Jagannath University Journal of Science
Volume 11, Number I, Jun. 2024, pp. 135−156
https://jnu.ac.bd/journal/portal/archives/science.jsp
ISSN 3005-4486 (Online), ISSN 2224-1698 (Print)

Enhancing Accuracy of Diabetic Retinopathy Detection Using a Hybrid Approach with the Fusion of Inceptionv3 and a Stacking Ensemble Learner

Research Article

Intifa Aman Taifa[1], Tania Islam[1,*], Md Aminul Islam[2], Md Mahbub E Noor[1] and Tazizur Rahman[3]

[1]*Department of Computer Science and Engineering, University of Barishal, Barishal- 8254, Bangladesh*

[2]*Department of Computer Science and Engineering, Jagannath University, Dhaka-1100 Bangladesh*

[3]*Department of Management Studies, University of Barishal, Barishal- 8254, Bangladesh*

DOI: https://doi.org/10.3329/jnujsci.v11i1.76699

Received: 19 March 2024, Accepted: 23 June 2024

ABSTRACT

Diabetic retinopathy (DR) is a severe global problem that affects millions of people worldwide and gets worse over time. If left untreated, DR can lead to blindness. Early and precise DR identification is necessary to address this developing challenge. Traditional approaches include applying machine learning or deep learning algorithms directly on the dataset or the preprocessed dataset, which has shown very good results recently. Very few focused on combining machine learning and deep learning-based algorithms. Extracting a good set of features is very crucial to getting higher performance from any machine learning-based or deep learning-based algorithm. This work presents a new method for DR detection by fusing a convolution neural network-based feature extraction method before feeding the data to a stacking ensemble learner, which uses several machine learning algorithms to make it more robust. Predictions from several classifiers, including decision trees, random forests, support vector machines, logistic regression, and others, have been used in previous studies of DR. In our work, we used these classifiers for our hybrid model. First, retinal image features are extracted using InceptionV3. Then, several fine-tuned machine learning-based classifiers have been used. Finally, all the classifier models are stacked together to create an ensemble model. Our hybrid approach showed promising performance in classifying binary (98.64%) and multi-class (94.95%) on the APTOS 2019 Blindness Detection dataset. This finding proves that our

Corresponding Author: Tania Islam
E-mail: tania.bd.09@gmail.com

hybrid technique is more capable than the traditional approach for the early diagnosis of diabetic retinopathy and offers a great hope for better medical intervention.

Keywords: *Diabetic retinopathy, Hybrid model, Stacking, Machine learning, Inception V3.*

1. Introduction

Diabetic Retinopathy (DR), a complicated ocular disorder that can result in vision loss, is a shadow of diabetes, a common metabolic condition, and persistent (Daniel et al. et al., 2016). DR is a major cause of visual impairment, especially in working-age individuals, as it gradually attacks the small blood vessels in the retina. The primary cause of this condition is chronically high blood sugar, which harms the retina's delicate blood vessels (Renu et al., 2007). These changes to blood vessels, which can take many different shapes, indicate diabetic retinopathy.

DR can be classified into five main types based on the alterations in blood vessels: No DR, Moderate NPDR, Severe NPDR, Mild Non-Proliferative DR, and Proliferative DR. The different stages represent different degrees of microvascular changes and retinal degeneration. The absence of noticeable retinal degeneration is a defining feature of the first stage, No-DR (Zihan Sun et al., 2019). Microaneurysms or tiny patches of balloon-like swelling in the retinal blood vessels are common in mild nonproliferative diabetic retinopathy (NPDR) and frequently indicate the beginning of more severe stages of the illness (Mo-Chi Yang et al., 2020). Because of obvious microvascular damage, such as blockages and vessel closure, there is a reduced blood flow to the retina with moderate NPDR (Janice et al. et al., 2020). Exacerbated blood artery obstructions in Severe NPDR considerably restrict blood supply to particular retinal locations, which may lead to aberrant growth of blood vessels (Nevill et al., 2022). The appearance of brittle, malformed blood vessels on the retina's surface characterizes PDR, the highest stage, and vessel leaks carry a significant risk of blindness (Le et al., 2021).

Additionally, hypertension, hyperlipidemia, genetic predisposition, long-term diabetes, and insufficient glycemic management are risk factors for the development of diabetic retinopathy (Clermont et al., 2007). Because it can significantly reduce a person's vision and quality of life, diabetic retinopathy is becoming a serious worry for those who have the disease. To enable proper clinical intervention, it is critical to identify and categorize the severity of DR (Taifa et al., 2024). Prompt diagnosis and treatment can significantly reduce the risk of vision impairment and blindness.

Diabetic retinopathy is currently receiving a lot of attention from researchers. To distinguish between various eye conditions, they are investigating both conventional image processing and machine learning techniques (Kar et al., 2017). Convolutional neural networks (CNNs) (Islam et al., 2022), in particular, are becoming increasingly popular deep learning models for analyzing ocular images. By identifying intricate patterns and problems in retinal pictures, these models can aid in the early detection of diabetic retinopathy (Varun et al., 2016). A clever method of using deep learning for early diabetic retinopathy identification from eye pictures was presented by Tymchenko et al. (Tymchenko et al., 2020). Increasing the sensitivity and specificity of the detection procedure is their primary objective. However, they also point out several drawbacks, such as the requirement that these models function effectively across various datasets, possible problems with labeled datasets, and an excessive reliance on specialists.

In the following related works section, we have mentioned several traditional approaches to detect this DR effectively. They used different methods, such as machine learning approaches, deep learning architectures, transfer learning and fine-tuning, federated learning and privacy-preserving techniques, and ensemble learning and hybrid architectures. Ensemble learning and hybrid architectures (Lahmar et al., 2023; Alyoubi et al., 2021; Wong et al., 2023) have shown better overall results. However, they didn't focus on better feature extraction parts, which is crucial for any machine learning or deep learning-based models. This study thoroughly investigates binary and multi-class classification for diabetic retinopathy (DR). Using a stacking ensemble learning strategy, our method presents a hybrid approach that combines predictions from various machine learning (ML) classifiers and a state-of-the-art deep learning model-based feature extractor. A variety of classifiers, such as Support Vector Machine (SVM) (Chandra et al., 2021), Logistic Regression (LR) (Denœux et al., 2019), Random Forest (RF) (G. Biau et al., 2016), K-Nearest Neighbors (KNN) (Gongde et al., 2003), Naive Bayes (NB) (I. Rish et al., 2001), Extra Tree (ET) (Shafique et al., 2019), Multilayer Perceptron (MLP) (Windeatt et al., 2008), Passive Aggressive (PA) (Chien-Chung et al., 2010), Perceptron (Stephen et al., 1990), Ridge Classifier (Chong et al., 2020), Stochastic Gradient Descent (SGD) (Fasihul et al., 2015), Nearest Centroid (NC) (Sonika et al., 2021), and Decision Tree (DT) (Yan-Yan Song et al., 2015) are used in an attempt to guarantee the reliability of our approach and making it more robust. Furthermore, retina scan pictures have been fed to a cutting-edge feature extractor, InceptionV3 (Gaurav et al., 2023), to extract discriminative features that capture minute details and patterns. Several preprocessing techniques, including data augmentation (Van Dyk et al., 2001), data normalization (Murali et al., 1996), and SMOTE oversampling (Nitesh et al., 2002), are included in order to enhance the performance of the models. Moreover, each classifier's performance is optimized by applying hyperparameter tuning (James et al., 2011).

Henceforth, this study's main contributions are as follows:

- Firstly, the study examines the effectiveness and performance assessment of a hybrid model and a stacking ensemble learning in the classification of diabetic retinopathy.
- Secondly, this research carries out in-depth research to enhance the hybrid model performance by fusing a cutting-edge deep learning-based feature extraction method, InceptionV3.
- Finally, we applied our hybrid model and several preprocessing methods to the APTOS 2019 Blindness Detection Kaggle dataset and benchmarked new and improved results.

The structure of this paper is as follows: Section II presents a thorough review. Section III describes the experimental setup. Section IV presents the experimental outcomes of this investigation. Section V summarizes the findings and offers suggestions for further study directions.

2. Related Works

Numerous research endeavors have been undertaken to enhance the accuracy and efficacy of diabetes-associated visual impairments, including diabetic retinopathy (DR), a grave ocular ailment.

M. U. Emon et al. (Emon et al., 2021) employed a wide variety of machine learning algorithms, including logistic regression, Naive Bayes, bagging classifiers, J48 classifiers, random forest classifiers, and stochastic gradient descent (SGD). The maximum accuracy of logistic regression was 75%, and its ROC curve value is relatively high at 83%.

De Sousa et al. (T.F. de Sousa et al., 2023) demonstrated a binary hierarchical combination of four convolutional neural networks. The approach obtained a Kappa score of 0.911 and an accuracy of 0.853 on the Kaggle APTOS dataset. Pavate Aruna et al. (Aruna et al., 2020) applied a MobileNet-based architecture trained on retinal fundus pictures from the Aptos 2019 challenge dataset to address the problem of early diagnosis of DR. With a binary classification technique, the model achieved strong recall, F-1, and precision scores (0.95, 0.98, and 0.97, respectively), resulting in a 95% accuracy rate.

Wong et al. (WK Wong et al., 2023) presented a Diabetic Retinopathy (DR) diagnosis method based on Transfer Learning (TL). This approach for classification integrated features from ResNet-18 and ShuffleNet with an Error Correction Output Code (ECOC) ensemble. With the use of the EyePac + Messidor-2 datasets, automated parameter setting using Adaptive Differential Evolution (ADE) produced grading accuracy of 82% for APTOS 5-class DR, 96% for APTOS 2-class grading, and 75% for 3-class grading. Although handling a variety of datasets and DR detection performance variances between classes could be challenging, increased accuracy might be achieved.

When utilizing Federated Learning in a decentralized training environment, Nasajpour et al. (M. Nasajpour et al., 2022) placed a high premium on privacy protection. They employed three models: Federated Averaging (FedAVG), Federated Proximal (FedProx), and conventional transfer learning. They obtained 92.19%, 90.07%, and 85.81% DR detection accuracies for each model.

Ömer Kasim (Ömer Kasim, 2023) concentrated on applying preprocessed images to MobileNet transfer learning and ensemble learning with Minimum Redundancy and Maximum Relevance feature selection. Using the Messidor, Aptos 2019, and DDR datasets, respectively, they obtained accuracies of 98.6%, 88.95%, and 97.46% for DR's early-stage detection. Furthermore, multiclass classification of all stages of DR was reported to be 96.06%, 89.86%, and 82.74% on the Messidor, Aptos 2019, and DDR datasets, respectively.

Deep learning is frequently used in research for disease-related classification. However, its combination with machine learning is not very common. Extant literature is deficient in thorough analyses, especially binary and multi-class classifications. This work suggests merging deep learning and machine learning techniques for reliable feature extraction to improve the overall performance.

3. Methodology

This section breaks down the technique for diagnosing diabetic retinopathy (DR) in its early stages into eight distinct phases. Figure 1 provides a comprehensive road map for comprehending the formation of our hybrid model, graphically illustrating the links between these steps. The multiphase model, which prioritizes accuracy and precision, performs well in detecting early signs of DR.

The investigation begins with presenting our workflow diagram and then goes into great detail to examine each important stage. Each stage is crucial to our all-encompassing strategy for the timely identification of diabetic retinopathy, ensuring total and effective preservation of retinal integrity. For our experiment, we first collected our data. The data preparation stages are done. In this stage, data was preprocessed to find better performance for our model. After that, feature extraction is done using a deep learning model named InceptionV3. Several traditional machine learning models were chosen as the base classifiers for our stacking ensemble model. For training those models, we adjusted the hyperparameters and then optimized them to get better performances. After getting distinct prediction outputs from each base classifier, we made another training set from those predictions for our final meta-model for the final prediction from our ensemble model. The above steps are described in detail below.

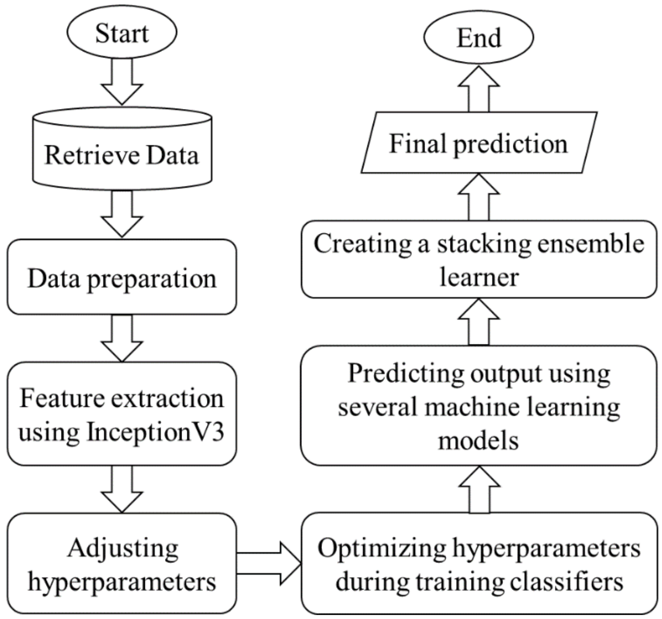


Fig 1. Workflow diagram

3.1. Data Collection

This work uses the APTOS 2019 dataset, which is freely accessible on Kaggle and consists of excellent retinal images divided into five groups that correspond to different severity levels of diabetic retinopathy. This diversified dataset makes a comprehensive study across different severity levels possible, improving model performance and evaluation. The dataset's severity levels are shown in greater detail in Figure 2.

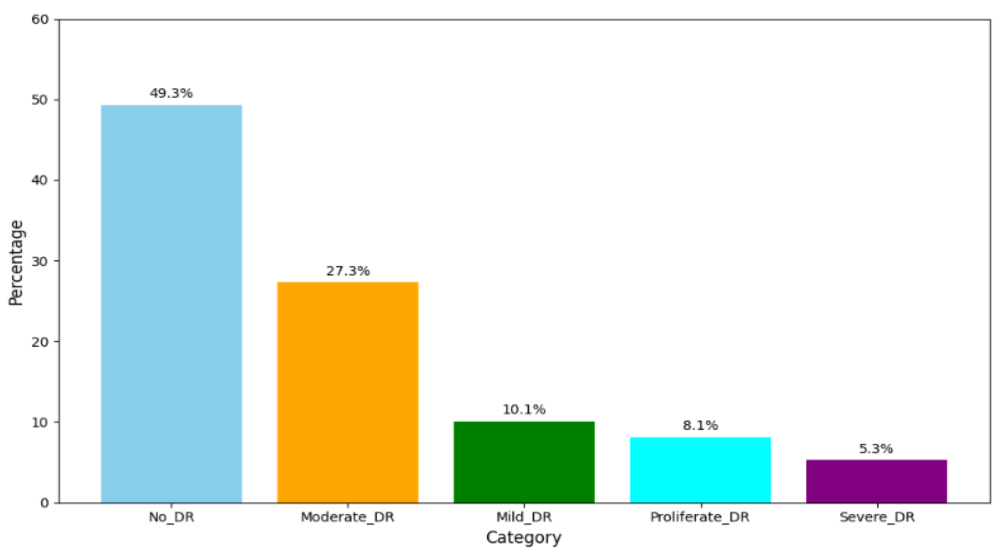


Fig 2. The APTOS dataset's severity levels

We have worked on the initially publicly available training set of 3,662 images for proper comparison with other existing works.

3.2. Data Preprocessing

For reliable analysis, preprocessing is necessary, and this is achieved by applying a Gaussian filter to ensure dataset consistency. Resizing all photographs to the common 224x224 pixel size will make them more compatible. Important modifications such as rotation, zoom, and color corrections are incorporated into the stages to improve training efficiency and increase the variety of datasets. Photos before preprocessing are displayed in Figure 3, whereas many preprocessed photos are displayed in Figure 4.

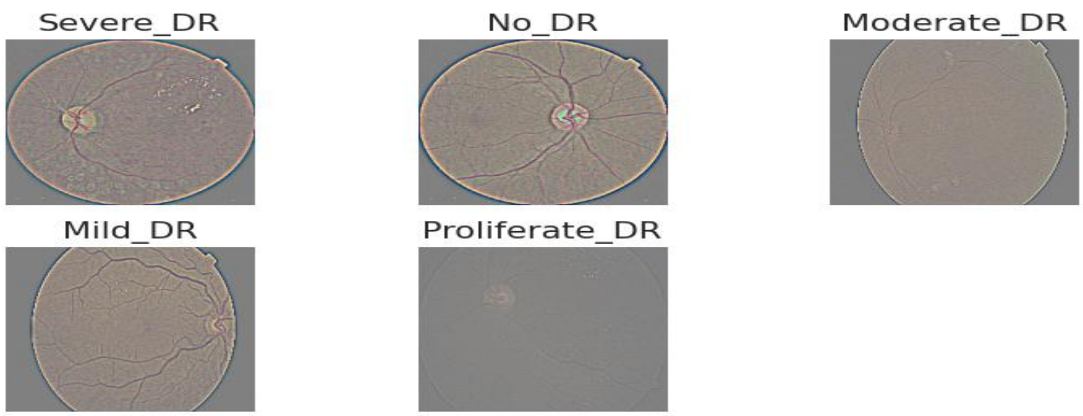


Fig 3. Images right before preprocessing

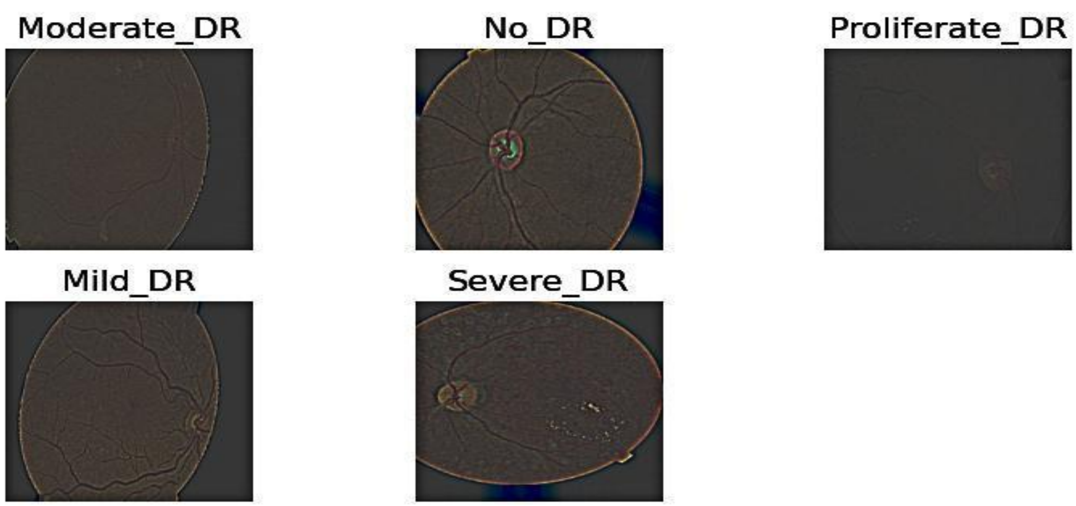


Fig 4. Preprocessed images

It is vital to rectify imbalances in the class distribution within a dataset. This step describes the resampling of the dataset also. Figure 5 shows a schematic representation of the phases. Some of the resampled photos are displayed in Figure 6.

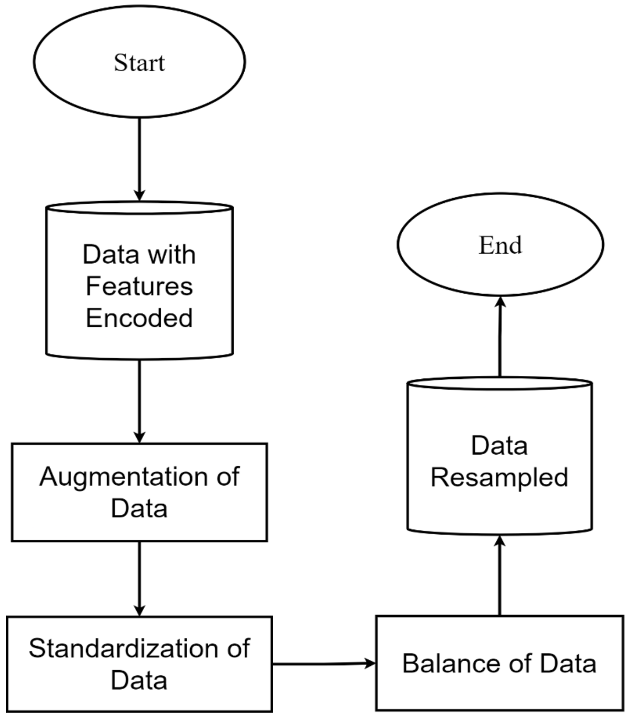


Fig 5. Data preprocessing steps before training

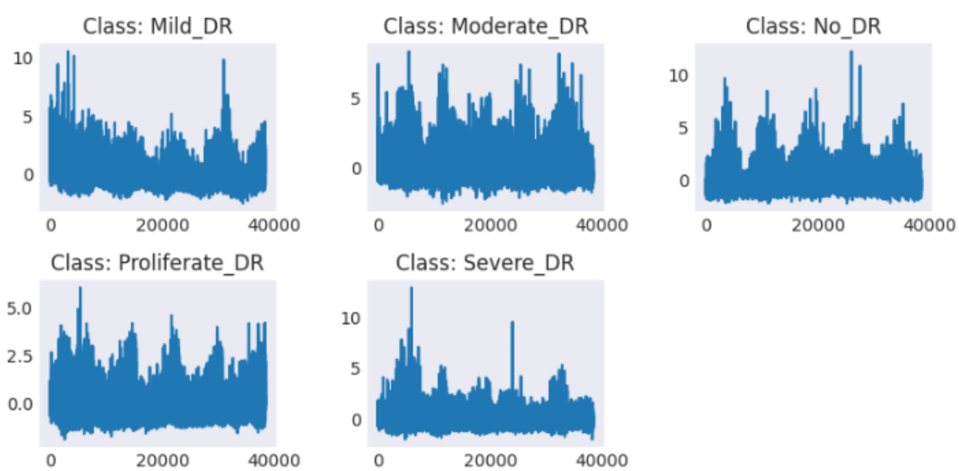


Fig 6. Resampled Datasets

Here is an explanation of the steps required in data resampling:

3.2.1. Data Augmentation The current study employed the data augmentation approach (Van Dyk et al., 2001) to enhance dataset diversity and mitigate the likelihood of overfitting. This involved starting fresh in the dataset by adding different effects to pre-existing photos. Dynamic elements, such as random rotations (between -30 and +30 degrees), were used to change the orientations of the objects. Zoom levels (up to 15%) were changed to ensure the model could recognize items of various sizes. The model's ability to adjust to changes in item placement was enhanced by adding randomness from arbitrary rotations (up to 20% in both directions) and shearing (up to 15 degrees).

The orientation-independent properties of the model were described, and an environment of similarity was established by horizontal flipping.

3.2.2. Data Standardization

To ensure dataset uniformity, data standardization (Murali et al., 1996) scales values for feature homogeneity. To guarantee that feature vectors rescaled to a mean of 0 and a standard deviation of 1 are executed by machine learning algorithms that assume features are on the same scale. Standardized features keep the learning process from being dominated by a single feature, increasing the model's overall resilience and efficacy during training and evaluation.

3.2.3. Data Balancing

Synthetic Minority Over-sampling Technique (SMOTE) (Nitesh et al., 2002) addresses imbalanced datasets in classification problems. The training dataset is increased by using this technique, which encourages diversity and improves the model's ability to generalize to untested data. Producing synthetic samples guarantees more effective minority-class learning without majority-class bias, which enhances classification performance, particularly for minority classes.

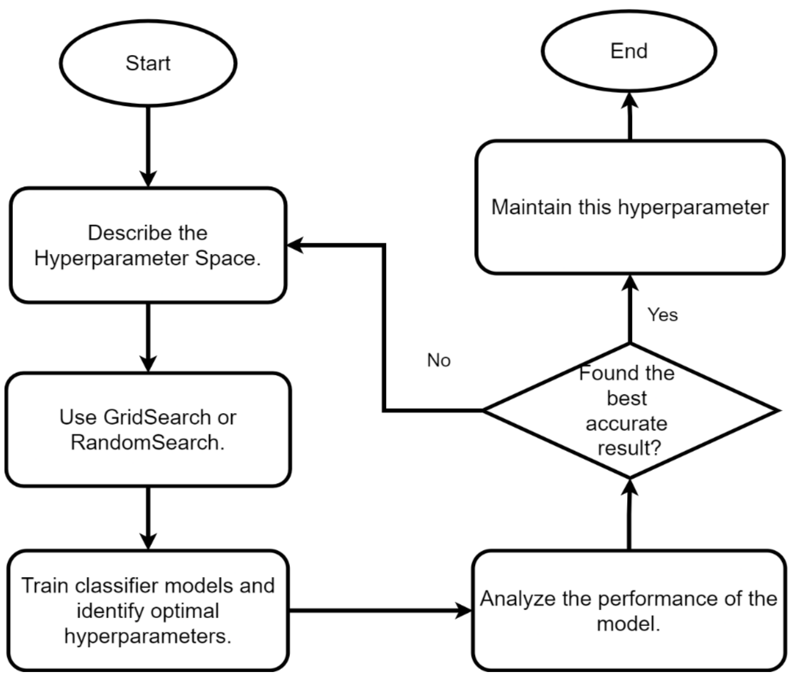


Fig 7. Flowchart of hyperparameter tuning steps

3.3. Hyperparameter Tuning

Hyperparameter tuning (James et al., 2011), sometimes called hyperparameter optimization or hyperparameter search, is crucial to training machine learning models. It involves figuring out which hyperparameters work best for a specific machine-learning algorithm and dataset. The data do not determine hyperparameters or settings made before the model is trained. They have an impact on the learning process of the model, which has an impact on the model's generalization, computing performance, and efficiency. Both Grid Search and Random Search were employed in this study to adjust hyperparameters for various machine-learning algorithms. Figure 7 shows this method in further detail**.**

3.4. Training and Testing Model

The dataset is thoughtfully split into testing and training chunks, with nice 80:20 ratios. To observe the data's separability, we separately applied t-SNE in this section to the features collected from the training and testing datasets shown in Figure 8. In this case, the groupings "No_DR," "Mild_DR," "Moderate_DR," "Severe_DR," and "Proliferate_DR" are represented by classes 0 through 4.

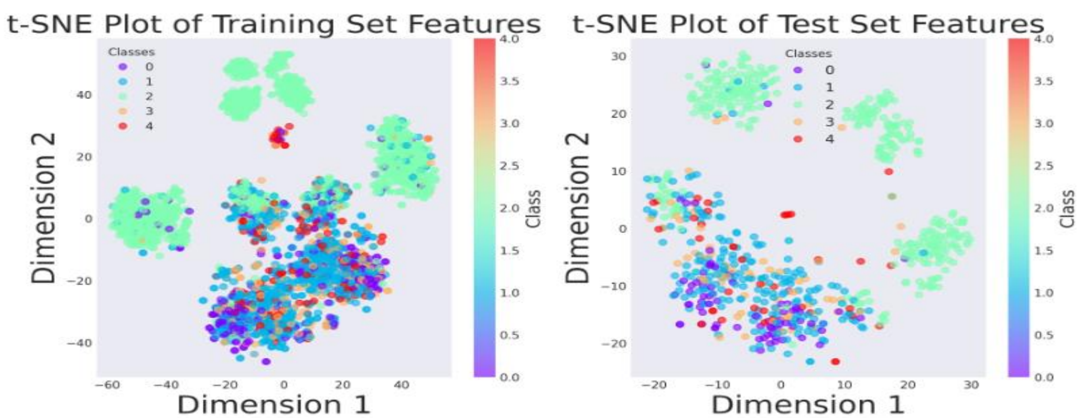


Fig 8. t-SNE plot of the multi-class training and testing dataset

3.5. Label Reduction to Binary-Class Model

This study evaluates binary classification in addition to multi-class classification. To make the classification procedure easier, a distinction is made between "DR" and "No_DR" situations. The issue is reduced to binary classification by classifying the label "No Diabetic Retinopathy" as "No_DR" and the remaining categories as "DR." The Binary Classification procedure is shown in Table 1.

Table 1. Simplified Binary Classification

Grades	Severity Levels of DR	Label Reduction
0	No Diabetic Retinopathy	No_DR
1	Mild_DR	DR
2	Moderate_DR	DR
3	Severe_DR	DR
4	Proliferate_DR	DR

Figure 9 shows the separability of the two-class separations for the testing and training dataset. Class 0 denotes the "No_DR" group, while class 1 denotes the "DR" group.

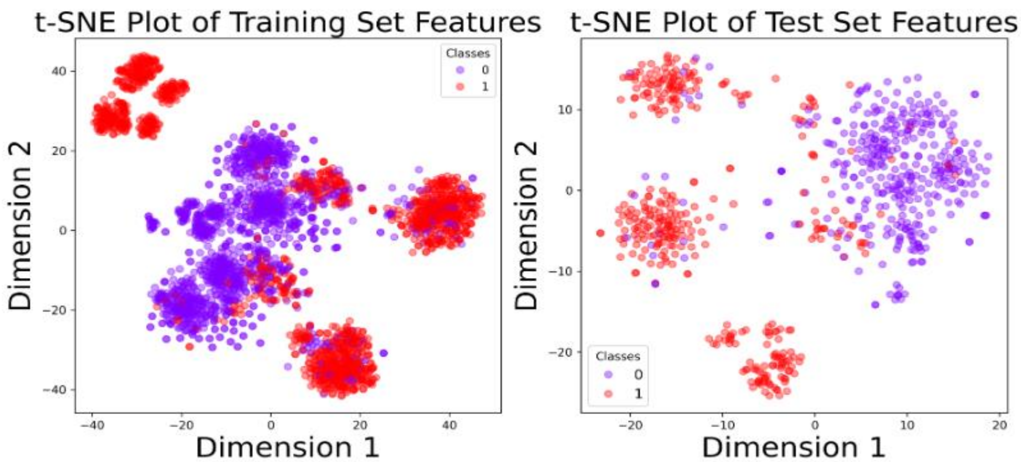


Fig 9.t-SNE plot of the binary class training and testing data

3.6. Feature Extraction

The process of feature extraction is essential to the classification of diabetic retinopathy. The primary objective is to transform raw retinal images into relevant and discriminative features that classifier systems may utilize to identify diseases precisely. In this work, InceptionV3, a pre-trained deep learning network, is used to extract features that recognize DR-relevant patterns in retinal imagery. In addition to improving the overall effectiveness of the study, this feature extractor is crucial for diagnosing and categorizing diabetic retinopathy.

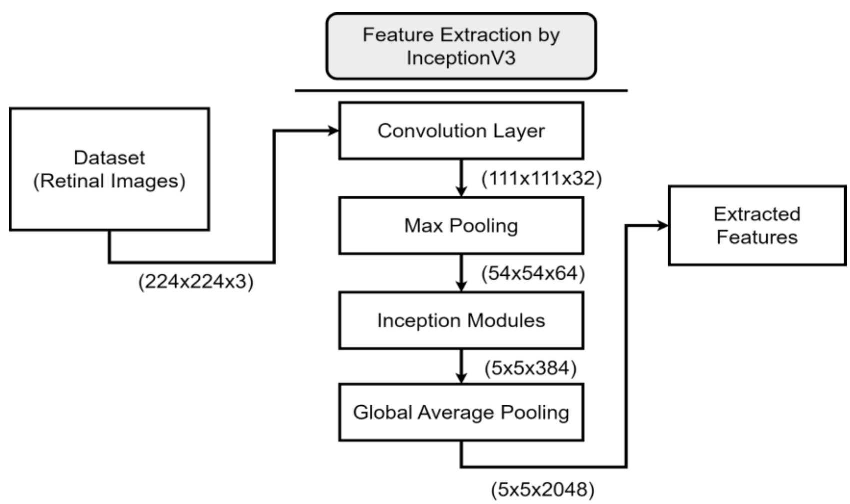


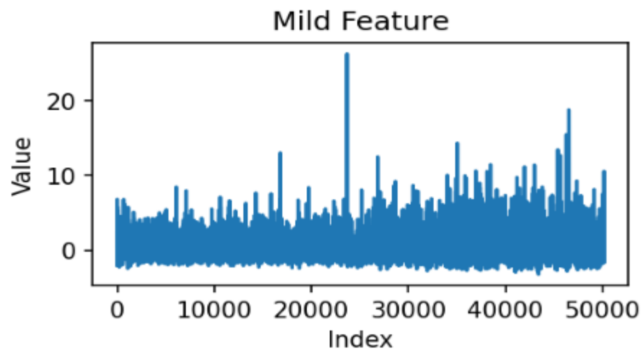
Fig 10. Structure of InceptionV3 model

3.6.1. InceptionV3

A well-known deep convolutional neural network architecture, InceptionV3 is especially popular for its performance on various computer vision tasks, most notably image classification. Its design complexity resides in the use of inception modules, which enable effective parameter usage and feature extraction. These modules, exhibited in the architecture shown in Figure 10, allow features to be explored at different spatial scales using various convolutional filters, including 1x1, 3x3, and 5x5 convolutions, in addition to max pooling procedures. This kind of thorough examination of image characteristics improves the model's ability to identify complex patterns in the data. Taking this technique eliminates any classification-specific computations from the model and leaves it focused only on extracting rich and discriminative features from input images.

The input consists of the preprocessed photos from the dataset. Figure 11 shows the input image and the features that were retrieved for index 0.

(a) Input image



(b) Extracted Feature Using InceptionV3

Fig 11. Using InceptionV3, the input image's features were retrieved for Index 0.

To optimize the identification of diabetic retinopathy (DR), InceptionV3 was selected as the feature extractor for this investigation. Its dense connectivity patterns enhance the capture of complicated properties.

3.7. Stacking Ensemble Model

This study uses a sophisticated ensemble technique termed stacking, initially reported by Wolpert in 1992 (Saso Džeros et al., 2004), to construct a robust model for Diabetic Retinopathy (DR) diagnosis. The stacking strategy enhances model performance by letting each basic classifier showcase its capabilities by judiciously mixing its prediction powers. This method has been widely accepted in the field of ensemble learning and demonstrates the capacity to integrate distinct ideas from different classifiers.

Stacking stands out from other ensemble techniques because it can integrate distinct insights from several traditional and basic classifiers. DT, RF, SVM, KNN, LR, ET, MLP (Islam et al., 2023), NC, PA, Perceptron, Ridge Classifier, and SGD models are chosen for this ensemble approach. Using the diverse viewpoints of various classifiers, the stacking technique ensures a comprehensive understanding of the underlying data and contributes to producing an exhaustive DR classification.

After the feature extraction step by InceptionV3, the predictions produced by each base classifier are extracted individually from the same training and testing sets of the dataset. This consecutive extraction allows for the creation of a new training set in the following stage. The distinct predictions generated by each basic classification method have been carefully combined to create this set.

In this manner, the enchantment happens at the next level. The meta-model, a potent MLP classifier with certain hyperparameters optimized for peak performance, is chosen for our stacking ensemble model for the final classification of the newly created training data. In this scenario, stacking works well because every base classifier offers a different viewpoint, and the meta-model learns how to combine these clever concepts into a final prediction mechanism. To attain this combination, the MLP Classifier is meticulously trained on the complete set of predictions, guaranteeing a complete understanding of the complexity contained in the dataset.

The last phase involves a thorough study based on the final forecast of our hybrid model. The evaluation explains how successfully the stacked ensemble distinguishes between the different phases of diabetic retinopathy. The meta-model's careful selection of basic classifiers demonstrates the effectiveness of the stacking strategy for improving DR classification.

Thus, our stacking approach of the ensemble model for classifying DR was created and illustrated in Figure 12.

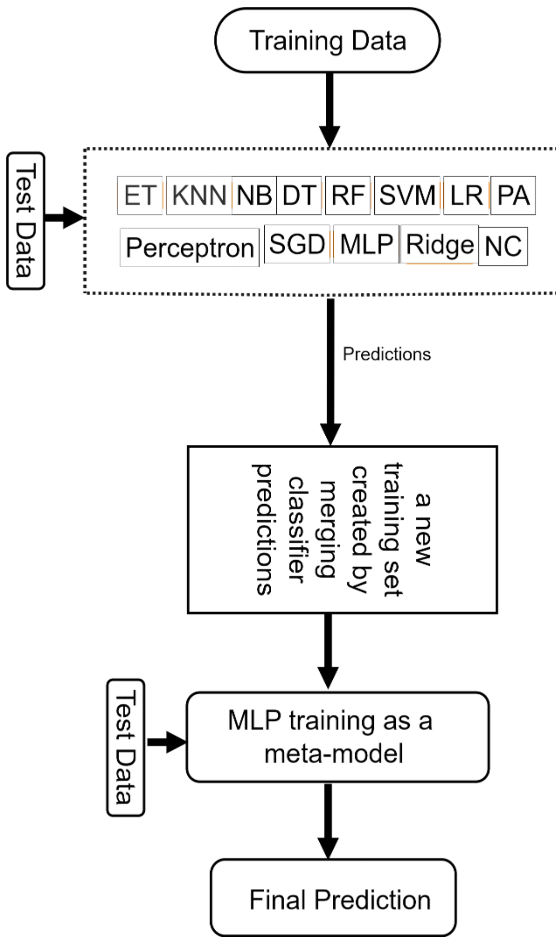


Fig 12. Steps of the stacking approach for the ensemble learner

4. Results and Discussion

This section will include the precise outcomes of this procedure in classifying diabetic retinopathy. The performance of a hybrid model is thoroughly examined, accounting for both binary and multiple classifications. This research will look at the effects of various feature extractors and investigate how stacking might be used to enhance classification performance.

4.1. Performance Results for Multi-Class Classification

Table 2 shows the outcomes of the multi-class classification experiment conducted using features extracted from the InceptionV3 model. This set of eighteen retrieved features goes through around thirteen classifiers before appearing in a hybrid model we created. Compared to the previous ML techniques, the suggested hybrid model provides multi-class classification with a maximum accuracy of 94.95%.

Table 2. Overview of Algorithm Performance for Multi-Class Using InceptionV3

No	Algorithm	Accuracy
1	SVM	80.22%
2	Logistic Regression	79.26%
3	Random Forest	78.85%
4	KNN	56.48%
5	Naive Bayes	63.57%
6	Extra Trees	78.85%
7	MLP	78.99%
8	Passive Aggressive	76.94%
9	Perceptron	76.13%
10	Ridge	80.90%
11	SGD	79.54%
12	Nearest Centroid	65.89%
13	Decision Tree	63.85%
14	**Proposed Hybrid Model**	**94.95%**

In addition, Table 3 displays the F1-score, recession, recall, and support for multiclass classification utilizing InceptionV3.

Table 3. Performance Summary of the Proposed Model for Multi-Class Using InceptionV3

	Precision	Recall	f1-score	support
Mild	93	88	90	74
Moderate	91	95	93	200
No_DR	98	100	99	361
Proliferate_DR	96	81	88	59
Severe	92	85	88	39
accuracy			95	733
macro avg	94	90	92	733
weighted avg	95	95	95	733

Figure 13 illustrates the accuracy vs. epoch and loss vs. epoch curves while training our model for multi-class classification.

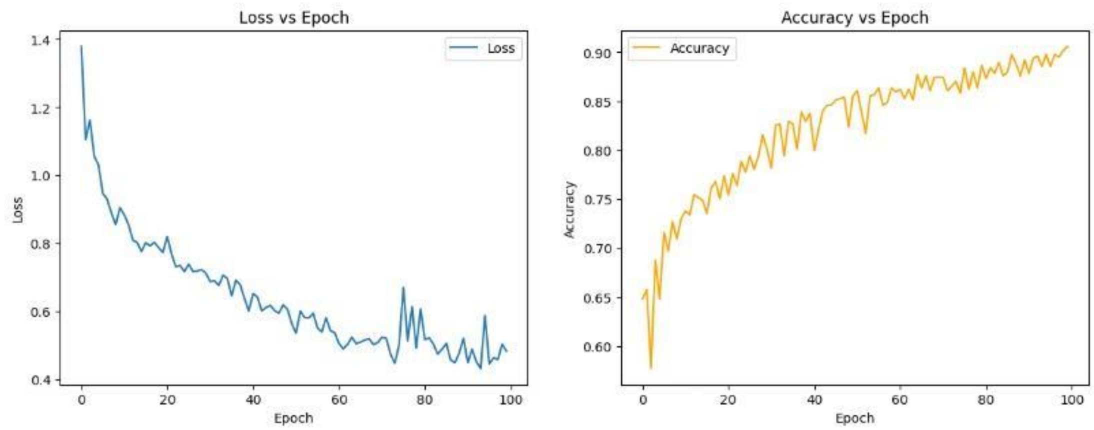


Fig 13. Loss vs. epoch and accuracy vs. epoch curves while training the Proposed Model with InceptionV3 for Multi-Class classification.

Figure 14 displays the ROC-AUC curve for InceptionV3, the accuracy of each class, and the total number of wrong predictions. Class Severe and No_DR had the greatest AUC values (0.99), while Mild, Proliferate_DR, and Moderate have the lowest values (see Figure).

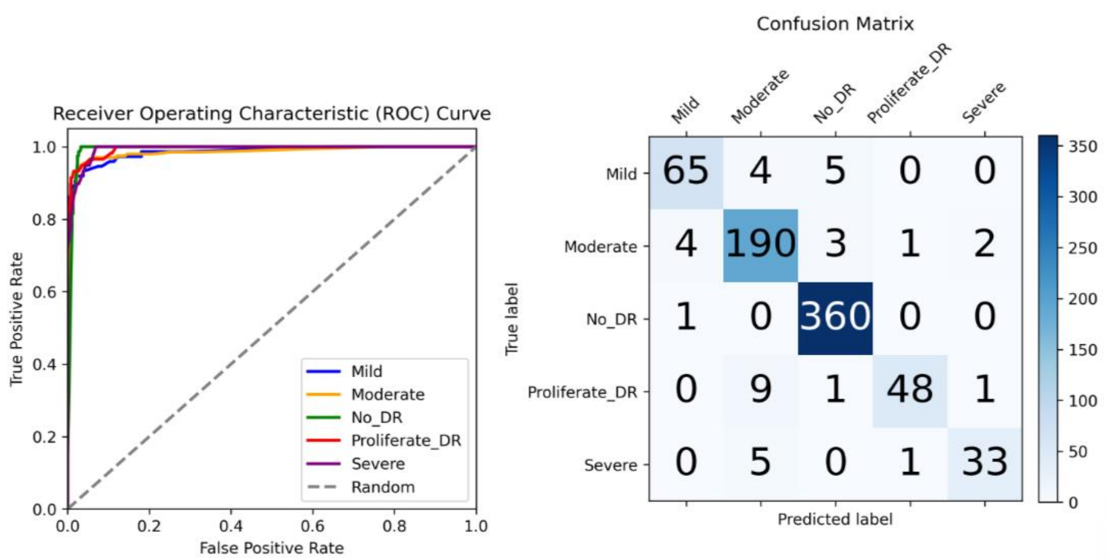


Fig 14. The Proposed Model's ROC-AUC Curve and Confusion Matrix for Multi-Class Classification Using InceptionV3.

The intricate design of InceptionV3 facilitates effective feature extraction and interpretability. It does a good job of spotting complex patterns in diagnosing diabetic retinopathy but has trouble with fine-grained classification. Erroneous categorizations underscore the necessity for continuous advancement in comprehending feature representations. Even with these drawbacks, InceptionV3 is still a useful diagnostic tool.

4.2. Performance Results for Binary Classification

This section explains binary classification's performance outcomes. Table 4 shows the results of methods that use the InceptionV3 model as a feature extractor. With a maximum accuracy of 98.64%, the suggested hybrid model performs better in binary classification than the other ML approaches in this case. It performs well in Table 5 with precision, recall, and f1 scores of 99%, 99%, and 99% for DR and No_DR classes, respectively.

Table 4. Overview of Algorithm Performance for Binary Classification Using InceptionV3

No	Algorithm	Accuracy
1	SVM	96.86%
2	Logistic Regression	96.86%
3	Random Forest	95.91%
4	KNN	93.04%
5	Naive Bayes	90.18%
6	Extra Trees	95.91%
7	MLP	96.45%
8	Passive Aggressive	96.59%
9	Perceptron	94.00%
10	Ridge	96.86%
11	SGD	96.32%
12	Nearest Centroid	89.50%
13	Decision Tree	89.77%
14	Proposed Hybrid Model	98.64%

Table 5. Performance Summary of the Proposed Model for Binary Classification Using InceptionV3

	Precision	Recall	f1-score	support
DR	99	99	99	372
No_DR	99	99	99	371
accuracy			99	733
macro avg	99	99	99	733
weighted avg	99	99	99	733

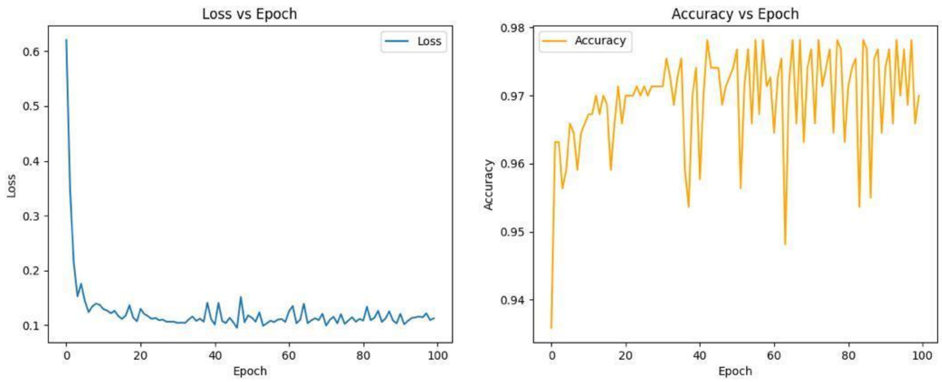


Fig 15. Loss vs. epoch and accuracy vs. epoch curves while training the Proposed Model with InceptionV3 for binary Class classification.

Figure 16 shows the confusion matrix of the suggested model using InceptionV3 for binary classification. Additionally, it shows the InceptionV3 ROC curve for binary classification, which has AUC values of 0.98 for both Class DR and Class No_DR, demonstrating the classifier's potent discriminating power. The model does remarkably well When differentiating between positive and negative examples of these classes.

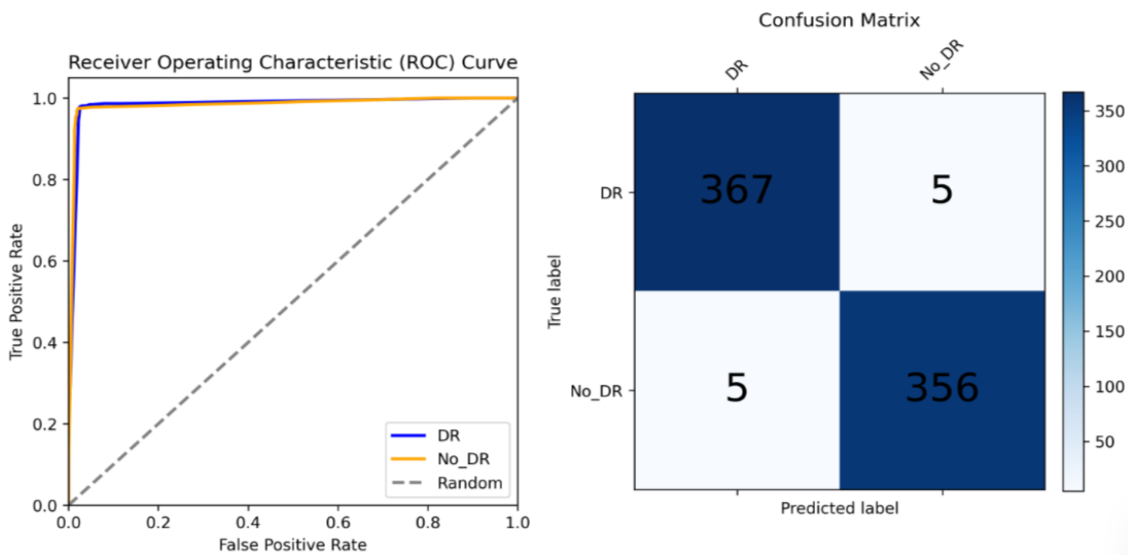


Fig 16. Confusion Matrix and ROC Curve for the Proposed Model for Binary Classification Using InceptionV3

The hybrid model obtained 98.64% accuracy in differentiating between the DR and No_DR categories using InceptionV3 as the feature extractor. However, the DR severity level imbalance in the dataset continues to be a problem for the generalization of the model. When determining the degree of diabetic retinopathy, InceptionV3 is a dependable choice.

4.3. Computational Efficiency of Each Model

This study used training/prediction times and memory use to evaluate the deep learning model's performance for the hybrid model. In the binary classification job, the "2-class-with-InceptionV3" model memory use was composed of the following: basic models (78,024 bytes), stacked data (76,232 bytes), meta-model (90,224 bytes), and stacked test data (76,232 bytes). The forecast time for stacking operations was 0.621 seconds, while the meta-model training was 0.0016 seconds.

When it transitioned to the multi-class classification scenario, the "5-class-with-InceptionV3" model showed memory utilization of 71,824 bytes for base models, 76,360 bytes for stacked data, 48 bytes for the metamodel, and 76,360 bytes for stacked test data. The metamodel's training time was calculated to be 1.56 seconds, and the stacking operation prediction time was found to be 0.0029 seconds.

Impressively, InceptionV3 obtains a binary classification accuracy of 98.64% and a multi-class classification accuracy of 94.95%. Because of the architecture's high degree of interaction, which facilitates efficient information movement across layers, complex patterns can be found to assess the extent of data loss.

The optimal model, however, would vary depending on certain needs and trade-offs, including interpretability, processing power, and accuracy. Before choosing the best model for the situation, it is essential to measure these factors precisely.

4.4. Comparative Analysis

We demonstrate the effectiveness of our proposed model with a thorough comparative analysis of our work in this part.

The study's experimental results show that this hybrid model, which includes InceptionV3, has the highest accuracy rates for detecting diabetic retinopathy compared to other models. A properly labeled dataset, multiple preprocessing stages, and feature extraction lead to an efficient model execution. We outperform other models with accuracy rates of 94.95% in multi-class classification and 98.64% in binary classification.

The proposed methodology is distinguished by its thoughtful dataset gathering, innovative stacking strategy, careful integration of binary and multi-class classifications, and extensive data pre-processing. By setting the foundation for an effective hybrid model, this thorough and innovative method shows the potential to significantly advance the field of Diabetic Retinopathy diagnosis. Table 6 compares the multiclass with other current efforts, and Table 7 shows the binary class.

Table 6. Comparative Study of DR for Multi-Class Categorization

Author	Model	Dataset	Accuracy
Emon et al., 2021	Logistic	UCI	75%
Wong et al., 2023	ECOC Ensemble	APTOS	82%
T.F. de Sousa et al., 2023	Combination of 4 CNN	APTOS	85.3%
Lahmar et al., 2023 Hybrid (SVM+MobileNetV2)		APTOS	88.80%
		Kaggle DR	84.01%
		Messidor-2	84.05%

Alyoubi et al., 2021	Fusion of CNN512 & YOLOv3	APTOS	89%
This study	**Hybrid model (stacking ensemble learner + InceptionV3)**	**APTOS**	**94.95%**

Table 7. Comparative Study of DR for Multi-Class Categorization

Author	**Model**	**Dataset**	**Accuracy**
Wong et al., 2023	ECOC Ensemble	APTOS	96%
M. Nasajpour et al., 2022	Standard, FedAVG, and FedProx	-	92.19%, 90.07%, and 85.81%
Aruna et al., 2020	MobileNet-based architecture	APTOS	95%
Ömer Kasim, 2023	Ensemble Learning	APTOS	88.95%
This study	Hybrid model (stacking ensemble learner + InceptionV3)	APTOS	98.64%

This paper examines various classifiers and their performance when used with various feature extractors, including SVM and SGD with InceptionV3. Strategic preprocessing (data augmentation, SMOTE) improves classification accuracy by addressing the class imbalance. The uneven distribution of diabetic retinopathy severity levels presents difficulties, although careful preprocessing techniques, including data augmentation and standardization, are used. While acknowledging drawbacks, this understanding fosters ongoing development, guaranteeing the model's flexibility and efficacy in healthcare settings.

This hybrid model is remarkable because of its meticulous approach and well-executed hyperparameter tuning. The model's capacity to combine predictions from thirteen different machine learning classifiers significantly increases accuracy. This study improves the field and provides comprehensive insights into the properties and suitability of different classifiers with different feature extractors, paving the way for additional research in therapeutic applications.

5. Conclusions

This work establishes a robust framework for categorizing diabetic retinopathy (DR) by fusing state-of-the-art deep-learning models with traditional machine-learning techniques. Regarding binary and multi-class classification tasks, various classifiers, feature extractors, and data pretreatment methods are used to attain high accuracy. These techniques yield remarkable outcomes. Remarkably, the hybrid model that takes advantage of InceptionV3 specifically attains the highest accuracy of 98.64% for binary classification and 94.95% for multi-class classification. The field of medical image analysis is greatly advanced by this work, which provides a reliable tool for the early identification of DR.

Direction for future work: In our upcoming work, we will investigate other novel deep-learning architectures and include domain-specific expertise, especially with larger datasets. Even though the present model shows encouraging results, confirming that it can be applied to various datasets is crucial. Furthermore, we can merge different smaller-sized datasets to make a larger dataset. Testing on more diverse datasets will improve the model's suitability for real-world scenarios, which will assess the model's resilience and performance under varied conditions.

Data availability

The dataset used in this study is accessible on Kaggle at the following link[1]. Privacy, secrecy, and moral standards are prioritized above everything else in data collection and use.

Funding

The study was planned, executed, and published without the assistance of outside sponsors or funding sources as part of the adherence to ethical standards.

Author Contributions

Regarding author contributions, each author played a specific role in the paper.

Intifa Aman Taifa: Writing—original draft, Research synthesis, Critical analysis, Methodology, Figures, and Tables. **Dr. Tania Islam:** Validation, Supervision, and Incorporating Feedback. **Dr**. **Md Aminul Islam:** Incorporating Feedback and Framework Development. Also, **Md. Mahbub-E-Noor and Dr. Tazizur Rahman**: Framework development, Monitoring Work Progress, and Reviewing manuscript.

Conflict of Interest

The authors declare the absence of conflicts of interest to guarantee openness and honesty in the study process.

6. References

Ting DSW, Cheung GCM, & Wong TY. 2016. Diabetic retinopathy: Global prevalence, major risk factors, screening practices, and public health challenges: a review. *Clinical & Experimental Ophthalmology, 44(4),* 260-277. https://doi. org/10.1111/ceo.12696

Kowluru RA, & Chan PS. 2007. Oxidative stress and diabetic retinopathy. *Experimental Diabetes Research*, Hindawi.

Sun Z, Tang F. Wong R, Lok J, Szeto SKH, Chan J. CK, Chan CKM, Tham CC, Ng DS, & Cheung CY. 2019. OCT angiography metrics predict progression of diabetic retinopathy and development of diabetic macular edema: a prospective study. *Ophthalmology, 126(12),* 1675-1684. https://doi.org/10.1016/j.ophtha.2019.06.016

Yang M-C, Zhu X-B, Wang Y-X, Wu S-L, Wang Q, Yan Y-N, Yang X, Yang J-Y, Chen M-X, Lei Y-H, et al. 2020. Influencing factors for peripheral and posterior lesions in mild non-proliferative diabetic retinopathy—the Kailuan eye study. *International Journal of Ophthalmology, 13(9)*, 1467. https://doi.org/10.18240/ ijo.2020.09.20

Ong JX, Kwan CC, Cicinelli MV, & Fawzi AA. 2020. Superficial capillary perfusion on optical coherence tomography angiography differentiates moderate and severe nonproliferative diabetic retinopathy. *PLoS One, 15(10),* e0240064. https://doi.org/ 10.1371/journal.pone.0240064

Nevill CR, Stratton IM, Maruti SS, Massó-González EL, Sivaprasad S, Bailey C, Ehrlich M., Chong V, & Scanlon PH. 2022. Epidemiology of moderately severe and severe non-proliferative diabetic retinopathy in South West England. *Eye, 36(2),* 433-440. https://doi.org/10.1038/ s41433-021-01471-8

Le NT, Kroeger ZA, Lin WV, Khanani AM, & Weng CY. 2021. Novel treatments for diabetic macular edema and proliferative diabetic retinopathy. *Current Diabetes Reports, 21*, 1-11. https://doi.org/10.1007/s11892-021-01412-5

Clermont AC, & Bursell, S.-E. (2007. Retinal blood flow in diabetes. *Microcirculation, 14(1)*, 49-61. https://doi.org/ 10.1080/10739680601072164

Taifa I. A, Setu, DM, Islam T, Dey SK, & Rahman, T. 2024. A hybrid approach with customized machine learning classifiers and multiple feature extractors for enhancing diabetic retinopathy detection. *Healthcare Analytics*, 100346.

Kar SS, & Maity SP. 2017. Automatic detection of retinal lesions for screening of diabetic retinopathy. *IEEE Transactions on Biomedical Engineering, 65(3)*, 608-618. https://doi.org/10.1109/TBME.2017.2707578

Islam, M. A., Ismail, M., Boyaci, O., Atat, R., & Shannigrahi, S. 2022. Graph Neural Network Based Prediction of Data Traffic in Cyber-Physical Smart Grids. In 2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) (pp. 322-327). IEEE. https://doi.org/10.1109/SmartGridComm52983.2022.9960963

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. 2016. Developing and validating a deep learning algorithm for detecting diabetic retinopathy in retinal fundus photographs. *JAMA, 316(22)*, 2402-2410. https://doi.org/10.1001/jama.2016.17216

Tymchenko, B., Marchenko, P., & Spodarets, D. 2020. Deep learning approach to diabetic retinopathy detection. *arXiv preprint arXiv:2003.02261*.

Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in image classification. *International Journal of Information Technology, pp. 13*, 1–11. https://doi.org/10.1007/s41870-017-0080-1

Denœux, T. 2019. Logistic regression, neural networks, and Dempster-Shafer theory: A new perspective. *Knowledge-Based Systems, pp. 176*, 54–67. https://doi.org/10.1016/j.knosys.2019.03.030

Biau, G., & Scornet, E. 2016. A random forest-guided tour. *Test, pp. 25*, 197–227. https://doi.org/10.1007/s11749-016-0481-7

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. 2003. KNN model-based approach in classification. In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings (pp. 986–996). Springer. https://doi.org/10.1007/978-3-540-39964-3/_62

Rish, I., & others. 2001. An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41–46).

Shafique, R., Mehmood, A., Choi, G. S., & others. 2019. Cardiovascular disease prediction system using extra trees classifier. https://doi.org/10.21203/rs.2.14454/v1

Windeatt, T. 2008. Ensemble MLP classifier design. In Computational Intelligence Paradigms: Innovative Applications (pp. 133–147). Springer. https://doi.org/10.1007/978-3-540-79474-5_6

Chang, C.-C., Lee, Y.-J., & Pao, H.-K. 2010. A passive-aggressive algorithm for semi-supervised learning. In 2010 International Conference on Technologies and

Applications of Artificial Intelligence (pp. 335-341). IEEE. https://doi.org/10.1109/TAAI.2010.61

Gallant, S. I., & others. 1990. Perceptron-based learning algorithms. *IEEE Transactions on Neural Networks, 1(2)*, 179–191.

Peng, C., & Cheng, Q. 2020. Discriminative ridge machine: A classifier for high-dimensional data or imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems, 32(6),* 2595-2609. https://doi.org/10.1109/TNNLS.2020.3006877

Kabir, F., Siddique, S., Kotwal, M. R. A., & Huda, M. N. 2015. Bangla text document categorization using stochastic gradient descent (SGD) classifier. In 2015 International Conference on Cognitive Computing and Information Processing (CCIP) (pp. 1-4). IEEE. https://doi.org/10.1109/CCIP.2015.7100687

Johri, S., Debnath, S., Mocherla, A., Singk, A., Prakash, A., Kim, J., & Kerenidis, I. 2021. Nearest centroid classification on a trapped ion quantum computer. *npj Quantum Information, 7(1),* 122. https://doi.org/10.1038/s41534-021-00456-5

Song, Y.-Y., & Ying, L. 2015. Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry, 27(2),* 130. https://doi.org/10.11919/j.issn.1002-0829.215044

Meena, G., Mohbey, K. K., & Kumar, S. 2023. Sentiment analysis on images using convolutional neural networks based on the Inception-V3 transfer learning approach. *International Journal of Information Management Data Insights, 3(1)*, 100174. Publisher: Elsevier.

Van Dyk, D. A., & Meng, X.-L. 2001. The art of data augmentation. *Journal of Computational and Graphical Statistics, 10(1)*, 1-50. https://doi.org/10.1198/10618600152418584

Shanker, M., Hu, M. Y., & Hung, M. S. 1996. Effect of data standardization on neural network training. *Omega, 24(4)*, 385–397. https://doi.org/10.1016/0305-0483(96)00010-2

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321-357. https://doi.org/10.1613/jair.953

Bergstra, J., Bardenet, R., Bengio, Y., & Kegl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems, 24*.

Emon, M. U., Zannat, R., Khatun, T., Rahman, M., Keya, M. S., & others. (2021). Performance analysis of diabetic retinopathy prediction using machine learning models. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1048–1052). IEEE. https://doi.org/10.1109/ICICT50816.2021.9358612

De Sousa, T. F., & Camilo, C. G. 2023. Deep: Hierarchical Deep Learning Combination for Detection of Diabetic Retinopathy. *Procedia Computer Science, 222*, 425-434. https://doi.org/10.1016/j.procs.2023.08.181

Pavate A., Mistry, J., Palve, R., & Gami, N. 2020. Diabetic retinopathy detection-MobileNet binary classifier. *Acta Sci Med Sci, 4(12)*, 86-91.

Wong, W. K., Juwono, F. H., & Capriono, C. 2023. Diabetic Retinopathy Detection and Grading: A Transfer Learning Approach Using Simultaneous Parameter Optimization and Feature-Weighted ECOC Ensemble. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3301618

Nasajpour, M., Karakaya, M., Pouriyeh, S., & Parizi, R. M. 2022. Federated transfer learning for diabetic retinopathy detection using CNN architectures. In SoutheastCon 2022 (pp. 655-660). IEEE. https://doi.org/10.1109/SoutheastCon48659.2022.9764031

Kasim Ö. 2023. Ensemble classification-based optimized transfer learning feature method for early-stage diagnosis of diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12. https://doi.org/10.1007/s12652-023-04648-z

Džeroski S, & Ženko B. 2004. Is combining classifiers with stacking better than selecting the best one? *Machine Learning, 54*, 255-273. https://doi.org/10.1023/B:MACH.0000015881.36452.6e

Islam MA, Ismail M, Atat R, Boyaci O, & Shannigrahi S. 2023. Software-Defined Network-Based Proactive Routing Strategy in Smart Power Grids Using Graph Neural Network and Reinforcement Learning. *E-Prime-Advances in Electrical Engineering, Electronics and Energy, 5*, 100187. Publisher: Elsevier.

Lahmar, C., & Idri, A. 2023. Deep hybrid architectures for diabetic retinopathy classification. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(2)*, 166-184. https://doi.org/10.1080/21681163.2022.2060864. Publisher: Taylor & Francis.

Alyoubi WL, Abulkhair, M. F., & Shalash, W. M. 2021. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors, 21(11)*, 3704. https://doi.org/10.3390/s21113704. Publisher: MDPI.