements, a critical gap remains unaddressed: the exploration of non-NN-based deep learning techniques for this task. While the prevailing research predominantly focuses on NN models, there is a conspicuous absence of studies that investigate alternative deep learning frameworks. These NN-based approaches often rely heavily on extensive hyper-parameter tuning, which can be computationally expensive and time-consuming, while also presenting challenges related to model interpretability and generalizability.

In this study, we present a deep learning-based classification framework for DR severity grading, exploring the use of non-NN-based models. Our method is designed to address the limitations identified in existing approaches, as presented in the following sections of this article.

## 3 Proposed method

DR detection can be defined as a binary or multi-class image classification task that involves automatically labeling each retinal image by the corresponding severity stage based on its contents. Let $T$ be the training set of $N$ images, where $T = \{(x_i, y_i)\}_{i=1}^{n}$. Each image $x_i \in T$ has the corresponding label $y_i \in Y$, where $Y = \{1, 2, \ldots K\}$ and $K$ represents the total number of the DR classes. Our objective is to classify $N'$ test images from the test set $T'$ where $T' = \{x_j\}_{j=1}^{N'}$.

Our proposed method combines the strengths of deep learning for feature extraction with advanced machine learning techniques for feature processing and classification, culminating in an efficient framework that enhances diagnostic accuracy. The method consists of five key steps: feature extraction, dimensionality reduction, feature fusion, data augmentation, and feature classification. These steps are visually illustrated in Figure 1, which provides an overview of the workflow.

Deep embedding features are extracted from input retinal images using pre-trained CNN models. These features are then fed to MWPCA to extract the most informative features, reducing the dimensionality of the

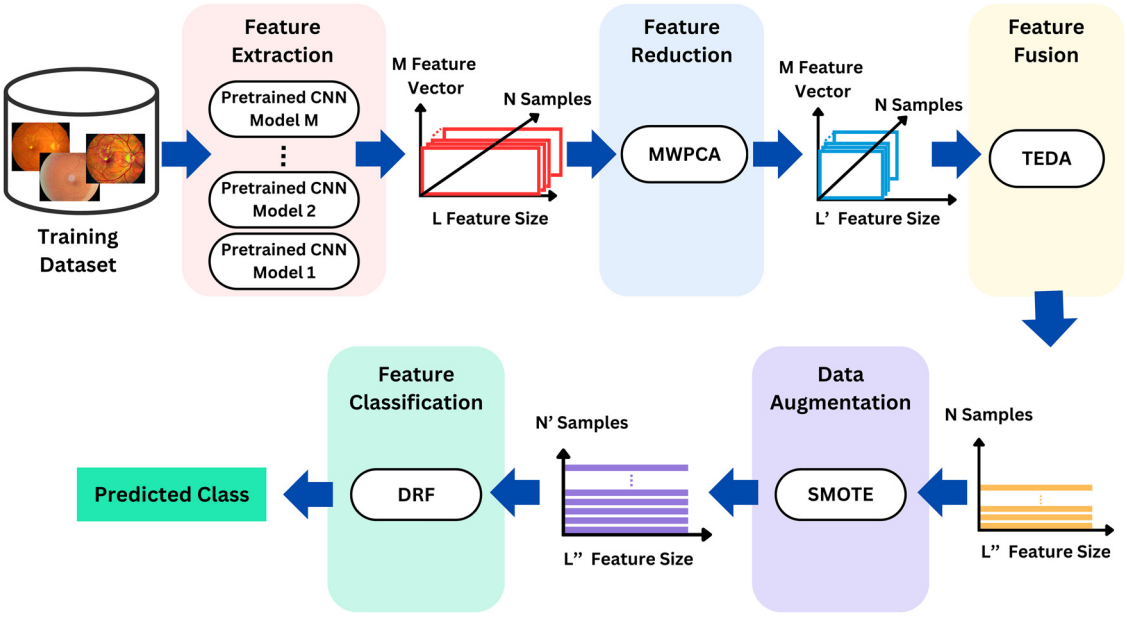


**Figure 1:** Key processing steps in the proposed DR classification framework. Source: This figure has been created by the authors.

feature space. Next, based on the TEDA method, we strive to improve feature representation discrimination by augmenting the inter-class variance and reducing the intra-class independence. The TEDA method converts the input image feature matrix to a one-dimensional feature vector for each image.

To overcome the limitations of unbalanced data for DR detection, we propose a data augmentation phase based on SMOTE, which aims to adjust the class distribution of a dataset by generating new synthetic feature points in the feature space. Finally, we train a DRF classifier for the classification phase. DRF is a non-neuronal deep architecture classifier that uses the ensemble learning method based on a decision tree as a basic computational unit. In a layer-by-layer feature processing, the information is forwarded from the first layer to the last layer, which outputs the predicted DR severity class. The details of each of these steps are presented in the subsequent sections.

## 3.1 Feature extraction

The accuracy and reliability of DR classification are largely contingent upon the ability to effectively detect and represent key retinal biomarkers. These crucial indicators, which include MAs, hemorrhages, and exudates, are instrumental in determining both the presence and progression of the disease. CNNs have become a powerful tool, offering state-of-the-art performance in generating rich, discriminative image representation. However, the limited availability of labeled retinal images for DR classification poses a significant challenge, making it difficult to train CNNs from scratch. To address this, our proposed method leverages transfer learning, utilizing pre-trained CNN models as robust feature extractors. Transfer learning capitalizes on the knowledge embedded in models pre-trained on large, diverse datasets, enabling these models to extract meaningful features even when applied to new, smaller datasets.

As shown in Figure 1, the first critical step in our proposed method is the extraction of features from the retinal images. We employ two highly effective pre-trained CNN models, ResNet-50 [39] and EfficientNet [40], as the backbone for this task. ResNet-50, with its deep architecture and residual connections, is renowned for its ability to learn hierarchical features across varying levels of abstraction, making it particularly adept at capturing complex patterns within retinal images. EfficientNet, on the other hand, utilizes a compound scaling method that balances network depth, width, and resolution, offering a highly efficient yet powerful feature extraction process.

The output of this step is a 3D tensor, where $N$ represents the number of input retinal images. Each image is represented by a matrix ($M \times L$), where $M$ is the number of the used pre-trained CNN models, and $L$ is the size of the extracted embedding features from each CNN model. By integrating features from multiple CNN models, this step not only enhances the representational power of the extracted features but also ensures that the extracted features are well suited for subsequent analysis, setting the foundation for improved DR classification.

## 3.2 Feature reduction

Following feature extraction, the dimensionality of the resulting feature space must be reduced to improve computational efficiency and focus on the most informative aspects of the data. We employ MWPCA [41] for this purpose, chosen for its effectiveness in handling multidimensional data.

MWPCA is particularly adept at minimizing noise and eliminating irrelevant variations within the feature space, which enhances the representation of retinal images. By preserving the most significant features while reducing dimensionality, MWPCA ensures that the retained features are both relevant and compact. The output of this step is a reduced 3D tensor, represented by a matrix of size ($M \times L'$) , where $L'$ is significantly smaller than the original feature size $L$ (i.e., $L' \ll L$).

## 3.3 Feature fusion

To enhance the discriminative capability of the extracted features, we employ TEDA [41]. As illustrated in Figure 1, TEDA processes the tensor output generated by MWPCA by applying projection matrices through contraction operations along each $M \times L'$ feature matrix.

This operation compresses the high-dimensional tensor into a more compact form while preserving the most critical discriminative information. The outcome of this transformation is a set of one-dimensional feature vectors for each input image $x_i$ ($i \in \{1 \dots N\}$), with a significantly reduced dimensionality $L''$, such that $L'' \ll l' \times M$. This compressed representation retains the essential characteristics of the retinal images as well as significantly reduces the computational costs, thereby facilitating more efficient and accurate DR classification.

## 3.4 Data augmentation

Addressing class imbalance is critical in machine learning, particularly in DR classification, where minority classes are often underrepresented, leading to biased decision boundaries and reduced model performance. To mitigate this issue, we employ the SMOTE [42], a widely recognized approach for rebalancing imbalanced datasets.

SMOTE generates synthetic samples for the minority class by interpolating between existing data points and their KNNs. By creating new data points along the lines connecting these neighbors, SMOTE effectively expands the decision boundary of the minority class, enabling the classifier to learn more robustly from the underrepresented instances. This method enhances the model's ability to detect subtle patterns specific to the minority class.

Also, one of the key advantages of SMOTE is its ability to improve the quality of the dataset without requiring additional manual labeling, which is often resource-intensive. The augmentation process results in an expanded feature set $N'$ with $N' \gg N$, significantly boosting the representational capacity of the minority class and, consequently, the overall classification performance.

## 3.5 Feature classification

In this step, we aim to train a DRF classifier [43]. The motivation behind using DRF is that it is a deep architecture based on a decision tree as a base unit instead of the artificial neural. That makes the RDF

classifier much easier to train than other DNNs with interpretable results. Furthermore, DRF does not require extensive hyperparameter tuning, which is one of the most challenging in the majority of our related works. Additionally, the DRF can work effectively even with small training data. On the other hand, the DRF is based on an ensemble learning technique, which is proven to outperform the use of a single classifier for various ranges of classification tasks [26,43–48].

Inspired by the deep network structure of the DNN, the RDF is based on in-model feature processing. The DRF structure uses layer-by-layer stacking to pass the information within each layer from the input layer to the last layer that outputs the final predicted class. Each level in the DRF architecture obtains the input received by its previous level and outputs the processing results to the next level.

Each level is an ensemble of forest classifiers, and each forest is an ensemble of decision trees. At a cascade level, each forest generates a class distribution vector. Then, it concatenates all the generated class distribution vectors from all the level's forests with the original input feature vector. The obtained vector is passed to the next level till the final level. The final cascade level takes as input the class distribution vectors from the previous layer. It outputs the final predicted class by averaging the received vectors and taking the class that has the highest prediction probability.

The number of cascade levels is determined automatically during the training phase based on the early stopping technique: after adding a new level, the training process will be discontinued if the classification performance does not increase sufficiently.

# 4 Experiments and results

This section details the experimental procedures carried out to evaluate our proposed method. We begin by introducing the setup, followed by a description of the used dataset. Then, we define the employed evaluation metrics and present the obtained results and the accompanying discussion.

## 4.1 Experimental setup

To ensure robust and unbiased assessment, we employed a 5-fold cross-validation evaluation protocol in our experiments. The dataset was divided into five equally sized subsets, with class distributions preserved to reflect the original dataset's balance. In each iteration, one subset served as the test set, while the remaining four subsets were used for training. This process was repeated five times, and the final performance was reported as the average across all folds.

In our study, we employed two pre-trained CNN models, namely, ResNet-50 [39] and EfficientNet [40], to extract deep retinal features. Specifically, we extracted the feature vectors from the last fully connected layer of these pre-trained models. For each input image, we obtained the feature vectors from both models, resulting in a 2D matrix of size $2{,}048 \times 2$. Subsequently, we collected the matrices of all the training samples and formed a third-order tensor. The purpose of this tensor was to estimate the projection matrices for subspace projection. To ensure convergence during the tensor projection process, we empirically set the maximum number of iterations to 16. Additionally, we set the convergence threshold for the MWPCA algorithm to 106, as employed in the study of Ouamane et al. [41].

For the classification phase, we employed the default parameters recommended in the original DRF paper [43], which have been shown to provide robust performance across a variety of tasks without extensive tuning [26,43–48]. These settings include using two RFs and two completely RFs for each level in the cascade structure, with each forest containing 500 trees. These trees are grown until they reach a pure leaf or a specified maximum depth of 100. The number of levels in the cascade is determined automatically based on performance via the early stopping mechanism; training halts when no further improvement is observed in the validation set, ensuring robust classification performance without requiring extensive hyperparameter tuning.

In our baseline comparison experiments, we used Scikit-Learn's [49,50][1] implementation of SVM and RF classifiers with their default parameter settings. For the RF model, we set the number of trees to 100, maximum depth to 30, and the minimum samples per leaf to 2. The SVM was trained using a RBF kernel with a regularization parameter ($C$) of 1 and a gamma value of 0.01. This approach was chosen to ensure a fair and consistent comparison across methods under standardized conditions, as these defaults are widely accepted as reasonable baselines for machine learning tasks. All the experiments were run on a machine with an Intel Core i7-7700 CPU (2.80 GHz) and 16GB RAM.

## 4.2 APTOS dataset

In this study, we utilized the APTOS 2019 dataset, a publicly available resource from the APTOS Blindness Detection competition on Kaggle. Curated by the Asia Pacific Tele-Ophthalmology Society, this dataset comprises 3,662 high-resolution retinal images, captured using fundus photography at the Aravind Eye Hospital in India. The dataset is instrumental in advancing research on DR detection, providing a diverse set of real-world retinal images essential for developing robust diagnostic models.

The dataset is categorized into five distinct classes: no DR (Class 0), mild DR (Class 1), moderate DR (Class 2), severe DR (Class 3), and proliferative DR (Class 4). The class distribution reveals a significant imbalance: 49% of the images are classified as no DR, while only 5% represent severe DR, and 8% are categorized as proliferative DR. Specifically, out of the 3,662 images, 1,805 are labeled as normal (no DR), and the remaining 1,857 are distributed across the four DR classes. Figure 2 provides a representative sample of images from each class, highlighting the varying severity levels of DR.

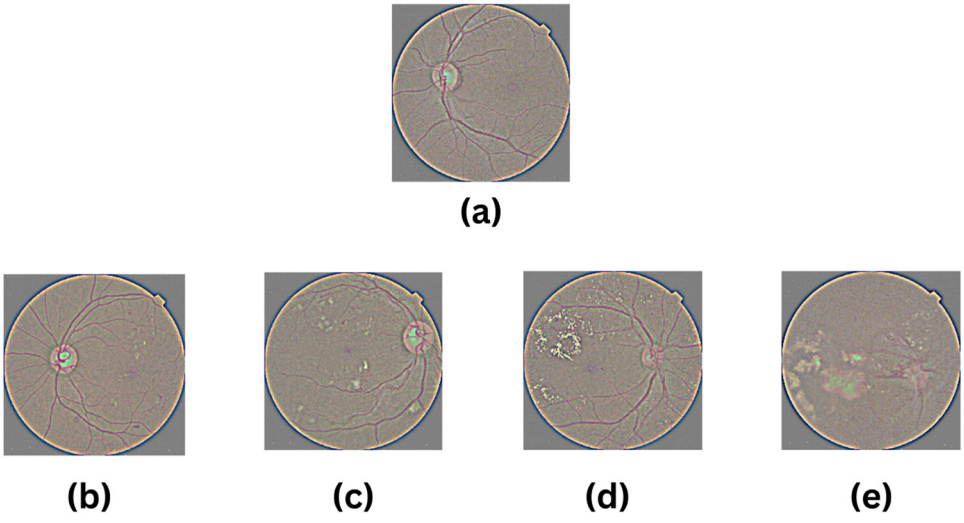


**Figure 2:** Sample images from the five Kaggle APTOS 2019 dataset classes: (a) no DR, (b) mild, (c) moderate, (d), proliferate, and (e) severe.

## 4.3 Evaluation metrics

To evaluate the classification performance of our proposition, we use the well-known evaluation metrics as follows.

1 https://scikit-learn.org/stable/.

**Accuracy** is a commonly used metric to evaluate the performance of a classification model. The accuracy measures the proportion of correct predictions out of the total number of predictions made. The accuracy is calculated as follows:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}. \tag{1}$$

**Precision** is a metric used to measure the proportion of correctly predicted positive instances (true positives) out of all instances that were predicted as positive (true positives and false positives). The precision is calculated as follows:

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}. \tag{2}$$

A high precision value indicates that the model has a low rate of false positives, meaning it is accurate when predicting positive instances. Conversely, a low precision value suggests a high rate of false positives, indicating that the model is prone to labeling negative instances as positive. Precision is particularly useful in situations where false positives are costly or have significant consequences. For example, in medical diagnostics, precision is valuable because it measures the ability of a model to accurately identify true positives (actual cases of disease) and avoid false positives (misdiagnosing healthy individuals as diseased).

**Recall:** (also called sensitivity and true-positive rate), the fraction of correctly classified positive observations over all the positive observations:

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}. \tag{3}$$

***F*1 score** is a metric commonly used to evaluate the performance of a classification model, particularly in situations where class imbalance exists. It is a single value that combines precision and recall into a balanced measure of the model's overall accuracy. The *F*1-score is calculated as the harmonic mean of precision and recall, providing a single value that takes into account both metrics. It is calculated using the following formula:

$$F1\,\text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

The *F*1 score ranges from 0 to 1, with 1 representing the best possible performance. A higher *F*1 score indicates a better balance between precision and recall, suggesting that the model is effective at both correctly identifying positive instances and avoiding false positives. Conversely, a lower *F*1-score suggests that the model may have issues with either precision or recall, or both.

**Weighted averaged *F*1 score** calculates the metric by assigning different weights to each class based on the number of samples in each class, making it useful to evaluate the model's performance while considering class imbalances. As an example, the weighted averaged *F*1 score can be estimated as the following equation:

$$\text{Weighted averaged } F1\,\text{score} = \sum_{(i=1)}^{N} (Ni/N)^{*}F1\,\text{score}_i, \tag{5}$$

where $N$ is the total number of samples, $N_i$ is the number of samples in class$_i$, and $F1$ score$_i$ is the $F1$ score of class$_i$.

**Macro-averaged *F*1 score** calculates the metric for each class independently and then computes the average of these individual class metrics, making it useful when assessing the model's performance without considering class imbalances. As an example, the macro-averaged *F*1 score can be estimated as the following equation:

$$\text{Macro averaged } F1 \text{ score} = 1/N \sum_{(i=1)}^{N} F1\,\text{score}_i, \tag{6}$$

where $N$ is the total number of classes and $F1$-score$_i$ is the $F1$ score for class$_i$.

## 4.4 Results

We conducted a comprehensive set of experiments on the APTOS dataset, following a standardized evaluation protocol. Although various classifiers were explored, we present the results from our proposed DRF classifier, which consistently outperformed conventional machine learning classifiers.

Our first set of experiments focused on evaluating the effectiveness of both single-view and multi-view deep feature representations (Table 1). The obtained results demonstrate that the fusion of deep features from ResNet50 (M1) and EfficientNet (M2) consistently outperformed individual feature sets across most DR severity classes. These results demonstrate the advantage of using complementary information from multiple deep learning models, improving the classifier ability to distinguish fine-grained changes in retinal images.

**Table 1:** Classification performance of single and multi-view deep features ($M$ = macro-averaged, $W$ = weighted averaged)

| Metric | Model | 0 | 1 | 2 | 3 | 4 | M | W |
|---|---|---|---|---|---|---|---|---|
| Precision | M1 | 88.69 | 81.27 | 89.01 | 75.24 | **82.56** | 83.35 | 86.83 |
| | M2 | **92.67** | 84.02 | 88.84 | 73.59 | 80.56 | 83.94 | 88.76 |
| | M1+M2 | 92.59 | **86.11** | **93.25** | **78.44** | 82.54 | **86.59** | **90.56** |
| Recall | M1 | 91.69 | 79.73 | 84.28 | 80.31 | 78.64 | 82.93 | 86.81 |
| | M2 | 91.54 | 82.43 | 85.91 | 88.08 | 87.12 | 87.02 | 88.54 |
| | M1+M2 | **94.18** | **83.78** | **87.09** | **88.60** | **88.14** | **88.36** | **90.42** |
| *F*1-score | M1 | 90.17 | 80.49 | 86.58 | 77.69 | 80.56 | 83.10 | 86.78 |
| | M2 | 92.10 | 83.22 | 87.35 | 80.19 | 83.71 | 85.32 | 88.60 |
| | M1+M2 | **93.38** | **84.93** | **90.06** | **83.21** | **85.25** | **87.37** | **90.43** |

Bold values indicate the best-performing results.

In the second phase of our experimental analysis, we evaluated the impact of integrating MWPCA for dimensionality reduction on the performance of our proposed method. The application of MWPCA resulted in a substantial enhancement across all evaluated metrics, as detailed in Table 2. This improvement can be primarily attributed to MWPCA's ability to effectively reduce noise and eliminate irrelevant variations within the feature space. By focusing on the most informative feature, MWPCA enhances the model's ability to generalize, thereby leading to more accurate predictions.

**Table 2:** Classification performance with and without MWPCA ("+" = with, "—" = without, $M$ = macro-averaged, $W$ = weighted averaged)

| Metric | MWPCA | 0 | 1 | 2 | 3 | 4 | M | W |
|---|---|---|---|---|---|---|---|---|
| | — | 92.59 | **86.11** | 93.25 | 78.44 | 82.54 | **86.59** | 90.56 |
| **Precision** | + | **94.73** | 80.10 | **94.58** | **79.17** | **84.14** | 86.54 | **91.54** |
| | — | **94.18** | 83.78 | 87.09 | 88.60 | 88.14 | 88.36 | 90.42 |
| **Recall** | + | 93.63 | **89.19** | **89.09** | **88.60** | **88.14** | **89.73** | **91.23** |
| | — | 93.38 | **84.93** | 90.06 | 83.21 | 85.25 | 87.37 | 90.43 |
| ***F*1-score** | + | **94.18** | 84.40 | **91.75** | **83.62** | **86.09** | **88.01** | **91.32** |

Bold values indicate the best-performing results.

In the third set of experiments, we investigated the impact of incorporating TEDA as a feature fusion technique. The results, summarized in Table 3, demonstrate a significant improvement in classification accuracy across all evaluated metrics after the application of TEDA. The primary factor contributing to this enhancement is TEDA's ability to increase the discriminative power of the features, effectively increasing the separability between different classes. Additionally, TEDA plays a crucial role in dimensionality reduc-

tion, which updates the feature space by efficiently combining multiple feature vector representations. This feature reduction and fusion strategy using TEDA method provides a more robust foundation for accurate classification.

**Table 3:** Classification performance with and without TEDA ("+" = with, "−" = without, $M$ = macro-averaged, $W$ = weighted averaged)

| Metric | TEDA | 0 | 1 | 2 | 3 | 4 | M | W |
|---|---|---|---|---|---|---|---|---|
| Precision | — | **94.73** | 80.10 | 94.58 | 79.17 | 84.14 | 86.54 | 91.54 |
| | + | 94.42 | **87.53** | **94.69** | **82.87** | **88.93** | **89.69** | **92.75** |
| Recall | — | 93.63 | 89.19 | 89.09 | 88.60 | 88.14 | 89.73 | 91.23 |
| | + | **94.74** | **89.19** | **90.99** | **92.75** | **89.83** | **91.50** | **92.65** |
| *F*1-score | — | 94.18 | 84.40 | 91.75 | 83.62 | 86.09 | 88.01 | 91.32 |
| | + | **94.58** | **88.35** | **92.80** | **87.53** | **89.38** | **90.53** | **92.68** |

Bold values indicate the best-performing results.

We conducted a series of experiments to assess the effectiveness of integrating SMOTE as a data augmentation strategy. Our results, summarized in Table 4, demonstrate that SMOTE significantly enhances the model's performance across all evaluation metrics. Notably, the *F*1-score experienced marked improvements of 2.54% for class 0 (no DR), 3.26% for class 1 (mild DR), 1.48% for class 2 (moderate DR), 1.4% for class 3 (severe DR), and 0.8% for class 4 (proliferative DR). These gains underscore SMOTE's effectiveness in addressing data imbalance, leading to a more robust and generalized classifier able to better distinguish between different severity levels of DR.

**Table 4:** Classification performance with and without SMOTE ("+" = with, "−" = without, $M$ = macro-averaged, $W$ = weighted averaged)

| Metric | SMOTE | 0 | 1 | 2 | 3 | 4 | M | W |
|---|---|---|---|---|---|---|---|---|
| Precision | — | 94.42 | 87.53 | **94.69** | 82.87 | **88.93** | 89.69 | 92.75 |
| | + | **98.19** | **90.29** | 94.56 | **84.26** | 88.56 | **91.17** | **94.89** |
| Recall | — | 94.74 | 89.19 | 90.99 | 92.75 | 89.83 | 91.50 | 92.65 |
| | + | **96.07** | **92.97** | **93.99** | **94.30** | **91.86** | **93.84** | **94.76** |
| *F*1-score | — | 94.58 | 88.35 | 92.80 | 87.53 | 89.38 | 90.53 | 92.68 |
| | + | **97.12** | **91.61** | **94.28** | **89.00** | **90.18** | **92.44** | **94.80** |

Bold values indicate the best-performing results.

Next, we conducted a performance comparison between our proposed DRF classifier and traditional classifiers such as SVM and RF. The experimental results, summarized in Table 5, clearly demonstrate the superiority of the DRF classifier. The DRF achieved notable improvements in classification performance, with an increase in the *F*1 score ranging from 2.13 to 3.23% for the macro-averaged metrics, and from 1.38 to 2.2% for the weighted average metrics. This enhanced performance can be attributed to DRF's ability to effectively manage small datasets while leveraging the complementary strengths of ensemble learning and deep learning. Specifically, DRF combines the robustness and generalization capabilities of RF with the layer-by-layer in-model feature processing power of deep learning networks, resulting in a more resilient and accurate classifier. For a more comprehensive understanding, the overall confusion matrix of our DRF-based classification approach is illustrated in Figure 3. This matrix provides a detailed view of the classifier performance across different DR severity classes, further validating the effectiveness of the proposed method.

Finally, we conducted a comparative analysis to evaluate the effectiveness of our proposed method against recent related works. Specifically, we benchmarked our approach against traditional machine learning, deep learning, and ensemble learning-based methods. The results, summarized in Table 6, demonstrate that our

**Table 5:** Comparison of classification performance of different classifiers ($M$ = macro-averaged, $W$ = weighted averaged)

| Metric | Classifier | 0 | 1 | 2 | 3 | 4 | M | W |
|---|---|---|---|---|---|---|---|---|
| Precision | RF | 96.67 | 87.86 | 94.51 | 80.19 | 86.60 | 89.17 | 93.51 |
| | SVM | 96.43 | 85.93 | 94.50 | 78.87 | 83.87 | 87.92 | 92.91 |
| | DRF | **98.19** | **90.29** | **94.56** | **84.26** | **88.56** | **91.17** | **94.89** |
| Recall | RF | 94.96 | 91.89 | 93.09 | 88.08 | 89.83 | 91.57 | 93.36 |
| | SVM | 94.40 | 90.81 | 92.79 | 87.05 | 88.14 | 90.64 | 92.71 |
| | DRF | **96.07** | **92.97** | **93.99** | **94.30** | **91.86** | **93.84** | **94.76** |
| *F*1-score | RF | 95.81 | 89.83 | 93.80 | 83.95 | 88.19 | 90.31 | 93.42 |
| | SVM | 95.41 | 88.30 | 93.64 | 82.76 | 85.95 | 89.21 | 92.78 |
| | DRF | **97.12** | **91.61** | **94.28** | **89.00** | **90.18** | **92.44** | **94.80** |

Bold values indicate the best-performing results.

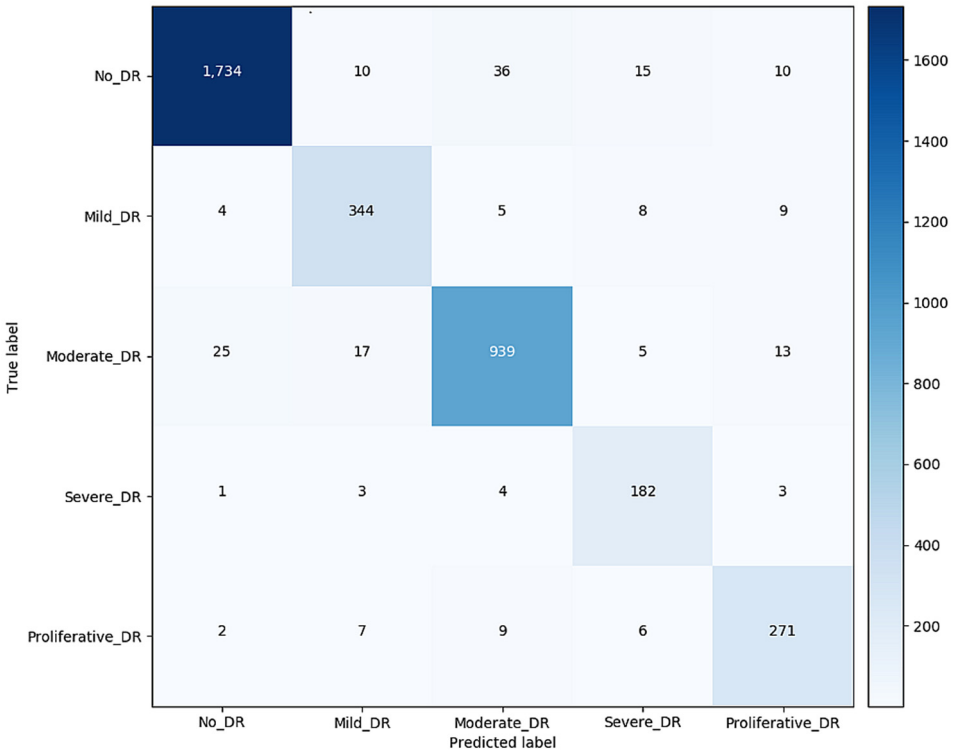


**Figure 3:** Confusion matrix of the proposed method. Source: This figure has been created by the authors.

method consistently outperforms existing approaches across multiple metrics. One of the significant advantages of our proposed method is its reduced dependency on large volumes of labeled training data, a common limitation in many DL-based methods. While most contemporary DL approaches rely heavily on NNs, our method leverages an ensemble of RF. This not only simplifies the training process but also enhances the model's robustness, avoiding the complexities associated with DNN architectures, and benefits of the advantages of ensembled learning technique.

Overall, as illustrated in Table 6, our method achieved the highest accuracy, weighted-averaged *F*1, and macro-averaged *F*1 scores of 94.8, 94.8, and 92.4%, respectively. Notably, the performance gains range from 2.9 to 13.6% in accuracy, 2.8 to 13.4% in weighted-averaged *F*1 score, and 5.2 to 17.9% in macro-averaged *F*1 score. To further validate the effectiveness of our proposition, we performed a paired *t*-test to compare the performance of our proposed method against the related works methods across all train/test folds. Experimental

**Table 6:** Performance comparison of existing DR-related works

| Work | 0 | 1 | 2 | 3 | 4 | *M* | *W* | *A* |
|---|---|---|---|---|---|---|---|---|
| [51]* | 95.6 | 85.4 | 92.7 | 77.5 | 85.2 | 87.3 | 92.0 | 91.8 |
| [22]* | 94.5 | 84.1 | 91.6 | 76.0 | 83.2 | 85.9 | 90.8 | 90.7 |
| [31]* | 94.0 | 82.7 | 91.0 | 73.5 | 83.6 | 84.9 | 90.1 | 90.0 |
| [30]* | 94.8 | 81.5 | 91.8 | 76.2 | 82.6 | 85.4 | 90.7 | 90.6 |
| [32]* | 93.1 | 78.2 | 91.5 | 71.8 | 80.1 | 83.0 | 89.0 | 88.8 |
| [9]* | 89.1 | 77.0 | 85.5 | 74.1 | 80.1 | 81.2 | 85.4 | 85.4 |
| [23]* | 88.5 | 76.0 | 84.5 | 70.6 | 78.0 | 79.5 | 84.4 | 84.3 |
| [33]* | 87.8 | 73.7 | 83.5 | 67.0 | 73.6 | 77.1 | 82.9 | 82.8 |
| [7]* | 86.3 | 72.0 | 82.3 | 61.1 | 72.5 | 74.8 | 81.3 | 81.1 |
| Ours | **97.1** | **91.6** | **94.3** | **89.0** | **90.2** | **92.4** | **94.8** | **94.8** |

*A* = accuracy, *M* = macro-averaged, *W* = weighted averaged, * indicates a statistically significant difference between our proposition and that related work (*t*-Student test, $p < 0.05$).
Bold values indicate the best-performing results.

results demonstrated that the differences were significant at ($p < 0.05$), indicating that our approach consistently outperforms existing methods across different evaluation metrics, proving its effectiveness in DR classification.

# 5 Discussion

Previous DL-based approaches for DR classification have shown remarkable success, but their reliance on large, well-labeled datasets often results in high computational costs. This dependency limits their applicability, especially where labeled data is scarce. In contrast, the method proposed in this study achieves better performance without the need for extensive labeled datasets. This flexibility makes our approach viable for both small-scale and large-scale applications, offering a more cost-effective alternative to conventional DL models.

Moreover, many current machine learning-based classification techniques are dependent on handcrafted features, which are prone to issues such as data sparsity and the curse of dimensionality. In contrast, our approach can automatically learn features from the input retinal images, effectively mitigating these challenges.

To further substantiate the effectiveness of our approach, we conducted a comprehensive comparative analysis against the relevant related works. The results consistently demonstrated that our proposition outperforms existing methods, achieving an accuracy rate exceeding 94%. This finding underscores the potential of our framework as a reliable and efficient tool for early DR detection.

While the experimental results demonstrate the high performance of the proposed framework in DR classification, it is important to consider the computational costs for practical deployment in clinical settings. The use of dimensionality reduction through MWPCA and feature fusion via TEDA ensures that the computational cost is minimized. These techniques help reduce the feature space, resulting in faster training times and lower memory usage. Additionally, the DRF model training ability without back-propagation and with fewer parameters reduces the resource demands typically associated with deep learning models. Our experiments show that the framework can be efficiently run on standard hardware, making it suitable for real-time applications in healthcare.

Although this study focuses on DR classification, the modular design of our classification framework makes it adaptable to other medical image classification tasks. The multi-view deep feature fusion approach can be applied to conditions such as skin cancer or lung disease, where multi-modal imaging data are

available. Additionally, the use of SMOTE to address data imbalance is highly relevant in many medical contexts, where certain classes (e.g., rare diseases) are underrepresented in the dataset.

## 6 Conclusion

DR represents a significant and progressive ocular complication of diabetes, characterized by damage to the retinal blood vessels, which can culminate in severe vision impairment or blindness if not promptly managed. Early detection and intervention are paramount in mitigating the progression of DR, thereby preserving vision. This study introduced a deep learning-based classification framework to facilitate the early diagnosis of DR, focusing on enhancing diagnostic accuracy and efficiency.

The proposed approach consists of five key phases: feature extraction, dimensionality reduction, feature fusion, data augmentation, and feature classification. Emphasis was placed on the critical role of data quality and sophisticated feature processing techniques in optimizing classification performances. The integration of effective algorithms for data augmentation and feature extraction was highlighted in addressing the challenges posed by imbalanced training datasets and the need for discriminative features in retinal image analysis. Additionally, the deployment of an ensemble learning-based classifier significantly enhances the classification accuracy and computational efficiency.

The experimental results demonstrated the effectiveness of the proposed approach in early DR detection. Our proposed approach achieved high and robust classification performance, outperforming most relevant related works. These results affirm the potential of the developed framework as a reliable and valuable tool for clinicians and ophthalmologists, aiding in the early detection and management of DR.

In future work, two potential avenues can be explored to enhance the deep learning approach for early DR detection: (1) employing synthetic image generation techniques for data augmentation to enrich the diversity of the training datasets, thereby improving the model's adaptability to the complex characteristics of DR; and (2) fine-tuning DRF classifier parameters to further optimize performance. By pursuing these avenues, we can continue to advance the capabilities of deep learning in the early detection of DR, ultimately contributing to better patient care and the preservation of vision for those affected by diabetes.

**Funding information**: This research was made possible thanks to a grant from the General Directorate of Scientific Research and Technological Development of the Algerian Ministry of Higher Education and Scientific Research through the National Research Project (PNR) entitled: Automatic Screening of Diabetic Retinopathy based on Artificial Intelligence Techniques. In addition, to the financial support from Blekinge Institute of Technology, Karlskrona, Karlskrona, Sweden.

**Author contributions**: Yaakoub Boualleg: conceptualization, methodology design, data analysis, and contribution to writing the original draft. Kheir Eddine Daouadi: formal analysis, software development, and manuscript drafting. Oussama Guehairia: data curation, experimental investigation, and manuscript drafting. Chawki Djeddi: validation of results and critical manuscript revision. Supervision of the research process, and funding acquisition. Abbas Cheddad: supervision, manuscript review, significant manuscript editing, and funding acquisition. Imran Siddiqi: resources provision, data interpretation, and significant manuscript editing. Brahim Bouderah: overall project supervision, guidance, and final approval of the manuscript.

**Conflict of interest**: The authors declare that there are no conflicts of interest related to this work.

**Data availability statement**: This study is evaluated on the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 dataset. This dataset is publicly available from the APTOS Blindness Detection competition on: Kaggle, https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered.

# References


[1] Madarapu S, Ari S, Mahapatra KK. A deep integrative approach for diabetic retinopathy classification with synergistic channel-spatial and self-attention mechanism. Expert Syst Appl. 2024;249:123523. doi: 10.1016/j.eswa.2024.123523.

[2] Wong CYT, Liu T, Wong TL, Tong JMK, Lau HHW, Keane PA. Development and validation of an automated machine learning model for the multi-class classification of diabetic retinopathy, central retinal vein occlusion and branch retinal vein occlusion based on color fundus photographs. JFO Open Ophthalmol. 2024;7:100117. doi: 10.1016/j.jfop.2024.100117.

[3] Kannan A, Palanivel S, Karthikeyan S, Mholds V, Joseph J. Detecting diabetic retinopathy using a hybrid ensemble XL machine model with dual weighted-kernel ELM and improved mayfly optimization. Expert Syst Appl. 2024;253:124221. doi: 10.1016/j.eswa.2024.124221.

[4] Biran A, Bidari PS, Lakshminarayanan AAV, Raahemifar K. Automatic detection and classification of diabetic retinopathy using retinal fundus images. Int J Comput Inform Eng. 2016;10(7):1308–11. doi: 10.5281/zenodo.1125443.

[5] Roy A, Dutta D, Bhattacharya P, Choudhury S. Filter and fuzzy c means-based feature extraction and classification of diabetic retinopathy using support vector machines. In 2017 International Conference on Communication and Signal Processing (ICCSP), IEEE. 2017, April. p. 1844–8.

[6] Cao W, Shan J, Czarnek N, Li L. Microaneurysm detection in fundus images using small image patches and machine learning methods. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. 2017, November. p. 325–31.

[7] Emon MU, Zannat R, Khatun T, Rahman M, Keya MS. Performance analysis of diabetic retinopathy prediction using machine learning models. In 2021 6th International Conference on Inventive Computation Technologies (ICICT). IEEE. 2021, January. p. 1048–52.

[8] Safitri DW, Juniati D. Classification of diabetic retinopathy using fractal dimension analysis of eye fundus image. In International conference on mathematics: Pure, applied and computation: Empowering engineering using Mathematics, Surabaya, Indonesia. AIP Conference Proceedings. 2017, August; Vol. 1867, No. 1. AIP Publishing. doi: 10.1063/1.4994414.

[9] Riaz H, Park J, Choi H, Kim H, Kim J. Deep and densely connected networks for classification of diabetic retinopathy. Diagnostics. 2020;10(1):24. doi: 10.3390/diagnostics10010024.

[10] Singh RK, Gorantla R. DMENet: diabetic macular edema diagnosis using hierarchical ensemble of CNNs. Plos One. 2020;15(2):e0220677. doi: 10.1371/journal.pone.0220677.

[11] Sunkari S, Sangam A, Suchetha M, Rajiv R, Rajalakshmi R, Tamilselvi S. A refined ResNet18 architecture with Swish activation function for diabetic retinopathy classification. Biomed Signal Process Control. 2024;88:105630. doi: 10.1016/j.bspc.2023.105630.

[12] Raiaan MAK, Fatema K, Khan IU, Azam S, Rashid MRU, Mukta MSH, et al. A lightweight robust deep learning model gained high accuracy in classifying a wide range of diabetic retinopathy images. IEEE Access. 2023;11: 42361–88.

[13] Luo X, Wang W, Xu Y, Lai Z, Jin X, Zhang B, et al. A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. CAAI Trans Intel Tech. 2024;9(1):153–66. doi: 10.1049/cit2.12155.

[14] Ashwini K, Dash R. Grading diabetic retinopathy using multiresolution-based CNN. Biomed Signal Proces Control. 2023;86:105210. doi: 10.1016/j.bspc.2023.105210.

[15] Kommaraju R, Anbarasi MS. Diabetic retinopathy detection using convolutional neural network with residual blocks. Biomed Signal Proces Control. 2024;87:105494. doi: 10.1016/j.bspc.2023.105494.

[16] Khan U, Khan M, Elsaddik A, Gueaieb W. Ddnet: Diabetic retinopathy detection system using skip connection-based upgraded feature block. In 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA). IEEE; 2023, June. p. 1–6. doi: 10.1109/MeMeA57477.2023.10171958.

[17] Li YH, Yeh NN, Chen SJ, Chung YC. Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network. Mobile Inform Syst. 2019;2019:6142839. doi: 10.1155/2019/6142839.

[18] Murugappan M, Prakash NB, Jeya R, Mohanarathinam A, Hemalakshmi GR, Mahmud M. A novel few-shot classification framework for diabetic retinopathy detection and grading. Measurement, 2022;200:111485. doi: 10.1016/j.measurement.2022.111485.

[19] Islam SMS, Hasan MM, Abdullah S. Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images. 2018. arXiv Preprint. http://arXiv.org/abs/arXiv:1812.10595.

[20] Nguyen QH, Muthuraman R, Singh L, Sen G, Tran AC, Nguyen BP, et al. Diabetic retinopathy detection using deep learning. In Proceedings of the 4th International Conference on Machine Learning and Soft Computing. 2020, January. p. 103–7.

[21] Jian M, Chen H, Tao C, Li X, Wang G. Triple-DRNet: A triple-cascade convolution neural network for diabetic retinopathy grading using fundus images. Comput Biol Med. 2023;155:106631. doi: 10.1016/j.compbiomed.2023.106631.

[22] Wan S, Liang Y, Zhang Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. Comput Electr Eng. 2018;72:274–82. doi: 10.1016/j.compeleceng.2018.07.042.

[23] Patel R, Chaware A. Transfer learning with fine-tuned MobileNetV2 for diabetic retinopathy. In 2020 International Conference for Emerging Technology (INCET). IEEE; 2020, June. p. 1–4. doi: 10.1109/INCET49848.2020.9154014.

[24] Sugeno A, Ishikawa Y, Ohshima T, Muramatsu R. Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning. Comput Biol Med. 2021;137:104795. doi: 10.1016/j.compbiomed.2021.104795.

[25] Chen CY, Chang MC. Using deep neural networks to classify the severity of diabetic retinopathy. In 2022 IEEE International Conference on Consumer Electronics-Taiwan. IEEE; 2022, July. p. 241–2.

[26] Guehairia O, Ouamane A, Dornaika F, Taleb-Ahmed A. Deep random forest for facial age estimation based on face images. In 2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP). IEEE; May 2020. p. 305–9. doi: 10.1109/CCSSP49278.2020.9151621.
[27] Bidwai P, Gite S, Patwa K, Maheshwari K, Bais TS, Batavia K. Detection of diabetic retinopathy using deep learning. In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT). IEEE; 2023, April. p. 1–8.
[28] Mondal SS, Mandal N, Singh KK, Singh A, Izonin I. EDLDR: An ensemble deep learning technique for detection and classification of diabetic retinopathy. Diagnostics; 2022;13(1):124. doi: 10.3390/diagnostics13010124.
[29] Gangwar AK, Ravi V. Diabetic retinopathy detection using transfer learning and deep learning. In Evolution in computational intelligence: Frontiers in intelligent computing: theory and applications (FICTA 2020). Vol. 1. Singapore: Springer; 2021. p. 679–89. doi: 10.1007/978-981-15-5788-0_64.
[30] Zia F, Irum I, Qadri NN, Nam Y, Khurshid K, Ali M, A multilevel deep feature selection framework for diabetic retinopathy image classification. Comput Mater Contin. 2022;70:2261–76. doi: 10.32604/cmc.2022.017820.
[31] Butt MM, Iskandar DA, Abdelhamid SE, Latif G, Alghazo R. Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features. Diagnostics. 2022;12(7):1607. doi: 10.3390/diagnostics12071607.
[32] Khojasteh P, Júnior LAP, Carvalho T, Rezende E, Aliahmad B, Papa JP, et al. Exudate detection in fundus images using deeply-learnable features. Comput Biol Med. 2019;104:62–9. doi: 10.1016/j.compbiomed.2018.10.031.
[33] Taufiqurrahman S, Handayani A, Hermanto BR, Mengko TLER. Diabetic retinopathy classification using a hybrid and efficient MobileNetV2-SVM model. In 2020 IEEE Region 10 Conference (TENCON). IEEE; 2020, November. p. 235–40.
[34] Dhir S, Bala R, Goel N, Sharma A. Improved transfer learning approach for diabetic retinopathy screening. In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE; 2023. p. 451–6. doi: 10.1109/SPIN57001.2023.10117173.
[35] Jabbar MK, Yan J, Xu H, Ur Rehman Z, Jabbar A. Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. Brain Sci. 2022;12(5):535. doi: 10.3390/brainsci12050535.
[36] Kassani SH, Kassani PH, Khazaeinezhad R, Wesolowski MJ, Schneider KA, Deters R. Diabetic retinopathy classification using a modified xception architecture. In 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE; 2019, December. p. 1–6. doi: 10.1109/ISSPIT47144.2019.9001846.
[37] Dondeti V, Bodapati JD, Shareef SN, Naralasetti V. Deep convolution features in non-linear embedding space for fundus image classification. Revue d'Intell Artif. 2020;34(3):307–13. doi: 10.18280/ria.340308.
[38] Jiang H, Yang K, Gao M, Zhang D, Ma H, Qian W. An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2019, July. p. 2045–8. doi: 10.1109/EMBC.2019.8857160.
[39] Koonce B. ResNet 50. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization. 1st ed. Berkeley, CA: Apress; 2021. p. 63–72. doi: 10.1007/978-1-4842-6168-2.
[40] Koonce B. EfficientNet. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization. 1st ed. Berkeley, CA: Apress; 2021. p. 109–23. doi: 10.1007/978-1-4842-6168-2.
[41] Ouamane A, Chouchane A, Boutellaa E, Belahcene M, Bourennane S, Hadid A. Efficient tensor-based 2d+3d face verification. IEEE Trans Inform Forensics Security. 2017;12(11):2751–62. doi: 10.1109/TIFS.2017.2718490.
[42] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intel Res. 2002;16:321–57. doi: 10.1613/jair.953.
[43] Zhou ZH, Feng J. Deep forest. National Science Review, 2019;6(1):74–86. doi: 10.1093/nsr/nwy108.
[44] Boualleg Y, Farah M, Farah IR. Remote sensing scene classification using convolutional features and deep forest classifier. IEEE Geosci Remote Sens Lett. 2019;16(12):1944–8. doi: 10.1109/LGRS.2019.2911855.
[45] Daouadi KE, Rebaï RZ, Amous I. Optimizing semantic deep forest for tweet topic classification. Inform Syst. 2021;101:101801. doi: 10.1016/j.is.2021.101801.
[46] Guehairia O, Ouamane A, Dornaika F, Taleb-Ahmed A. Feature fusion via Deep Random Forest for facial age estimation. Neural Networks. 2020;130:238–52. doi: 10.1016/j.neunet.2020.07.006.
[47] Daouadi KE, Boualleg Y, Guehairia O. Deep random forest and AraBert for hate speech detection from arabic tweets. J Univ Comput Sci. 2023;29(11):1319–35. doi: 10.3897/jucs.112604.
[48] Daouadi KE, Rebaï RZ, Amous I. Bot detection on online social networks using deep forest. In Artificial Intelligence Methods in Intelligent Algorithms: Proceedings of 8th Computer Science On-line Conference 2019, Vol. 985. Cham: Springer; 2019. p. 307–15. doi: 10.1007/978-3-030-19810-7_30.
[49] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
[50] Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. arXiv preprint arXiv:1309.0238. 2013.
[51] Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, et al. A deep learning ensemble approach for diabetic retinopathy detection. IEEE Access. 2019;7:150530–9.