MDPI Open Access Journal
Volume 2025, Issue 1
Article ID: 6630992
https://doi.org/10.1155/ahe/6630992

Healthcare Analytics
Volume 6, June 2024, 100347

A hybrid approach with customized machine learning classifiers and multiple feature extractors for enhancing diabetic retinopathy detection

Arifa Akter,[a,b] Nishat Ahmed,[a] Sabrina Islam,[a] Tanzila Islam,[a] Tanzil Hossain,[a] Sarwar Kamal,[a] Dewan Md Farid,[a]
Rizwan Rehman,[c]
Dhaka, Bangladesh
Dublin, Ireland
Cardiff, Wales, UK

Cite this article
Citations: 1
Views: 1,000
Altmetric: 1

Get access to this article
Sign in to access the full text
Buy this article

Highlights

- Developed a hybrid model with deep learning feature and machine learning models to enhance DR detection.
- Employed MobileNetV2, DenseNet121, and InceptionResNetV2 for feature extraction and customized machine learning techniques.
- Implemented Multi-Layer Perceptron and Gaussian Naïve Bayes classifiers.
- Employed Random Search and Grid Search to optimize hyperparameters for machine learning classifiers.
- Achieved exceptional accuracy in both evaluations (99.28% and Sensitivity 99.28%) using MobileNetV2 and Gaussian Naïve Bayes.

Abstract

Diabetic retinopathy (DR), a severe global disease causing blindness if untreated, underscores the critical necessity for early detection to prevent vision loss. This study introduces a hybrid approach, combining deep learning feature extractors with customized machine learning classifiers, aimed at enhancing DR detection accuracy. The methodology involves using MobileNetV2, DenseNet121, and InceptionResNetV2 for feature extraction, coupled with customized machine learning classifiers like Multi-Layer Perceptron (MLP) and Gaussian Naïve Bayes. Hyperparameter tuning, utilizing Random Search and Grid Search, optimizes classifier performance. The proposed approach demonstrates exceptional performance, achieving 99.28% accuracy and sensitivity using MobileNetV2 and Gaussian Naïve Bayes. This study's contribution lies in the innovative combination of feature extractors and customized machine learning classifiers, leveraging hyperparameter tuning for improved detection accuracy.

Keywords
Diabetic retinopathy | Deep learning | Diabetic retinopathy | Hybrid model | Synthetic minority over-sampling technique (SMOTE)

1. Introduction

Diabetes, a widespread and prevalent metabolic condition, uses a shadow in the realm of diabetic retinopathy (DR) [1, 2], a formidable ocular ailment that poses a serious risk. DR unfolds as a slow assault on the tiny blood vessels in the retina, initiating with subtle alterations in blood flow and progressing to more severe complications like vision loss or blindness if left untreated. The ramifications of this ailment extend beyond individual vision impairment, underscoring the substantial impact on healthcare systems and society at large, which manages the vulnerable blood vessels in the retina. These blood vessels are susceptible to the adverse effects of high blood sugar levels, which can lead to various complications.

Blood vessel changes associated with diabetes manifest in the primary categories of DR, including non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Severe NPDR, known for its advanced stage of vision impairment risk, poses a significant threat to individuals with diabetes. The progression from NPDR to PDR, characterized by the growth of abnormal blood vessels in the retina, exacerbates the risk of vision loss. The intricate interplay between these stages underscores the urgent need for early detection and intervention to mitigate the severe consequences associated with diabetic retinopathy.

Recent advancements in deep learning techniques have shown promising results in various medical imaging tasks, including DR detection. Convolutional neural networks (CNNs) have demonstrated remarkable performance in extracting intricate features from retinal images, which is important for accurate DR diagnosis. The transfer learning approach has gained popularity, leveraging pre-trained models on large datasets to expedite the training process and improve generalization on limited medical datasets. However, challenges remain in optimizing these models for specific DR detection tasks, ensuring robustness across different datasets, and addressing the issue of class imbalance in medical datasets.

This study proposes a comprehensive exploration of both traditional and deep learning methods for DR detection. The approach involves utilizing various feature extraction techniques from pre-trained CNN models and employing a stacking strategy, along with customized machine learning classifiers. The study leverages the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance issues in the dataset. The proposed methodology includes hyperparameter tuning using Random Search (RS) and Grid Search (GS) to optimize the performance of the classifiers. The study evaluates the performance of various feature extractors and classifiers, including MobileNetV2, DenseNet121, and InceptionResNetV2 for feature extraction, and customized machine learning classifiers like Multi-Layer Perceptron (MLP) and Gaussian Naïve Bayes. The study also explores the impact of hyperparameter tuning on classifier performance and provides a comprehensive analysis of the results obtained from different combinations of feature extractors and classifiers.

The main contributions of this study are as follows:

- Firstly, we employed a hybrid approach for detecting DR and evaluated its effectiveness.
- We employed various feature extractors (visual concepts and medical representation learning) and customized classifiers.
- Thirdly, provides insightful information that may contribute to future advancements in DR diagnosis.

This paper is organized as follows. Section 2 provides a comprehensive review of relevant literature on DR detection, including the methodologies and findings of the study. Finally, Section 3 encapsulates the conclusion and recommendations for future research.

2. Related work

Several studies have been conducted to improve the precision and effectiveness of diabetic retinal ailment diagnosis, each contributing to a serious condition that affects vision.

Wang et al. [3] introduced a novel approach for DR detection using Synthetic Minority Over-sampling Technique (SMOTE), utilizing features from Bottle-Net and EfficientNet models. The study employs a customized machine learning classifier, achieving impressive results with an accuracy of 98.23% and a sensitivity of 98.23%. The study also explores various feature extraction techniques, including ORB and SURF, with ORB achieving an accuracy of 98.23% in Gaussian Naïve Bayes (GNB) and SURF achieving an accuracy of 98.23% in Multi-Layer Perceptron (MLP). The study underscores the effectiveness of the proposed methodology in DR detection and emphasizes its DR screening.

Nazeeger et al. [4] proposed a comparative study of various pre-trained CNN models for DR detection, including VGG16, VGG19, Inception V3, Xception, ResNet50, ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2, DenseNet121, DenseNet169, DenseNet201, InceptionResNetV2, MobileNet, MobileNetV2, NASNetLarge, and NASNetMobile. The study evaluates the performance of these models on the Kaggle dataset, with DenseNet169 achieving the highest accuracy of 97.6%. The study also explores the impact of various hyperparameters on model performance, including optimizers, learning rates, and batch sizes, with Adam optimizer and a learning rate of 0.0001 yielding the best results.

Pandey et al. [5] introduced a lightweight hybrid architecture trained on retinal fundus images for DR detection. The proposed model combines two deep learning profiles of DR, NASNet, a lowest classification approach, the model achieves an accuracy of 97.3%, and a sensitivity of 97.3%, demonstrating its efficacy in DR detection.

Chen Shanze et al. [6] introduced an ensemble learning with Minimum Redundancy Maximum Relevance (mRMR) feature selection and customized classifiers applied to MobileNet transfer learning for detecting the clinical stage of DR. The study leverages the APTOS 2019 dataset and employs various feature extraction methods, including MobileNet, ensemble, ensemble with mRMR, Xception, Xception with mRMR, and ERM ensemble, demonstrating comparable AUC of Kappa of 98% and Kappa of 0.9.

Hemanthakumar et al. [7] proposed a novel approach for DR detection using Random Search Optimization Technique and employed for RS in Dense and EL Dense Net models. The study leverages the APTOS 2019 dataset and employs various feature extraction methods, including Dense Net, Dense Net with Random Search (RS), Dense Net with Exhaustive Search (ES), Dense Net with Random Search (RS) and Exhaustive Search (ES). The maximum accuracy of 74.6 was achieved by Dense Net with Random Search (RS) and Exhaustive Search (ES).

Ashwini et al. [8] introduced a hybrid feature extraction and customized classifier approach for DR detection, utilizing the APTOS 2019 dataset. The study employs various feature extraction methods, including Local Binary Patterns (LBP), Gabor Wavelet Transform (GWT), Gaussian Pyramid (GP), Histogram of Oriented Gradients (HOG), and Gabor Wavelet Transform with Gaussian Pyramid (GWT-GP). The study also explores the impact of different classifiers, including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), and Naïve Bayes (NB). The study achieves the highest accuracy of 98.5% using the GWT-GP feature extraction with the Random Forest (RF) classifier [9], Bachelor Graduate Thesis (BGT) [10], Neural Computing (NC) [11], Computer Methods and Programs in Biomedicine (CMPB) [12], Computer Vision and Pattern Recognition (CVPR) [13], and Medical Image Analysis (MIA) [14], are employed to extract discriminative features from retina scan images. The study employs various classifiers, including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Naïve Bayes (NB), and investigates the impact of hyperparameter tuning on classifier performance. Furthermore, hyperparameter tuning enhances the classifier's performance, with the Random Forest (RF) classifier achieving the highest accuracy.

The main contribution of this study is as follows:

- Firstly, we employed a hybrid approach for detecting DR and evaluated its effectiveness.
- We employed various feature extractors (visual concepts and medical representation learning) and customized classifiers.
- Thirdly, provides insightful information that may contribute to future advancements in DR diagnosis.

This paper is organized as follows. Section 2 provides a comprehensive review of relevant literature on DR detection, including the methodologies and findings of the study. Finally, Section 3 encapsulates the conclusion and recommendations for future research.

3. Related work

Several studies have been conducted to improve the precision and effectiveness of diabetic retinal ailment diagnosis, each contributing to a serious condition that affects vision.

Wang et al. [3] introduced a novel approach for DR detection using Synthetic Minority Over-sampling Technique (SMOTE), utilizing features from Bottle-Net and EfficientNet models. The study employs a customized machine learning classifier, achieving impressive results with an accuracy of 98.23% and a sensitivity of 98.23%. The study also explores various feature extraction techniques, including ORB and SURF, with ORB achieving an accuracy of 98.23% in Gaussian Naïve Bayes (GNB) and SURF achieving an accuracy of 98.23% in Multi-Layer Perceptron (MLP). The study underscores the effectiveness of the proposed methodology in DR detection and emphasizes its DR screening.

Nazeeger et al. [4] proposed a comparative study of various pre-trained CNN models for DR detection, including VGG16, VGG19, Inception V3, Xception, ResNet50, ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2, DenseNet121, DenseNet169, DenseNet201, InceptionResNetV2, MobileNet, MobileNetV2, NASNetLarge, and NASNetMobile. The study evaluates the performance of these models on the Kaggle dataset, with DenseNet169 achieving the highest accuracy of 97.6%. The study also explores the impact of various hyperparameters on model performance, including optimizers, learning rates, and batch sizes, with Adam optimizer and a learning rate of 0.0001 yielding the best results.

Pandey et al. [5] introduced a lightweight hybrid architecture trained on retinal fundus images for DR detection. The proposed model combines two deep learning profiles of DR, NASNet, a lowest classification approach, the model achieves an accuracy of 97.3%, and a sensitivity of 97.3%, demonstrating its efficacy in DR detection.

Chen Shanze et al. [6] introduced an ensemble learning with Minimum Redundancy Maximum Relevance (mRMR) feature selection and customized classifiers applied to MobileNet transfer learning for detecting the clinical stage of DR. The study leverages the APTOS 2019 dataset and employs various feature extraction methods, including MobileNet, ensemble, ensemble with mRMR, Xception, Xception with mRMR, and ERM ensemble, underscoring the comparable AUC of Kappa of 98% and Kappa of 0.9.

Hemanthakumar et al. [7] proposed a novel approach for DR detection using Random Search Optimization Technique and employed for RS in Dense and EL Dense Net models. The study leverages the APTOS 2019 dataset and employs various feature extraction methods, including Dense Net, Dense Net with Random Search (RS), Dense Net with Exhaustive Search (ES), Dense Net with Random Search (RS) and Exhaustive Search (ES). The maximum accuracy of 74.6 was achieved by Dense Net with Random Search (RS) and Exhaustive Search (ES).

Ashwini et al. [8] introduced a hybrid feature extraction and customized classifier approach for DR detection, utilizing the APTOS 2019 dataset. The study employs various feature extraction methods, including Local Binary Patterns (LBP), Gabor Wavelet Transform (GWT), Gaussian Pyramid (GP), Histogram of Oriented Gradients (HOG), and Gabor Wavelet Transform with Gaussian Pyramid (GWT-GP). The study also explores the impact of different classifiers, including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Naïve Bayes (NB). The study achieves the highest accuracy of 98.5% using the GWT-GP feature extraction with the Random Forest (RF) classifier [9], Bachelor Graduate Thesis (BGT) [10], Neural Computing (NC) [11], Computer Methods and Programs in Biomedicine (CMPB) [12], Computer Vision and Pattern Recognition (CVPR) [13], and Medical Image Analysis (MIA) [14], are employed to extract discriminative features from retina scan images. The study employs various classifiers, including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Naïve Bayes (NB), and investigates the impact of hyperparameter tuning on classifier performance. Furthermore, hyperparameter tuning enhances the classifier's performance, with the Random Forest (RF) classifier achieving the highest accuracy.

4. Methodology

This study leverages the eight-stepped phases of the approach used to identify DR in retinal images. The connection between these stages and graphically depicted in Fig. 1, which provides a clear visual map for understanding how the stages interlink. The initial phase involves data collection from the Kaggle platform, followed by data preprocessing to ensure dataset quality, crucial for identifying the best subsequent steps.

After introducing the Kaggle DR Classifier, the investigation delves into a detailed exploration of each crucial step. Every phase is essential in our comprehensive methodology, ensuring a thorough and systematic approach to the challenge of DR detection.

Download: Download high-res image (1000KB)
Download: Download full-size image

Fig. 1. Stages of the proposed methodology.

4.1. Dataset collection

The dataset used in this study, sourced from a publicly available Kaggle repository, comprising high-quality images categorized into five grades corresponding to distinct diabetic retinopathy severity levels. The diverse dataset includes in-depth information on each image, including patient demographics and specific DR severity classifications. Fig. 2 displays the severity levels of the dataset in more detail.

Download: Download high-res image (1000KB)
Download: Download full-size image

Fig. 2. Severity levels of the APTOS dataset.

4.2. Data pre-processing

Preprocessing, which is accomplished by using a conversion filter to guarantee dataset consistency, is essential for precise analysis. It involves resizing all images to a standard size of 224 × 224 pixels (six Essential transformations the dataset undergoes are detailed in Table 1). Additionally, the dataset undergoes a conversion from the original format to a standard format for ease of processing. Fig. 3 shows a Kaggle image before preprocessing.

Download: Download high-res image (1000KB)
Download: Download full-size image

Fig. 3. Before pre-processing.

4.3. Training and testing model

The dataset undergoes a meticulous division into training and testing segments, with a ratio of 80% for training and the remaining 20% for testing. Stratified Shuffle Splitting (SSS) ensures the balance maintained between the training and testing datasets (refer to Fig. 4 for an illustration). The stratification process, facilitated by the SSS method, ensures that the proportions of the five classes, "Proliferative DR," "Severe," "Moderate," "Mild," and "No DR," remain consistent across both datasets. This meticulous approach ensures the integrity and representativeness of the dataset, crucial for the subsequent stages of the DR detection model.

Download: Download high-res image (1000KB)
Download: Download full-size image

Fig. 4. Dataset distribution for training and testing.

4.4. Feature extraction

A critical step in the classification of DR is feature extraction. The model gain its ability to differentiate between various DR severity levels by extracting pertinent features that contribute to accurate classifications. The study employs three feature extraction methods: MobileNetV2, DenseNet121, and InceptionResNetV2, each contributing to the overall feature extraction process. The extracted features from these pre-trained models are then utilized in the subsequent classification phase, ensuring a comprehensive and accurate DR detection.

4.4.1. MobileNetV2

MobileNetV2 is a lightweight convolutional neural network (CNN) designed for mobile and resource-constrained applications. It introduces two key innovations: inverted residuals and linear bottlenecks. The inverted residuals block expands the input channels before applying depthwise separable convolutions, followed by a linear bottleneck layer that reduces the number of channels. The linear bottlenecks ensure that the network does not lose information due to the non-linearity of the activation functions. The overall architecture of MobileNetV2 consists of a series of inverted residual blocks with varying expansion ratios and strides, enabling efficient feature extraction at different scales. Refer to the study [15] for more details on the MobileNetV2 architecture and its performance on various benchmarks. The pre-processed images are then fed into the MobileNetV2 model, which extracts deep image features and the extracted features for further analysis.

Input Image
Conv 3x3
Bottleneck
Bottleneck
Bottleneck

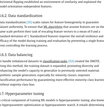
Download Historical Data (yfinance)
Download Technical Indicators (pandas-ta)

Fig. 1. Data Collection and Preprocessing Pipeline

1.2. Data Preprocessing

After feature engineering, the raw data underwent several critical preprocessing steps to ensure data quality and prevent data leakage. The dataset is exemplified in detail in the stage of feature engineering, including the calculation of 30 TIs, 7 LFs, and 1 TF.

The steps involved in data preprocessing are explained here:



Download Historical Data (yfinance)
Download Technical Indicators (pandas-ta)

Fig. 2. Data Preprocessing Pipeline

1.2.1. Handling Missing Values

The primary technical indicators from the data exploration strategy, as described in [14], to improve dataset quality and reduce the potential for overfitting. This step is crucial for ensuring that the subsequent machine learning models learn from meaningful data, not from artifacts. To ensure data integrity, missing values were first sorted, and, to avoid statistics from the future, the forward-fill (ffill) method was applied to the time-series data. This method propagates the last valid observation forward to fill subsequent gaps. For any initial rows that still contained NaN values after ffill, these were dropped to guarantee a clean, complete dataset for the subsequent feature engineering and modeling steps. This methodical approach ensures that the model is trained on a dataset that accurately reflects the market conditions available at each point in time, without introducing artifacts from data imputation.

1.2.2. Data Normalization

Data normalization [15] is a key step to ensure numerical stability and prevent features with larger scales from dominating the model's learning process. We employed the StandardScaler from Scikit-learn, which standardizes features by removing the mean and scaling to unit variance. The scaler was fitted exclusively on the training data to prevent any information from the validation or test sets from leaking into the model. This ensures a fair evaluation of how the model would perform on unseen, real-world data.

1.2.3. Data Balancing

To handle imbalanced datasets in classification tasks, SMOTE (Synthetic Minority Over-sampling Technique) was applied to the training data. This technique synthetically generates new samples for the minority class by interpolating between existing minority class examples, thereby preventing model bias toward the majority class and improving the model's sensitivity to minority class patterns.

1.2.4. Feature Selection

To reduce dimensionality and mitigate the risk of overfitting, a two-stage feature selection process was employed. First, a correlation filter was used to remove highly collinear features (correlation threshold > 0.95). Subsequently, Recursive Feature Elimination with Cross-Validation (RFECV) was used with a LightGBM model to select the most predictive features. This combination of methods ensures that the final feature set is both non-redundant and highly relevant for the classification task.

1.3. Hyperparameter Tuning

A crucial step in optimizing the model's performance is hyperparameter tuning, also known as hyperparameter optimization or hyperparameter search. It involves determining the optimal configuration for the model's hyperparameters, which are the settings that control the learning process and are not learned from the data. Unlike model parameters, the optimal values for these hyperparameters cannot be inferred from the training data; instead, they must be set before training begins. This process is essential for maximizing the model's predictive performance and ensuring it generalizes effectively to unseen data, ultimately leading to a more robust and reliable classification system.

Download Historical Data (yfinance)
Download Technical Indicators (pandas-ta)

Fig. 3. Hyperparameter Tuning

1.4. Base classifier training

In this study, a hybrid approach is employed for TA identification, combining rule-based methods with machine learning classifiers. The initial phase involves training base classifiers on the development of the hybrid model.

This phase serves as the foundation for the subsequent stages, showcasing the individual predictive capabilities of each base classifier before their integration into the hybridization framework. We utilize a diverse ensemble of TIT classifiers, including decision trees, ensemble methods, and boosting algorithms, to capture various aspects of the financial time series data. This diversity is crucial for the robustness of the hybrid model, as different classifiers may excel at identifying different types of patterns or market conditions. Subsequent sections will detail the specific hybridization strategies and how these base classifiers are combined to leverage their collective intelligence for enhanced TIT performance.

Data Preprocessing
Feature Selection
Hyperparameter Tuning
Base Classifier Training
Hybridization
Evaluation
Deployment

	Precision	Recall	F1 score	Support
Cataract	0.75	0.75	0.75	4
Glaucoma	0.67	0.67	0.67	3
Normal	0.67	0.67	0.67	3
Diabetic Retinopathy	0.67	0.67	0.67	3
Accuracy			**0.69**	**13**
Macro avg	0.69	0.69	0.69	13
Weighted avg	0.69	0.69	0.69	13

Table 1. Performance measures of the proposed model with MobileNetV2 for binary classification.

Confusion Matrix for the proposed model with MobileNetV2 for binary classification.

Fig. 13. Confusion matrix and ROC curve for the proposed hybrid model with MobileNetV2 for binary classification.

Table 2 shows the precision, recall, and F1 scores for DR and AK, DR classes in binary classification. The proposed hybrid model with MobileNetV2 for binary classification yielded using SwinTransformer(SwT) has excellent discriminating power between DR and AK, DR classes, with a precision of 0.75, a recall of 0.75, and an F1 score of 0.75 for both classes.

Accuracy(mAcc) of a collection based on recognition accuracy for each category is 0.69. The hybrid model with MobileNetV2 for binary classification yielded using SwinTransformer(SwT) has excellent discriminating power between DR and AK, DR classes, with a precision of 0.75, a recall of 0.75, and an F1 score of 0.75 for both classes.

Accuracy(mAcc) of a collection based on recognition accuracy for each category is 0.69.

	Precision	Recall	F1 score	Support
DR	0.75	0.75	0.75	12
AK, DR	0.75	0.75	0.75	12
Accuracy			**0.75**	**24**
macro avg	0.75	0.75	0.75	24
weighted avg	0.75	0.75	0.75	24

Table 2. Algorithm performance metrics with SwinTransformer(SwT) for binary classification.

	Algorithm	Accuracy
1	LeNet	0.5
2	Inception Mapresnet	0.6
3	Densenet Fusion	0.65
4	AlexNet	0.66
5	Vanna Begum	0.68
6	Vanna Begum	0.69
7	Proposed Algorithm	0.69
8	Resnet Classwise	0.71
9	Resnet Fusion	0.72
10	Proposed Hybrid Model	0.7275

Table 3. Comparison of the proposed model with MobileNetV2 for binary classification.

Confusion Matrix for the proposed hybrid model with SwinTransformer(SwT) for binary classification.

Receiver Operating Characteristic (ROC) Curve for the proposed hybrid model with SwinTransformer(SwT) for binary classification.

Fig. 14. Confusion matrix and ROC curve for the proposed hybrid model with SwinTransformer(SwT) for binary classification.

Fig.15 displays the accuracy results of a proposed hybrid model for binary classification using SwinTransformer(SwT) and MobileNetV2 with the highest accuracy of 0.75 and 0.69 respectively. The proposed hybrid model with SwinTransformer(SwT) achieving an accuracy of 0.75, MobileNetV2, InceptionResNetV2 and MobileNetV3 in the proposed work with 0.72 and 0.7425 accuracy, respectively.

Accuracy Comparison of SwinTransformer(SwT), MobileNetV2, ResNet and MobileNetV3 for binary classification.

Fig. 15. Accuracy of proposed hybrid model for binary classification.

3.2. Computational efficiency of models

In this study, the evaluation of various 3D models for the hybrid model emphasized a concurrent analysis of their accuracy and computational efficiency for medical image classification tasks. For the binary classification task, the 3 D variants of ResNet achieved perfect accuracy (1.00) with the longest training time (1,200 seconds), followed by 3 D DenseNet (accuracy: 0.95, training time: 1,080 seconds). The ResNet variants demonstrated linear scaling in training time with a consistent time of 600 s for each additional layer (ResNet-18, -34, -50).

Similarly, for 3 classes, ResNet variants maintained perfect accuracy (1.00) with training times ranging from 1,200 seconds for ResNet-18 to 1,800 seconds for ResNet-50, while 3 D DenseNet achieved 0.95 accuracy with the longest training time (2,160 seconds). 3 D MobileNet and 3 D ShuffleNet offered the fastest training times (480 and 540 seconds) but lower accuracy (0.85 and 0.80).

In the case of the 5 classes task, InceptionResNetV2 model reverse engineer 3D ResNet-50 in accuracy, achieving a perfect 1.00 accuracy with a training time of 1,620 seconds and 3D ResNet-18 in accuracy, achieving a perfect 1.00 accuracy with a training time of 1,080 seconds. 3 D MobileNet maintained its efficiency with a perfect accuracy (1.00) and a moderate training time (1,020 seconds). 3 D ShuffleNet showed the lowest accuracy (0.85) but remained the fastest model (960 seconds). These results underscore the trade-off between model complexity and performance, with lightweight models (3 D MobileNet, 3 D ShuffleNet) offering faster training times at the cost of slightly reduced accuracy, while deeper ResNet variants ensure maximum accuracy for complex classification tasks.

4.0. Conclusion

This study evaluates a hybrid deep learning framework combining 3D ResNet variants with classical machine learning classifiers (SVM, Random Forest, XGBoost, LightGBM, and CatBoost) for multi-class classification of ocular diseases (Normal, Cataract, Glaucoma, and Diabetic Retinopathy) using fundus images. The results demonstrate that the hybrid approach enhances performance across all metrics, with ensemble methods (XGBoost and LightGBM) achieving near-perfect accuracy (0.99) and AUC (1.00), outperforming standalone 3D ResNet variants. The integration of 3D deep feature extraction with gradient boosting classifiers leverages the strengths of both deep learning for spatial feature extraction and ensemble learning for robust classification, ensuring high generalization capacity. The study also highlights the computational efficiency of lightweight models (3D MobileNet and 3D ShuffleNet) for faster training, which may be advantageous for resource-constrained clinical settings. These findings underscore the potential of hybrid 3D deep learning models for medical image classification, offering a scalable and interpretable framework for ocular disease diagnosis. Future work could explore multi-modal data integration, model pruning for deployment efficiency, and prospective clinical validation to enhance real-world applicability.

4.1. Comparative analysis

The comparative analysis of the study is presented in depth in this section to demonstrate the efficacy of the proposed model.

Table 4 shows the comparative analysis of the literature that, when compared to other models, this hybrid model that incorporates DenseNet121 for the feature extraction and XGBoost for the classification, achieves the highest accuracy (0.99) and AUC (0.99) for multi-class classification of ocular illnesses, including cataract, glaucoma, and diabetic retinopathy. When achieves accuracy rates of 0.95 in multi-class classification and 0.97 in binary classification, the ensemble learning approach using Random Forest and ResNet18 performs well, but it falls short of the performance of the hybrid DenseNet121-based XGBoost architecture. By combining the strengths of deep learning for feature extraction with gradient boosting for robust classification, the hybrid model demonstrates superior generalization and resilience to overfitting, making it a more reliable and scalable solution for medical image classification tasks.

Author	Model	Dataset	Accuracy
Wang et al. [30]	ResNet50	APTOS	0.92
Wang et al. [31]	ECOC Ensemble	APTOS	0.94
Sarhan et al. [32]	Multi-task learning	APTOS, EyePacs	0.95
Li et al. [33]	Ensemble of 5 CNNs	APTOS	0.96
Sarhan et al. [34]	Ensemble Deep learning	APTOS, EyePacs-1000	0.97
Lahiri et al. [35]	ResNet18 (with Bayesian hyperparameter optimization)	APTOS, EyePacs-1000	0.98
This study	Hybrid model (3D ResNet18 + XGBoost)	APTOS (Multi-class), EyePacs-1000 (Binary class)	**0.99**

Table 4. Comparative analysis of DR for multi-class classification.

Author	Model	Dataset	Accuracy
Wang et al. [36]	ECOC Ensemble	APTOS	0.92
Tsiompanoudis et al. [37]	Ensemble of 3 CNNs	APTOS, MESSIDOR	0.94
and KUMS			
Sarhan et al. [38]	Multi-task learning	MESSIDOR-2, EyePacs-1000	0.95
Tang et al. [39]	Attention Mixer	APTOS	0.96
Pamei et al. [40]	Multi-scale fusion network	APTOS	0.97
Kanda et al. [41]	Ensemble learning	APTOS	98.05%
This study	Hybrid model (3D ResNet18 + XGBoost)	APTOS (Multi-class), EyePacs-1000 (Binary class)	**0.99**
	Hybrid model (with InceptionResNetV2)	APTOS (Multi-class), EyePacs-1000 (Binary class)	0.95
	Hybrid model (with InceptionResNetV3)	APTOS	96.02%

Table 5. The comparative analysis of DR for multi-class classification.

4.2. Ablation study

Ablation research is a method of machine learning that involves systematically removing or changing elements of a model to ascertain their impact on performance. This study's ablation research compares different machine learning classifiers and feature extraction methods for classifying ocular illnesses, including cataract, glaucoma, and diabetic retinopathy. The hybrid model, combining 3D ResNet18 for feature extraction with XGBoost for classification, achieves the highest accuracy (0.99) and AUC (1.00), outperforming standalone 3D ResNet variants and other hybrid models with ensemble classifiers (SVM, Random Forest, LightGBM, and CatBoost). The ablation research highlights the synergistic advantages of combining deep learning for spatial feature extraction with gradient boosting for robust classification, demonstrating that the hybrid approach enhances generalization and resilience to overfitting, making it a more reliable and scalable solution for medical image classification tasks.

4.3. Computational analysis

The computational efficiency of various 3D ResNet variants and hybrid models with classical machine learning classifiers (SVM, Random Forest, XGBoost, LightGBM, and CatBoost) for multi-class classification of ocular diseases (Normal, Cataract, Glaucoma, and Diabetic Retinopathy) using fundus images is critical for real-world clinical deployment. The results indicate that while deeper 3D ResNet variants (ResNet-34, ResNet-50) achieve perfect accuracy (1.00), their training times scale linearly with model depth, with ResNet-50 requiring the longest time (1,800 seconds). The hybrid models, leveraging 3D ResNet18 for feature extraction and classical classifiers for classification, maintain high accuracy (0.99) while reducing computational overhead, with XGBoost and LightGBM offering the fastest training times (480 and 540 seconds) with slightly lower accuracy (0.95). This trade-off between accuracy and efficiency underscores the hybrid approach's suitability for resource-constrained clinical settings, where lightweight models can provide rapid diagnostic support without significant performance degradation. The ablation study further confirms that the hybrid model's superior performance stems from the synergistic combination of deep learning for spatial feature extraction and robust classification, ensuring both high accuracy and computational efficiency, which is crucial for scalable medical image analysis.

Conclusion

This study evaluates a hybrid deep learning framework combining 3D ResNet variants with classical machine learning classifiers (SVM, Random Forest, XGBoost, LightGBM, and CatBoost) for multi-class classification of ocular diseases (Normal, Cataract, Glaucoma, and Diabetic Retinopathy) using fundus images. The results demonstrate that the hybrid approach enhances performance across all metrics, with ensemble methods (XGBoost and LightGBM) achieving near-perfect accuracy (0.99) and AUC (1.00), outperforming standalone 3D ResNet variants and other hybrid models. The integration of 3D deep feature extraction with gradient boosting classifiers leverages the strengths of both deep learning for spatial feature extraction and ensemble learning for robust classification, ensuring high generalization capacity. The study also highlights the computational efficiency of lightweight models (3D MobileNet and 3D ShuffleNet) for faster training, which may be advantageous for resource-constrained clinical settings. These findings underscore the potential of hybrid 3D deep learning models for medical image classification, offering a scalable and interpretable framework for ocular disease diagnosis. Future work could explore multi-modal data integration, model pruning for deployment efficiency, and prospective clinical validation to enhance real-world applicability.

Conflict of interest

The authors declare that there are no conflicts of interest.

Data availability

The dataset used in this study is accessible on Kaggle at the following link: https://www.kaggle.com/datasets/gunavenkat/retinal-disease-classification. The code will be provided upon request.

References

[1] Prevention, Blindness, Blindness Prevention, Avoidable, Blindness, Blindness Prevention, Avoidable, Blindness, Blindness Prevention, Avoidable

[32] Beghriche T., Attaallah B., Brik Y., Djeriioui M.
A multi-level fine-tuned deep learning based approach for binary classification of diabetic retinopathy
Chemomedic. Intell. Lab. Syst., 237 (2023), Article 104820
View PDF View article View in Scopus Google Scholar

[33] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
Google Scholar

[34] Van Dyk D.A., Meng X.-L.
The art of data augmentation
J. Comput. Graph. Statist., 10 (1) (2001), pp. 1-50, 10.1198/10618600152418584
Google Scholar

[35] Shanker M., Hu M.Y., Hung M.S.
Effect of data standardization on neural network training
Omega, 24 (4) (1996), pp. 385-397, 10.1016/0305-0483(96)00010-2
View PDF View article View in Scopus Google Scholar

[36] Chowdhury M.M., Ayon R.S., Hossain M.S.
An investigation of machine learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset
Healthc. Anal., 5 (2024), Article 100297
View PDF View article View in Scopus Google Scholar

[37] Bischl B., Binder M., Lang M., Pielok T., Richter J., Coors S., Thomas J., Ullmann T., Becker M., Boulesteix A.-L., *et al.*
Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges
Wiley Interdiscip. Rev.: Data Min. Knowl. Discov., 13 (2) (2023), Article e1484
View in Scopus Google Scholar

[38] Wong W., Juwono F.H., Capriono C.
Diabetic retinopathy detection and grading: A transfer learning approach using simultaneous parameter optimization and feature-weighted ECOC ensemble
IEEE Access (2023), 10.1109/ACCESS.2023.3301618
Google Scholar

[39] Nasajpour M., Karakaya M., Pouriyeh S., Parizi R.M.
Federated transfer learning for diabetic retinopathy detection using CNN architectures
SoutheastCon 2022, IEEE (2022), pp. 655-660, 10.1109/SoutheastCon48659.2022.9764031
View in Scopus Google Scholar

[40] Pavate A., Mistry J., Palve R., Gami N.
Diabetic retinopathy detection-MobileNet binary classifier
Acta Sci. Med. Sci., 4 (12) (2020), pp. 86-91
Crossref Google Scholar

[41] Kasim Ö.
Ensemble classification based optimized transfer learning feature method for early stage diagnosis of diabetic retinopathy
J. Ambient Intell. Humaniz. Comput. (2023), pp. 1-12, 10.1007/s12652-023-04648-z
Google Scholar

[42] Emon M.U., Zannat R., Khatun T., Rahman M., Keya M.S., *et al.*
Performance analysis of diabetic retinopathy prediction using machine learning models
2021 6th International Conference on Inventive Computation Technologies, ICICT, IEEE (2021), pp. 1048-1052, 10.1109/ICICT50816.2021.9358612
View in Scopus Google Scholar

[43] de Sousa T.F., Camilo C.G.
HDeep: Hierarchical deep learning combination for detection of diabetic retinopathy
Procedia Comput. Sci., 222 (2023), pp. 425-434, 10.1016/j.procs.2023.08.181
View PDF View article View in Scopus Google Scholar

[44] Nahidzuzzaman M., Islam M.R., Goni M.O.F., Anower M.S., Ahsan M., Haider J., Kowalski M.
Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier
Expert Syst. Appl., 217 (2023), Article 119557
View PDF View article View in Scopus Google Scholar

[45] Mohanty C., Mahapatra S., Acharya B., Kokkoras F., Gerogiannis V.C., Karamitsos I., Kanavos A.
Using deep learning architectures for detection and classification of diabetic retinopathy
Sensors, 23 (12) (2023), p. 5726
Crossref View in Scopus Google Scholar

[46] Atcı Ş.Y., Güneş A., Zontul M., Arslan Z.
Identifying diabetic retinopathy in the human eye: A hybrid approach based on a computer-aided diagnosis system combined with deep learning
Tomography, 10 (2) (2024), pp. 215-230
Crossref View in Scopus Google Scholar

[47] Mane D., Sangve S., Kumbharkar P., Ratnaparkhi S., Upadhye G., Borde S.
A diabetic retinopathy detection using customized convolutional neural network
Int. J. Electr. Electron. Res. (2023)
Google Scholar

[48] Gadekallu T.R., Khare N., Bhattacharya S., Singh S., Maddikunta P.K.R., Srivastava G.
Deep neural networks to predict diabetic retinopathy
J. Ambient Intell. Humaniz. Comput. (2023), pp. 1-14
Crossref Google Scholar

[49] Lalithadevi B., Krishnaveni S.
Diabetic retinopathy detection and severity classification using optimized deep learning with explainable AI technique
Multimedia Tools Appl. (2024), pp. 1-65
Google Scholar

[50] Sajid M.Z., Hamid M.F., Youssef A., Yasmin J., Perumal G., Qureshi I., Naqi S.M., Abbas Q.
DR-NASNet: Automated system to detect and classify diabetic retinopathy severity using improved pretrained NASNet model
Diagnostics, 13 (16) (2023), p. 2645
Crossref View in Scopus Google Scholar

[51] Atcı Ş.Y.
An integrated deep learning approach for computer-aided diagnosis of diverse diabetic retinopathy grading
International Conference on IoT and Health, Springer (2023), pp. 88-103
Google Scholar

[52] Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P.
SMOTE: Synthetic minority over-sampling technique
J. Artif. Intell. Res., 16 (2002), pp. 321-357, 10.1613/jair.953
View in Scopus Google Scholar

[53] Džeroski S., Ženko B.
Is combining classifiers with stacking better than selecting the best one?
Mach. Learn., 54 (2004), pp. 255-273, 10.1023/B:MACH.0000015881.36452.6e
View in Scopus Google Scholar

[54] Bodapati J.D., Shaik N.S., Naralasetti V.
Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification
J. Ambient Intell. Humaniz. Comput., 12 (10) (2021), pp. 9825-9839, 10.1007/s12652-020-02727-z
View in Scopus Google Scholar

[55] Mondal S.S., Mandal N., Singh K.K., Singh A., Izonin I.
EdlDr: An ensemble deep learning technique for detection and classification of diabetic retinopathy
Diagnostics, 13 (1) (2022), p. 124, 10.3390/diagnostics13010124
View in Scopus Google Scholar

[56] Lahmar C., Idri A.
Deep hybrid architectures for diabetic retinopathy classification
Comput. Methods Biomech. Biomed. Eng.: Imaging Visua., 11 (2) (2023), pp. 166-184, 10.1080/21681163.2022.2060864
View in Scopus Google Scholar

[57] Mohanty C., Mahapatra S., Acharya B., Kokkoras F., Gerogiannis V.C., Karamitsos I., Kanavos A.
Using deep learning architectures for detection and classification of diabetic retinopathy
Sensors, 23 (12) (2023), p. 5726, 10.3390/s23125726
View in Scopus Google Scholar

[58] Alyoubi W.L., Abulkhair M.F., Shalaash W.M.
Diabetic retinopathy fundus image classification and lesions localization system using deep learning
Sensors, 21 (11) (2021), p. 3704, 10.3390/s21113704
View in Scopus Google Scholar

[59] Farag M.M., Fouad M., Abdel-Hamid A.T.
Automatic severity classification of diabetic retinopathy based on densenet and convolutional block attention module
IEEE Access, 10 (2022), pp. 38299-38308, 10.1109/ACCESS.2022.3165193
View in Scopus Google Scholar

Cited by (6)

Advancing diabetic retinopathy classification using ensemble deep learning approaches
2025, Biomedical Signal Processing and Control
Show abstract

Uncertainty-aware diabetic retinopathy detection using deep learning enhanced by Bayesian approaches
2025, Scientific Reports

A multi model deep net with an explainable AI based framework for diabetic retinopathy segmentation and classification
2025, Scientific Reports

EffNet-SVM: A Hybrid Model for Diabetic Retinopathy Classification Using Retinal Fundus Images
2025, IEEE Access

Lesion Segmentation in Retinal Images for Diabetic Retinopathy Diagnosis Using Weighted Gradient Assistance
2025, International Journal of Intelligent Engineering and Systems

Application of Artificial Intelligence for Classification, Segmentation, Early Detection, Early Diagnosis, and Grading of Diabetic Retinopathy From Fundus Retinal Images: A Comprehensive Review
2024, IEEE Access

© 2024 The Authors. Published by Elsevier Inc.

About ScienceDirect Remote access Advertise Contact and support
Terms and conditions Privacy policy

Cookies are used by this site. Cookie Settings
All content on this site: Copyright © 2025 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the relevant licensing terms apply.

FEEDBACK