



Original Research Paper

Predicting Diabetic Retinopathy and Nephropathy Complications Using Machine Learning Techniques

G.Madhu

Asst.Professor, Department of Computer Science and Engineering
MVSR Engineering College,
Nadergul, Hyderabad-501510
madhu_cse@mvsrec.edu.in

Dr.K.Shyam Sunder Reddy

Professor,
Department of Computer Science and Engineering
MVSR Engineering College
Nadergul,
Hyderabad-501510
shyam_cse@mvsrec.edu.in

K.Padma

Asst.Professor
Department of computer science and engineering
Mvsr engineering College
koutarapupadmavathi@gmail.com

Krishna Dheeravath

Department of CSE (AI&ML)
CVR College of Engineering
Hyderabad, India.
kittu514@gmail.com

Abstract— Diabetes and its sequelae are major problems to worldwide healthcare systems especially Diabetic Retinopathy (DR) and Diabetic Nephropathy (DN). Thus, predictive models are needed for the early diagnosis and intervention. This is a typical problem of imbalanced data and complexity of features interaction in this work. The organized diabetes clinical datasets and APTOS 2019 retinal fundus images were used as publicly available datasets. The preprocessing involved the use of KNN imputer for missing values, outlier detection and handling, MinMax scaling and SMOTE oversampling to balance the data. The following several machine learning models were used for classification: Logistic Regression, Random Forest, XGBoost, LightGBM, CatBoost, Multi-Layer Perceptron, and hybrid ensemble models. Moreover, more complex models were constructed, such as StackingClassifier, as well as ensemble-of-ensemble and image classification models like ResNet50, DenseNet121, Xception, NasNetLarge and ensemble of Xception + DenseNet121. The assessment metrics used are: Accuracy, Precision, Recall, F1-Score, ROC-AUC, RMSE and LogLoss. DenseNet121 achieved the highest classification accuracy of 99.6% on the nephropathy challenge while StackingClassifier Oversampled and LightGBM Oversampled achieved 99.9% and 99.6% accuracy, respectively, for the nephropathy challenge. To obtain the model interpretability, the explainable AI models LIME, SHAP and Grad-CAM were used and the prediction deployment was provided to the user through an explainable interface based on Flask framework, which allowed to obtain accurate, transparent and actionable clinical decision support.

Keywords— Diabetes Mellitus, Diabetic Retinopathy, Diabetic Nephropathy, Machine Learning, Ensemble Models, Predictive Analytics.

I. INTRODUCTION

Diabetes mellitus is one of the most important public health issues all over the world. It is a metabolic disorder that is associated with chronic hyperglycemia due to metabolic dysfunctions that impact many organ systems. The global burden of diabetes has been rapidly growing in the past few decades, and by the early 1980s an estimated 108 million adults had diabetes, rising to over 422 million adults in the

next 50 years [1]. The disease has been associated with significant long-term sequelae like eye problems, kidney failure, cardiovascular diseases and neurological complications causing high morbidity and mortality [2]. The rise in the burden of disease not only affects developed and emerging nations but also leads to substantial health care expenditures and social consequences, necessitating more clinical surveillance and early diagnostic approaches [3].

*Corresponding Author:

849

Received: 15th Jul, 2026 | Revised: 25th Jul, 2026 | Accepted: 5th Aug, 2026 | Available Online: 12th Aug, 2026

(DOI): 10.70102/AEJ.2026.18.3s.103

Diabetic retinopathy and diabetic nephropathy are the most serious consequences of diabetes which can cause irreversible visual loss and chronic kidney failure if not detected in early stages. However, the traditional diagnostic methods are heavily reliant on clinical examination, laboratory tests and expert interpretation, which are often insufficient to detect subtle patterns of illness development [4]. These methods are further hampered by the increasing amount of heterogeneous patient data and the intricate interactions between clinical, demographic and physiological aspects [5]. A number of different analytical methods have been explored, but predictive strategies that have been used typically have been limited to their applicability with a specific population and do not adequately address issues of class imbalance or the changing nature of disease presentation over time [6].

The main objective of this research is to investigate the prediction methods to enhance early diagnosis of DR and DRN using big health data. This covers investigation of existing models of complex analysis that are capable of identifying hidden associations among the clinical manifestations, patient history and diagnostic imaging data. The study aims to assess the consistency, reliability and adaptability of the predicted properties of the study across various clinical contexts and to ensure the interpretability and applicability of the study in clinical practice [7]. In addition, the project also seeks to provide comparative information about the effectiveness of the models for different patient conditions, which will help improve the clinical assessment process and aid in the management and monitoring of patients. [8]

This work is meaningful because it could be used to enhance the accuracy of early diagnosis and help in making clinical decisions in advance to prevent the progression of diabetes-related complications. Enhanced prediction tools may assist health workers in identifying high-risk patients and prompt quicker interventions and treatment planning that can be tailored to each person's needs [9]. Moreover, predicting the use of data-driven prediction algorithms in the healthcare sector could reduce healthcare costs, improve patient health, and enable more extensive disease management initiatives. As more complex analytical technology becomes increasingly adopted in health care systems, these advances are expected to result in more efficient, transparent and patient-centric diabetes management approaches globally [10].

II. RELATED WORK

The application of sophisticated analytical approaches has made considerable advances in predictive healthcare, notably in the detection of chronic disease. Janiesch et al. [11] gave a detailed summary of the various applications of machine learning and DL techniques to process large-scale healthcare datasets and to recognize complex correlation between the variables. Their results revealed the potential of automated learning frameworks for revolutionizing clinical decision-making while also highlighting challenges in interpretability and data quality. Hasan et al. presented an ensemble-based predictive model which reduces the error rate in diabetes prediction and exhibits superior diabetes classification stability and performance over the single classifiers [12]. Their approach did require significant calibration, though, and problems with highly skewed medical data.

In several investigations, the risk of diabetes has been investigated by using structured health data and predictive modelling approaches. Birjaïs et al. developed a machine learning approach to foresee future risk of diabetes by clinical parameters of patients, and the approach achieved promising prediction results and sparked concerns about the generalization of the data to a more diverse patient population [13]. Likewise, Sadhu and Jadli [14] conducted a comparative study of classification techniques for the early diagnosis of diabetes, which showed better diagnostic accuracy when comparing models and pointed out the disadvantage of using clinical characteristics with a high number of dimensions. Xue et al. also advanced the techniques for discovering diabetes progression patterns from health-care data sets, focusing on improved diagnostic performance at the cost of higher complexity and reduced flexibility in dynamic medical settings [15].

Optimization of features and early diagnosis of disease have also been very well studied. Le et al. proposed a wrapper based feature selection method with the assistance of metaheuristic optimization techniques for better prediction accuracy and removing duplicate clinical features [16]. While this made the model more efficient, it was computationally intensive and needed a big training dataset. Shafi and Ansari concentrated on early detection of diabetes using predictive modeling techniques on healthcare decision systems and presented improved disease screening characteristics while highlighting the scaling and real-time implementation problems in clinical environments [17]. Put together, these studies demonstrate the potential of innovative analytical tools in enhancing predictive health care, while also acknowledging ongoing challenges with data heterogeneity and computational expenses.

Further research on the assessment of conventional classification approaches to predicting diabetes and comparative analysis of models. However, in a different study Sisodia and Sisodia assessed various classification algorithms with patient clinical data, and found higher prediction accuracy in various models without using diverse healthcare data sources [18]. Hassan et al. [19] examined classification based prediction systems, that showed satisfactory results in disease identification, but showed a limitation in terms of capturing complex interactions between multi-dimensional patient attributes. Kandhasamy and Balamurali have analyzed the performance of different models of classifiers and discussed the importance of selecting the optimum prediction models for medical diagnostics and presented the performance difference among classifiers on different dataset [20].

However, although great strides have been made, current studies have focused primarily on predictive modeling using a single dataset, relatively few feature integrations and only a few types of complications being evaluated. In addition, many of these frameworks are not interpretable and/or real-time utilizable for clinical decision-support purposes. The present study aims to overcome these challenges by (i) using heterogeneous clinical and imaging data, (ii) assessing the predictive consistency for different diabetes complications and (iii) incorporating interpretability and deployment-oriented frameworks for facilitating reliable and transparent healthcare decisions.

III. MATERIALS AND METHODS

The proposed approach is to accurately forecast the DR and DRN effects using publicly available diabetes clinical and retinal datasets. The framework starts with data ingestion and improving its quality, then moves on to standardizing the preparation processes for model building and evaluation. To further improve the predictive robustness, innovative balancing algorithms were used to address imbalanced class distribution, and feature optimization algorithms were used to keep the most informative clinical features and remove redundancies. For better generalization and prediction variance reduction, we proposed a hybrid Ensemble-of-Ensembles architecture that integrates many complimentary learning models in a hierarchical stacking structure. Furthermore, DL-based image classification architectures were added to capture discriminative features from the retinal fundus images for multimodal complication assessment. XAI methods were integrated for providing the explainability of the model predictions and to reveal the most important clinical and imaging parameters. Furthermore, a light-weight web-based deployment interface has been developed for providing the prediction support in an easy and real-time manner. In addition, a light-weight web-based deployment interface has been created to provide the prediction support easily and in real-time. The integrated system is designed to improve the diagnostic accuracy, reliability, interpretability and clinical usefulness in the early detection of diabetic complications.

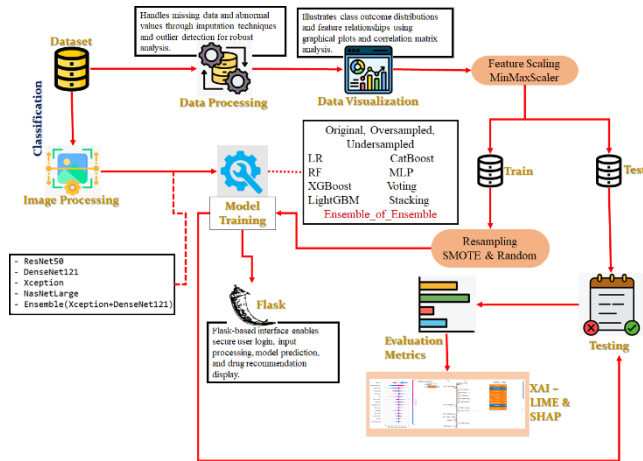


Fig. 1. System Architecture

The proposed system architecture is a comprehensive scheme for intelligent classification from data preparation to visualization, model training, evaluation and deployment. To address class imbalance and noise, raw datasets are preprocessed using resampling and feature scaling. To overcome class imbalance and noise in the raw datasets, resampling and feature scaling are performed. There are several ML and DL models trained and tested with cross validation techniques. Explainability of forecasts is provided by XAI approaches. Last but not least, a Flask-based interface for reading, delivering and displaying the results in real time, ensuring usability, transparency and scalability.

A) Dataset Collection

The Diabetic Retinopathy – Nephropathy dataset is taken from the Kaggle repository, containing 767 clinical records of patients with 22 structured features. Data includes demographic, lifestyle, physiological and biochemical

parameters including insulin use, prescription information, lipid profile, HbA1c, blood pressure and duration of diabetes. It provides binary class labels for two types of DR – Diabetic Retinopathy (DR) and Diabetic Nephropathy (DN). There are some missing values and moderate variances of class distribution in the dataset as in a real clinical case. It provides a very useful representation of many parameters, which makes it very useful for predictive modelling and risk assessment of complications.

	Sex	Age	Diabetes duration (y)	Diabetic retinopathy (DR)	Diabetic nephropathy (DN)	Smoking	Drinking	Height(cm)
0	Male	57	10.0	1	1	1	0	178.0
1	Male	50	8.0	1	1	1	1	172.0
2	Male	53	8.0	1	0	1	0	168.0
3	Male	52	20.0	1	1	0	0	175.0
4	Female	56	12.0	1	1	0	0	159.0

5 rows × 22 columns

Fig. 2. Diabetic Nephropathy v1

The APTOS 2019 dataset was obtained from the Kaggle repository and consists of retinal fundus images that have been captured for automated diagnosis of DR. The Dataset contains high resolution color retinal images annotated for two classes: DR and No DR. This makes supervised image classification possible. It simulates real-world screening with variation in image quality, illumination and severity of the illness. The dataset is highly diverse and clinically validated annotated, suitable for DL based retinal analysis and is efficient for building scalable and automated diabetic retinopathy diagnosis systems.

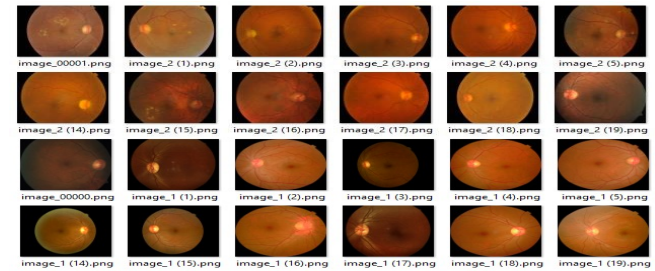


Fig.3 APTOS 2019

B) Pre-Processing

i. Data Pre-processing: To improve the quality of datasets and overcome the issue of missing and inconsistent data, data pre-processing was conducted. Data continuity and information loss were avoided by using missing data imputed by similarity. Any extreme and unusual values were also identified and removed to reduce noise and variations in the data set. This stage guarantees the overall consistency of data and helps create meaningful patterns from reliable clinical data in predictive models, thereby enhancing model stability and convergence and prediction reliability in diabetes complication categorization.

ii. Data Visualization: The features of the datasets were examined and analyzed using data visualization tools to understand the relations between the features prior to model development. Distribution analysis was performed to analyze the ratio of classification outcomes and find the extent of imbalance between complication classes. Interdependencies between clinical factors were also determined by correlation analysis to check the importance of features. This helps in

making educated decisions for data preparation, discovering influential medical indicators and also increases interpretability by giving insights into the relations between diabetic problems and the related patient health features.

iii. Feature Scaling: All variables were represented uniformly using feature scaling and the clinical qualities were scaled to a standard numeric scale. Medical qualities often have different scales and measurement units and therefore normalizing helps to prevent the domination of high valued traits in predictive learning. This step can be very helpful for distance and features sensitive algorithms. By maintaining balanced contributions to features, scaling can enhance optimization efficiency, stabilize learning, and accurately model the relationship between patient health measures and outcomes of diabetic complications.

iv. Data Splitting: The data set was randomly divided into various training and testing portions to objectively assess prediction of performance. The majority was used for training the patterns, and the rest to test their model on unseen patient data. The split guarantees fair assessment of the model and assesses generalization ability to avoid overfitting. Proper data partitioning increases reliability of performance evaluation and provides realistic insight into prediction model performance in real world clinical screening settings.

v. Data Resampling: Data resampling techniques were used to deal with the class imbalance that is typical in medical datasets. To address the rare complication outcome, we oversampled the minority classes to create additional minority classes. To reduce the influence of majority class samples and impart a balance to the entire dataset, undersampling was also carried out. In this way, the increase in prediction fairness, sensitivity to rare medical diseases and reduction in classification bias. Approaches to Class Balance can help the models learn to detect more comprehensive patterns, which in turn will improve detection accuracy for problems of DR and DRN.

C) Image processing:

i. Image Resizing: Standardized the size of the retinal images by downsizing the retinal images to a fixed input size. This step reduces the computational complexity, and enables compatibility with DL architectures. Standardization of picture sizes also enables batch processing and extraction of features with a higher degree of uniformity for all retinal samples.

ii. Image Rescaling: Normalization was performed to rescale the pixel intensity range in the image to a specified numeric range. This normalization eliminates the variance caused by the light variation, and improves the numerical stability during the model training. There are advantages in rescaling in terms of the effectiveness of gradient optimization and whether or not visual retinal information is learned consistently.

iii. Shear Transformation: Retinal images were slightly altered using slightly altering the image geometry (shear transformation). This technique simulates the small differences in picture acquisition angles and the position of the patient's eyes. This enhances the diversity of training and allows the model to be more resilient to the geometric aberrations that occur in real-world retinal pictures.

iv. Zooming the Image: Image zooming is simulated in retinal imaging to simulate the changes in the camera distance

and the focusing level. In this stage, the model is trained with images that are both magnified and reduced, allowing the model to recognize retinal lesions and abnormalities at various scales.

V. Horizontal Flip: The retinal photographs were flipped up and down and side to side to expand the data set. This change adds more diversity to data, assisting the model to learn the spatial invariance. This helps for generalization, as it allows to recognize retinal patterns independently of the image orientation.

vi. Image Reshaping: The image data was reshaped to be formatted to meet the requirements of the input of the deep learning model. This stage assures the uniformity of the channel configuration and the dimensional alignment, which allows for effective processing, better memory use and proper feature extraction in the categorization of retinal images.

D) Algorithms

Diabetic Retinopathy & Nephropathy Models:

Logistic Regression: LR is a simple probabilistic classification technique, that creates linear decision surfaces between the outcome classes. It employs weighted connections between features, to estimate class probabilities, and regularization to reduce overfitting, guaranteeing stable convergence, interpretability and good generalization to varied clinical prediction situations.

Random Forest: RF is a collection of decision trees which create multiple trees using random features and training data to enhance the accuracy of the predictions. This ensemble method can effectively highlight the interactions of the features that are not linear, reduce the variation and increase the stability of the model, and in the context of complex health care prediction scenarios, ensure that the model has high generalization ability and can provide stable predictions.

XGBoost: XGBoost is a gradient boosting framework that works on decreasing residual errors using several decision trees in an iterative manner to improve the predictions. Its regularization algorithms ensure high predictive accuracy, its training process is efficient and it produces consistent results in different categorization scenarios, modeling nonlinear feature interactions.

LightGBM: LightGBM is a highly efficient gradient boosting framework, based on histogram-based learning and leaf-wise tree growing techniques, to achieve faster training speed and higher prediction accuracy. It can effectively capture the non-linear relationship between features, is efficient in computation, memory usage and has good generalization performance for high dimensional and imbalanced classification problems.

CatBoost: CatBoost is a gradient boosting method that is specialized for reducing prediction bias and is based on ordered boosting and symmetric trees. It can effectively address such complex interactions of features, better regularization methods to enhance generalization and stability, and to give reliable probability estimation and performance in various classification domains.

Multi-Layer Perceptron (MLP): A MLP is one class of feed forward neural networks that can learn non-linear

correlations by performing hierarchical transformations of features. It supports increased flexibility of classification and increased prediction accuracy. It extends classical ML models by applying the adaptive learning process and controlled regularization process that learns complex relationships between the input variables.

Voting Classifier: Voting Classifier is an ensemble learning approach that utilizes a number of different classifiers, using either majority or probability voting to get the result. By combining the different predictive models in this ensemble, the effects on bias and stability of the single models are reduced and the overall classification accuracy is enhanced, due to the complementary strengths of the different predictive models.

Stacking Classifier: Stacking Classifier uses an ensemble of base level models with a meta level classifier which uses the predictions of base level models optimally to make the final predictions. This hierarchical ensemble structure will provide inter-model connections, enhance decision boundaries and increase the robustness of the predictions by trading off the bias and variance of various classification algorithms.

Advanced Stacking Classifier (RF, ETC, GBC, SVC, Ridge, XGB → LR): The stacked architecture is very complex, consisting of a multitude of base learners, including bagging, boosting, kernel-based models, linear models and a logistic regression meta-classifier. This is a layered structure of ensembles which optimizes decision fusion, robustness to noise, and generalization in challenging high-dimensional categorization scenarios.

Ensemble of Ensembles: In the Ensemble-of-Ensembles approach, many ensemble models are combined into a single higher order ensemble model. A multi-paradigm, hierarchical fusion of several learning paradigms can help to minimize correlated prediction errors, enhance robustness and stabilize classification – a highly valuable strategy for medical prediction applications where reliability is key.

Image classification Models:

ResNet50: ResNet50 is a deep CNN with residual connections, which enables the efficient training of deep designs. These skip connections increase gradient propagation and feature learning, which help to identify hierarchical spatial patterns and boost classification accuracy in complicated medical image processing tasks.

DenseNet121: In DenseNet121, features are passed to the previous layers in every layer due to dense connection. This architecture facilitates the reusability of features, gradient flow and representational efficiency, which results in accurate identification of minute visual patterns and increase in the classification reliability in diagnostic imaging environments.

Xception: Xception is a depthwise separable convolution (DWSCConv) based CNN, that separates the feature learning into channel-wise and spatial-wise. The architecture reduces the computational complexity of the system while maintaining a high level of representation ability and increases the convergence rate and classification accuracy in high-resolution medical image analysis applications.

NASNetLarge: The neural architecture search (NASNetLarge) is a CNN that automatically optimizes the configuration of the convolutional network by searching the optimal structure for image classification. Its scalable modular design is able to capture rich hierarchical visual data, thus allowing it to achieve high predictive accuracy, and improved generalization and stability in challenging diagnostic imaging contexts.

Ensemble (Xception + DenseNet121): The combination of Xception and DenseNet121 is complementary with both having their own strengths in terms of spatial-channel separation and rich contextual feature learning. This fusion mitigates the bias of individual models, improves the stability of classification and increases the prediction accuracy via learning various visual representation.

IV. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is the ability to correctly classify the patient and healthy patients. To determine the accuracy of a test, the percentage of true positive and true negative should be calculated for all cases it has been applied to. In maths, this is expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Precision: Precision is the ratio of the number of occurrences or samples that are correctly classified to the total number of occurrences or samples classified as positive. So the formula to compute the precision is given by:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (4)$$

Recall: Recall is a metric in ML to quantify how well a model can identify all relevant instances of a certain class. It is the proportion of the number of positive observations that are correctly predicted by the model, to the total number of positive observations, and provides an indication of the ability of a model to capture the cases of a particular class.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F1-Score: In ML, the F1 score is a metric used to measure the accuracy of a model. It offers recall and precision as in a model. The accuracy is the percentage of times that a model is correct for an entire data set.

$$F1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision} * 100 \quad (6)$$

“Table. 1. Performance Evaluation - Image Classification”

ML Model	Accuracy	Precision	Recall	F1_score
ResNet50	0.99545	0.99543	0.95420	0.99545
DenseNet121	0.99636	0.99636	0.99636	0.99636
Xception	0.99545	0.99543	0.99543	0.99545

NASNetLarge	0.98180	0.98229	0.98150	0.98178
Ensemble	0.99545	0.99546	0.99544	0.99545

Many DL models are evaluated and compared in the performance study with standardized metrics. The DenseNet121 model has the best accuracy and a fairly good balance of precision and recall and F1-score, which shows its effectiveness in classification.

“Table.2 Performance Evaluation Table – Nephropathy”

ML Model	Accuracy	Precision	Recall	F1 - Score	ROC - AUC	R MSE	Log Loss
LR Original	0.698	0.400	0.148	0.216	0.697	0.550	0.541
RF Original	0.698	0.375	0.111	0.171	0.619	0.550	0.591
XGBoost Original	0.719	0.500	0.407	0.449	0.676	0.530	0.621
LightGBM Original	0.698	0.450	0.333	0.383	0.611	0.550	0.986
CatBoost Original	0.719	0.500	0.407	0.449	0.645	0.530	0.708
MLP Original	0.729	0.520	0.481	0.500	0.724	0.520	0.915
Voting Classifier Original	0.729	0.529	0.333	0.409	0.696	0.520	0.557
Stacking Classifier Original	0.740	0.556	0.370	0.444	0.722	0.510	0.535
LR Oversampled	0.734	0.728	0.744	0.736	0.788	0.516	0.560
RF Oversampled	0.998	0.999	0.998	0.998	1.000	0.039	0.092
XGBoost Oversampled	0.998	0.999	0.998	0.998	1.000	0.039	0.012
LightGBM Oversampled	0.998	0.998	0.998	0.998	1.000	0.045	0.005

CatBoost Oversampled	0.998	0.998	0.998	0.998	1.000	0.045	0.020
MLP Oversampled	0.986	0.985	0.987	0.986	0.996	0.118	0.061
Voting Classifier Oversampled	0.998	0.998	0.998	0.998	1.000	0.045	0.092
Stacking Classifier Oversampled	0.999	1.000	0.998	0.999	1.000	0.032	0.010
LR UnderSampled	0.620	0.583	0.840	0.689	0.653	0.616	0.693
RF UnderSampled	0.600	0.586	0.680	0.630	0.618	0.632	0.669
XGBoost UnderSampled	0.680	0.655	0.760	0.704	0.650	0.566	0.725
LightGBM UnderSampled	0.700	0.667	0.800	0.727	0.702	0.548	0.763
CatBoost UnderSampled	0.680	0.645	0.800	0.714	0.659	0.566	0.647
MLP UnderSampled	0.540	0.533	0.640	0.582	0.613	0.678	0.730
Voting Classifier UnderSampled	0.660	0.633	0.760	0.691	0.661	0.583	0.643
Stacking Classifier UnderSampled	0.680	0.636	0.840	0.724	0.670	0.566	0.677
Ensemble_of_Ensemble	0.998	0.998	0.999	0.998	1.000	0.039	0.094

The table provides a comparison between different classifiers for the original setting and for the oversampled and undersampled settings, and it is seen that oversampling and ensemble methods significantly improve the accuracy, stability and error metrics, with the exception of the stacking-based ensemble classifiers that perform almost perfectly.

“Table.3 Performance Evaluation Table – Retinopathy”

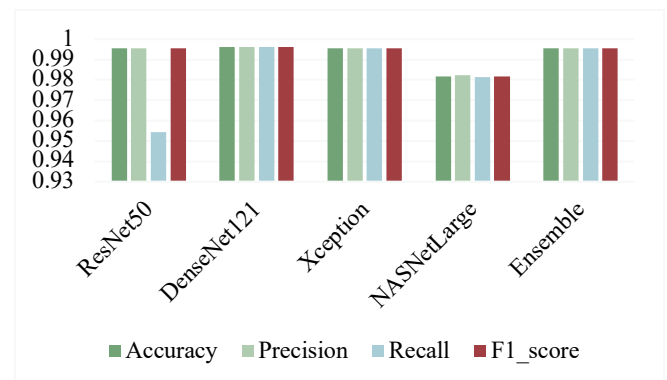
ML Model	Accuracy	Precision	Recall	F1 - Score	ROC - AUC	R MSE	Log Loss
----------	----------	-----------	--------	------------	-----------	-------	----------

				ore	Auc		
LR Original	0.604	0.655	0.404	0.500	0.636	0.629	0.690
RF Original	0.583	0.640	0.340	0.444	0.685	0.645	0.653
XGBoost Original	0.625	0.667	0.468	0.550	0.731	0.612	0.669
LightGBM Original	0.646	0.697	0.489	0.575	0.729	0.595	0.768
CatBoost Original	0.635	0.667	0.511	0.578	0.721	0.604	0.815
MLP Original	0.656	0.684	0.553	0.612	0.641	0.586	0.739
Voting Classifier Original	0.615	0.667	0.426	0.519	0.739	0.621	0.629
Stacking Classifier Original	0.604	0.645	0.426	0.513	0.725	0.629	0.628
LR OverSampled	0.690	0.691	0.685	0.688	0.747	0.556	0.596
RF OverSampled	0.991	0.986	0.996	0.991	1.000	0.095	0.102
XGBoost OverSampled	0.992	0.991	0.994	0.992	1.000	0.087	0.029
LightGBM OverSampled	0.996	0.993	0.998	0.995	1.000	0.067	0.015
CatBoost OverSampled	0.994	0.990	0.998	0.994	1.000	0.077	0.017
MLP OverSampled	0.974	0.967	0.981	0.974	0.994	0.161	0.090
Voting Classifier OverSampled	0.994	0.992	0.995	0.993	1.000	0.081	0.106
Stacking Classifier OverSampled	0.996	0.995	0.998	0.996	1.000	0.059	0.013

LR UnderSampled	0.525	0.520	0.650	0.578	0.539	0.689	0.693
RF UnderSampled	0.625	0.619	0.650	0.634	0.674	0.612	0.646
XGBoost UnderSampled	0.675	0.652	0.750	0.698	0.721	0.570	0.624
LightGBM UnderSampled	0.638	0.622	0.700	0.659	0.661	0.602	0.661
CatBoost UnderSampled	0.600	0.587	0.675	0.628	0.682	0.632	0.672
MLP UnderSampled	0.512	0.512	0.550	0.530	0.542	0.698	2.687
Voting Classifier UnderSampled	0.625	0.609	0.700	0.651	0.649	0.612	0.656
Stacking Classifier UnderSampled	0.612	0.588	0.750	0.659	0.614	0.622	0.684
Ensemble_of_Ensemble	0.996	0.995	0.997	0.996	1.000	0.063	0.100

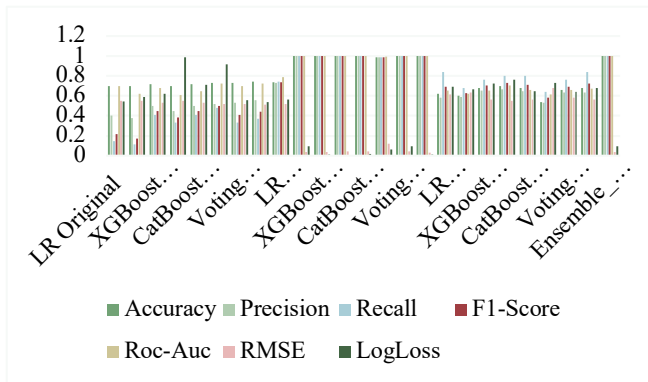
The performance of the original and oversampled/undersampled settings are compared in the table below. The results demonstrate that the oversampling and ensemble based methods attain predictive accuracy, discrimination and error reduction to a significant degree over the baseline models.

Graph.1 Comparison Graph – Image Classification



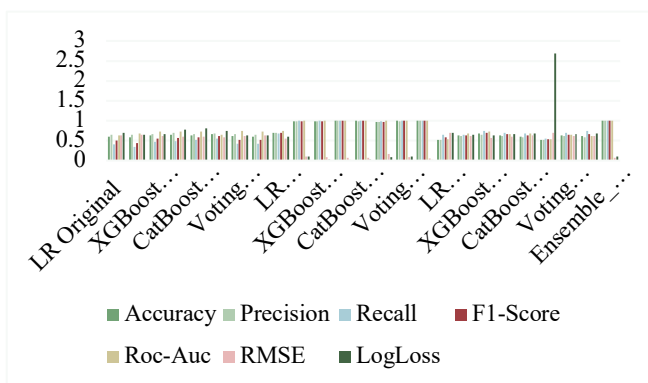
The performance metrics of DL models are compared in the comparison graph .1. Obviously, DenseNet121 and ensemble techniques are highly accurate, precise, recall and F1 score than other designs.

Graph.2 Comparison Graph – Nephropathy



The comparison graph depicts the performance difference between the different classifiers and sampling strategies. The oversampling and ensemble based models are more accurate and have less error metrics, better balanced precision-recall and higher ROC-AUC.

Graph.3 Comparison Graph – Retinopathy



The comparison graph shows error and performance metrics over the classifiers and sampling methods. It indicates that the oversampled approach and ensemble have the lower RMSE and log-loss, and that the classification performance is consistently better.

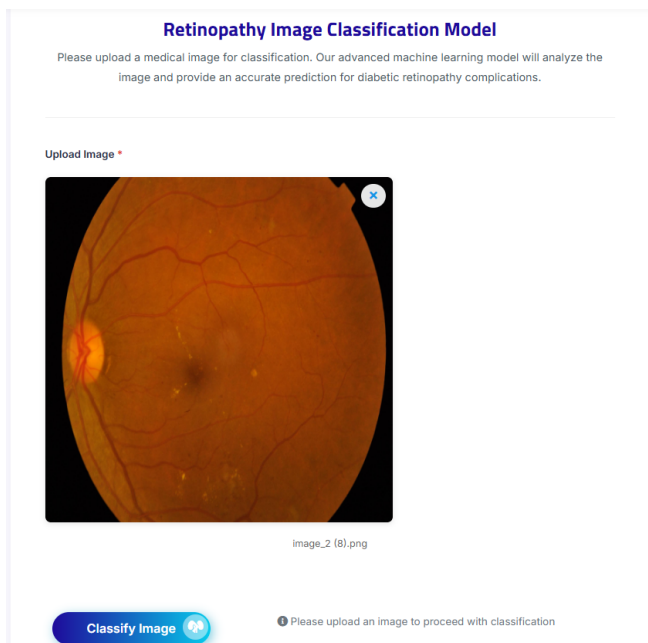


Fig. 4 Upload a input image

The input interface displays an uploaded retinal fundus image for DR analysis, and enables to upload medical photographs and begin automated classification with an interactive and intuitive prediction engine.

Image Classification Results

Detailed analysis of your retinal image classification

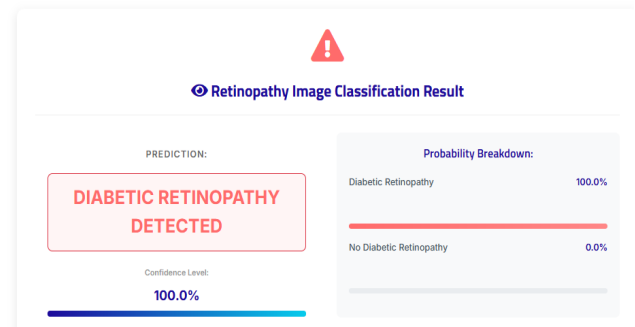


Fig. 5 Predicted Result

The classification output identified diabetic retinopathy at a confidence score of 100.0% which indicated a high-risk retinal condition and demonstrated the model's excellent predictive ability when used in clinical decision support.

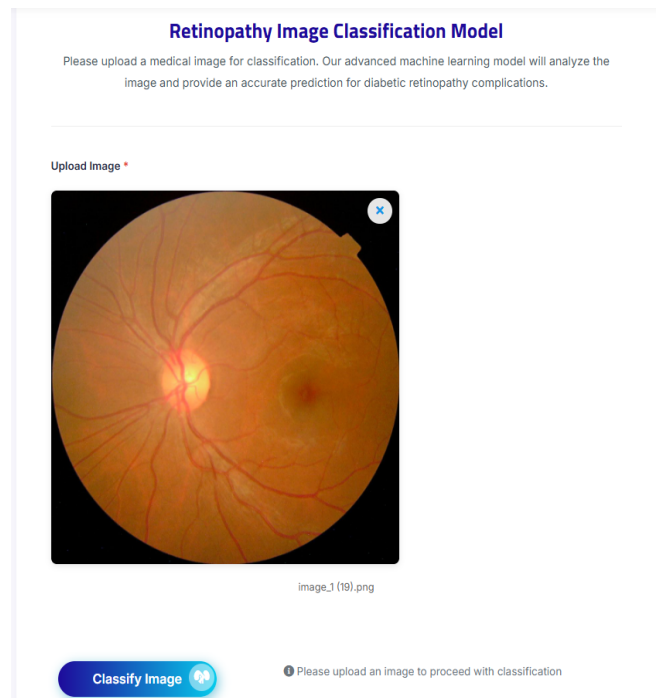


Fig. 6 Upload a input image

The user interface displays an uploaded retinal fundus image for analysis of DR using a simple and interactive classification system, with the user providing a photo of the eye.

Image Classification Results

Detailed analysis of your retinal image classification

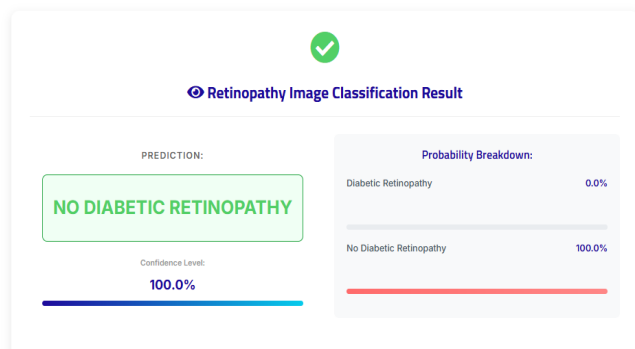


Fig. 7 Predicted Result

The classification output indicates no DR with 100.0% confidence, which indicates a healthy retina and indicates that the classification model is well suited and could be used for accurate ophthalmic screening support.

V. CONCLUSION

Diabetes mellitus and its complications, notably DR and DN have emerged as important challenges for healthcare systems around the world which stress the need for accurate prediction models for early diagnosis and management. The solution is built on publicly available datasets which include organized diabetes clinical data and APTOS 2019 retinal fundus images. The main techniques are: imputation using KNN, dealing with Outliers, MinMax scaling and class balancing through SMOTE. For the advanced models, they developed the following three more types: 1) StackingClassifier 2) Ensemble-of-Ensembles 3) Image classification architectures: ResNet50, DenseNet121, Xception, NasNetLarge, and ensemble of Xception + DenseNet121 To increase interpretability, we used explainable AI techniques such as LIME, SHAP, and Grad-CAM. Because we wanted to make the predictions easily deployable, we have built an interface using Flask. The system performed better with DenseNet121 achieving 99.6% of Accuracy in image classification, StackingClassifier Oversampled got 99.9% of Accuracy in prediction of nephropathy, LightGBM OverSampled got 99.6% of Accuracy in prediction of retinopathy. Overall, the system provides reliable, interpretable, and automated clinical decision support, which enhances diagnostic precision, facilitates informed clinical decisions, and is suitable for scaling up for diabetes related comorbidities.

Multi-modal data integration such as genomic, lifestyle and wearable sensors data enables further refinement of the system, leading to higher prediction accuracy and the development of individual intervention techniques. Real-time streamed data and cloud-based implementation may be the path to scalable and continuous monitoring of diabetes complications. Further improvements can be obtained via sophisticated DL architectures, e.g., vision transformers in image classification and attention-based models in tabular data. The interoperability feature with EHR and HIS would facilitate a streamlined clinical implementation. Furthermore, the expansion of Explainable AI techniques along with interactive visualization tools may enhance model

transparency, clinicians' trust and patient involvement that can help make better educated, data-driven healthcare decisions.

REFERENCES

- [1] Florence, S. (2025). Improving Patient Outcomes: Predicting Diabetic Retinopathy and Nephropathy with DL.
- [2] Li, T., Chen, J., Zhang, X., Wang, K., Zhao, X., Cao, Y., ... & Liang, X. (2025). A machine learning model for predicting the risk of diabetic nephropathy in individuals with type 2 diabetes mellitus. *Frontiers in Endocrinology*, 16, 1587932.
- [3] Wen, Y., Wan, Z., Ren, H., Wang, X., & Wang, W. (2025). Interpretable machine learning model for predicting and assessing the risk of diabetic nephropathy: Prediction model study. *JMIR medical informatics*, 13, e64979.
- [4] Kim, S., Park, J., Son, Y., Lee, H., Woo, S., Lee, M., ... & Rhee, S. Y. (2025). Development and validation of a machine learning algorithm for predicting diabetes retinopathy in patients with type 2 diabetes: algorithm development study. *JMIR Medical Informatics*, 13, e58107.
- [5] Wan, X., Zhang, R., Wang, Y., Wei, W., Song, B., Zhang, L., & Hu, Y. (2025). Predicting diabetic retinopathy based on routine laboratory tests by machine learning algorithms. *European Journal of Medical Research*, 30(1), 1-19.
- [6] H. Sun *et al.*, "IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Res. Clin. Pract.*, vol. 183, Jan. 2022.
- [7] Y. Wu *et al.*, "Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention," *Int. J. Med. Sci.*, vol. 11, no. 11, 2014.
- [8] V. Bellou *et al.*, "Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review," *PLoS ONE*, vol. 13, no. 3, 2018.
- [9] G. Daryabor *et al.*, "The effects of type 2 diabetes mellitus on organ metabolism and the immune system," *Frontiers Immunol.*, vol. 11, 2020.
- [10] S. Firdous *et al.*, "A survey on diabetes risk prediction using machine learning approaches," *J. Family Med. Primary Care*, vol. 11, 2022.
- [11] C. Janiesch *et al.*, "Machine learning and DL," *Electron. Mark.*, vol. 31, 2021.
- [12] M. K. Hasan *et al.*, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, 2020.
- [13] R. Birjais *et al.*, "Prediction and diagnosis of future diabetes risk: A machine learning approach," *Social Netw. Appl. Sci.*, vol. 1, no. 9, 2019.

- [14] A. Sadhu and A. Jadli, "Early-stage diabetes risk prediction: A comparative analysis of classification algorithms," *Int. Adv. Res. J. Sci. Eng. Technol.*, vol. 8, no. 2, 2021.
- [15] J. Xue *et al.*, "Research on diabetes prediction method based on machine learning," *J. Phys.: Conf. Ser.*, vol. 1684, 2020.
- [16] T. M. Le *et al.*, "Wrapper-based feature selection for early diabetes prediction enhanced with metaheuristics," *IEEE Access*, vol. 9, 2021.
- [17] S. Shafi and G. A. Ansari, "Early prediction of diabetes disease using machine learning," in *Proc. Int. Conf. Smart Data Intell.*, 2021.
- [18] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Comput. Sci.*, vol. 132, 2018.
- [19] A. S. Hassan *et al.*, "Diabetes mellitus prediction using classification techniques," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, 2020.
- [20] J. P. Kandhasamy and S. Balamurali, "Performance analysis of classifier models to predict diabetes mellitus," *Procedia Comput. Sci.*, vol. 47, 2015.
- [21] X.-H. Meng *et al.*, "Comparison of data mining models for predicting diabetes," *Kaohsiung J. Med. Sci.*, vol. 29, no. 2, 2013.
- [22] A. Sungheetha and R. Sharma, "Early detection and classification of diabetic retinopathy using CNN," *J. Trends Comput. Sci. Smart Technol.*, vol. 3, no. 2, 2021.
- [23] M. F. Aslan and K. Sabanci, "DL-based diabetes prediction by converting clinical data into image data," *Diagnostics*, vol. 13, no. 4, 2023.
- [24] P. A. Jadhav *et al.*, "Analysis of plasma biomarkers in diabetes patients with retinopathy and nephropathy," *Spectrochimica Acta A*, vol. 304, 2024.
- [25] L. Fregoso-Aparicio *et al.*, "Machine learning and DL predictive models for type 2 diabetes," *Diabetology Metabolic Syndrome*, vol. 13, 2021.
- [26] E. Dritsas and M. Trigka, "Data-driven machine-learning methods for diabetes risk prediction," *Sensors*, vol. 22, no. 14, 2022.
- [27] Y. Su *et al.*, "Age-adaptive diabetes mellitus risk prediction using machine learning," *Biomed. Signal Process. Control*, vol. 80, 2023.
- [28] L. Jiang *et al.*, "Community-based diabetes risk prediction using machine learning," *Preventive Med. Rep.*, vol. 35, 2023.
- [29] S.-J. Hahn *et al.*, "Prediction of type 2 diabetes using polygenic risk scores and metabolic profiles," *EBioMedicine*, vol. 86, 2022.
- [30] Y. Su *et al.*, "Multi-party diabetes mellitus risk prediction using secure federated learning," *Biomed. Signal Process. Control*, vol. 85, 2023.