

Predicting Diabetic Retinopathy And Nephropathy Complications Using Machine Learning Techniques

Ganapa Vyshnavi¹, Dr.Alaparathi Sudhir Babu²

¹Student, Department of Computer Science and Engineering Koneru lakshmaiah Education Foundation,Vaddeswaram,AP ,India. Email: vyshnaviganapa2002@gmail.com

²Professor Department of Computer Science and Engineering Koneru lakshmaiah Education Foundation,Vaddeswaram,AP ,India. Email: asudhirbabu@kluniversity.in

Abstract

Diabetes and its consequences, especially DR and DN are the major problems to the healthcare systems worldwide, demanding reliable prediction models for early diagnosis and intervention. Imbalanced datasets and complicated feature interactions are poorly handled by traditional methods. The Diabetes clinical data and APTOS 2019 retinal fundus photos were organized and publicly available datasets. Preprocessing included KNN imputer for missing values, outlier detection and handling, MinMax scaling and SMOTE oversampling to balance the data. In terms of classification several ML Algorithms were developed, which included various versions of LR, RF, XGBoost, LightGBM, CatBoost, Multi-Layer Perceptron and hybrid ensembles thereof. Also, advanced models such as StackingClassifier and Ensemble-of-Ensembles were constructed, as well as image classification models: ResNet50, DenseNet121, Xception, NasNetLarge and Xception + DenseNet121. Evaluation metrics used were Accuracy, Precision, Recall, F1-Score, ROC-AUC, RMSE and LogLoss. DenseNet121 was the best model in terms of classification with 99.6% Accuracy, while StackingClassifier Oversampled and LightGBM OverSampled were the best models in terms of nephropathy and retinopathy classification with 99.9% and 99.6% respectively. We used explainable AI approaches like LIME, SHAP and Grad-CAM to interpret the model and a user-friendly prediction interface with Flask to enable accurate, transparent and actionable clinical decision assistance.

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

I. INTRODUCTION

Diabetes mellitus is now a serious public health issue in the world. It is a condition that is associated with a continuous rise in blood glucose as a result of metabolic disturbances in multiple organ systems. Diabetes has become a growing problem worldwide; in the early 1980s, approximately 108 million people had this disease, and currently there are over 422 million adults with diabetes, and this figure is expected to rise over the next few decades [1]. The disease can have severe long-term consequences for vision impairment, kidney failure, cardiovascular and neurological diseases, and morbidity and mortality [2]. The increasing burden of disease is affecting developed and developing countries alike, and has important ramifications in terms of healthcare costs and social impacts, which require enhanced clinical surveillance and earlier diagnostic tools [3].

Two of the most severe complications of diabetes can be irreversible sight loss due to diabetic retinopathy and chronic kidney failure due to diabetic nephropathy. Reliable diagnostic methods of traditional periods involved in many comprehensive clinical examinations, laboratory measurements and professional interpretation, but often fail to detect subtle signs of disease development [4]. These techniques are also hampered by the increasing number of heterogeneous patient data and intricate interrelationships between clinical, demographic and physiological aspects [5]. Other analytical approaches have been investigated, however existing predictive approaches often fail to generalise to other populations and do not sufficiently handle class imbalance and dynamic illness presentations [6].

The primary aim of this study is to investigate predictive approaches for early detection of DR and DRN based on vast collections of health care data. This includes investigating the potential of sophisticated analytical models to find associations between clinical signs, patient history and diagnostic imaging data. The study will investigate the anticipated consistency, reliability, and flexibility in various clinical contexts, as well as the interpretability and applicability of the study in healthcare settings [7]. In addition, the aim of the work is to give comparative views on the effectiveness of the models on each patient state, therefore working towards better clinical evaluation methodologies, and better practices of patient care and monitoring [8].

This study is important because it can improve the early diagnosis and encourage early therapeutic decisions, reducing the development of diabetes-related issues. With improved prediction models it would be possible to help health care providers identify individuals at high risk, enabling earlier interventions and individual treatment

plans [9]. Furthermore, the incorporation of data-driven prediction algorithms into medical systems can help to reduce healthcare costs, improve patient health, and facilitate large-scale disease monitoring and management. The use of intelligent analytical technology in healthcare systems is anticipated to help in establishing more effective, transparent, and patient-centered diabetes management strategies around the world [10].

II. RELATED WORK

The application of intelligent analytical approaches has considerably contributed to the development of predictive healthcare, especially for diagnosing chronic diseases. Janiesch et al. have given a comprehensive overview on the applications of machine-learning and deep-learning, highlighting their abilities to process big data sets in healthcare and uncover intricate relationships among variables [11]. Their discoveries indicated that automated learning frameworks can transform healthcare determination making, but raised concerns in regards to interpretability and data quality. Hasan et al. presented the ensemble based predictive system which has many classifiers that increase the accuracy in diabetes prediction and has better stability and performance in the classification task than individual classifiers [12]. But their approach required extensive modifications of the model, and suffered from limitations of processing highly unbalanced medical data sets.

Prescriptive modelling methods for diabetes risk prediction from structured health-care data have been explored in several studies. Birjais et al. have constructed an innovative approach using ML for predicting the future risk of diabetes based on clinical measurements of patients and obtained good prediction results, but raised the issue of data generalisation across patient groups [13]. Likewise, Sadhu & Jadli (2019) compared the classification methods for early detection of diabetes to show how the comparison of the various models can yield greater accuracy in the diagnosis process and the issues associated with handling a high-dimensional clinical variables [14]. Otherwise, Xue et al. also introduced prediction approaches to identify the diabetes progress patterns in healthcare data sets. They noted the need for improved diagnosis accuracy, and reported more complexity in computation and less adaptability to dynamic clinical settings [15].

Furthermore, features optimisation and disease detection have been extensively studied. Le et al. proposed a wrapper-based feature selection strategy that was enhanced by metaheuristic optimisation methods and significantly enhanced prediction accuracy and the removal of redundant clinical features [16]. This approach made the model more efficient, but required significant amounts of computing power and extensive datasets for training. Shafi and Ansari (2018) used prescriptive modelling approach to a healthcare decision system for early detection of diabetes that demonstrated improved disease screening but lacked in scalability and real-time applicability in clinical use [17]. These inquiries underscore the potential of advanced analytical models for enhancing predictive health care solutions and the critical issues of data variety and calculation expenses.

Some more research have presented the classic categorisation strategies for diabetes prediction and comparative analysis of models. By analysing a variety of categorisation techniques, based on clinical patient information, Sisodia and Sisodia demonstrated the effectiveness of various models in predicting accuracy. Their study, however, did not involve integration of various sources of healthcare data [18]. Hassan et al. presented prediction algorithms using categorisation. They were able to perform well for disease identification but admitted that they were unable to represent complex relationships between multi-dimensional patient data. Kandhasamy and Balamurali [20] have performed the performance evaluation of different models of classifiers and emphasized the selection of best predictive model for healthcare diagnostics. They observed that there were differences in the performance of the models on different datasets.

Although significant advances have been made, the available studies largely focus on predictions of models based on a single dataset, limited feature integration, and challenging a select few sets of complications. There are also many frameworks with low interpretability and applicability in real-time clinical decision support scenarios. To address these challenges, we employ diverse clinical and imaging datasets, assess consistency of predictions across various diabetes complications, and incorporate frameworks for interpretability and deployment to enable reliable and transparent healthcare decision-making in this study.

III. MATERIALS AND METHODS

The suggested approach is aimed at reliable prediction of diabetic retinopathy and nephropathy consequences utilising publically available diabetes clinical and retinal information. The framework starts with the data ingestion and quality improvement, followed by the standardised data preparation methods to ensure the reliable model creation and evaluation. To address class imbalance and feature optimization methods were used to ensure that the most useful clinical features are kept and redundant features were dropped, which increased prediction robustness. A new Ensemble-of-Ensembles architecture was proposed, which uses a strong hierarchy of stacking

many different complimentary learning models to enhance generalisation and minimise variance in the prediction. In addition, multimodal classification of the complications was done using deep-learning based image classification architectures that were able to extract discriminative characteristics from the retinal fundus images. Explainable Artificial Intelligence methods were employed to give a transparent explanation of model predictions and identify important clinical and imaging characteristics. Additionally, a user-friendly web-based deployment interface was created to deliver easy and timely prediction support. The integrated system is expected to enhance the accuracy, reliability, interpretability and clinical usability of early diagnosis of diabetic complications.

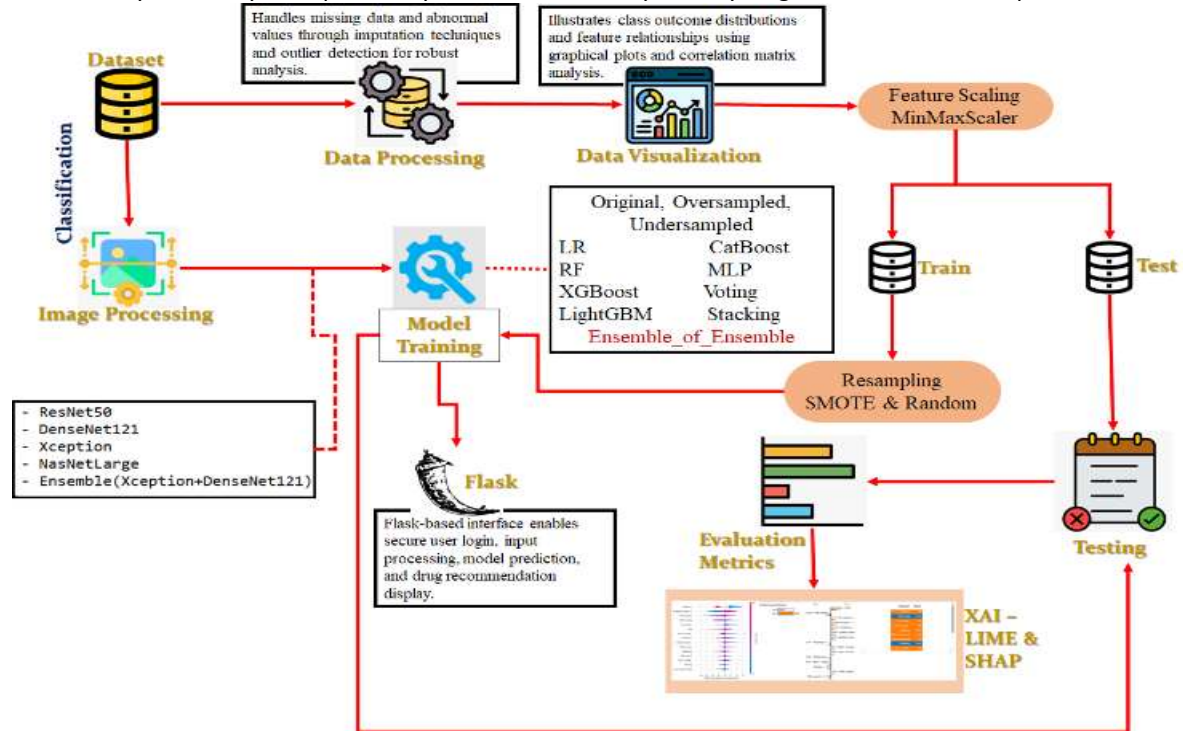


Fig. 1. System Architecture

The system architecture represents an overall intelligent classification system that encompasses data pre-processing, visualization, model training, evaluation and deployment. The raw data is preprocessed, scaled and resampled to handle the noise and imbalance of classes. Cross validation techniques are taught and validated for a number of DL and ML algorithms. In addition, we also include explainability by applying XAI approaches to explain the predictions. Lastly, it is a Flask-based interface, which enables real-time input processing, prediction delivery and result visualisation, ensuring usability, transparency, and scalability.

A) Dataset Collection

The Diabetic Retinopathy–Nephropathy dataset was taken from the Kaggle repository, and it consists of 767 clinical records of patients with 22 structured variables. The dataset contains demographic, lifestyle, physiological and biochemical characteristics such as age, diabetes duration, blood pressure, HbA1c, lipids, use of insulin and medication details. It gives binary class labels for DR and DN. The dataset has some missing values and a moderate variance in the class distribution, which is reflective of the real clinical situations. Being well represented with a number of parameters, it is very useful for forecast and complication risk assessment.

	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

APTOS 2019 is the dataset that was downloaded from the Kaggle repository and it has a number of retinal fundus images collected for the purpose of automated diabetic retinopathy screening. The data is a collection of high-resolution colour retinal images with labels indicating either No DR or DR. This will make supervised image classification easier. It is representative of real-life screening scenarios, differing in image quality, light and disease severity. A diverse dataset and clinically correct annotations are appropriate for DL retinal analysis and are valuable for creating scalable automated DR diagnosis systems.

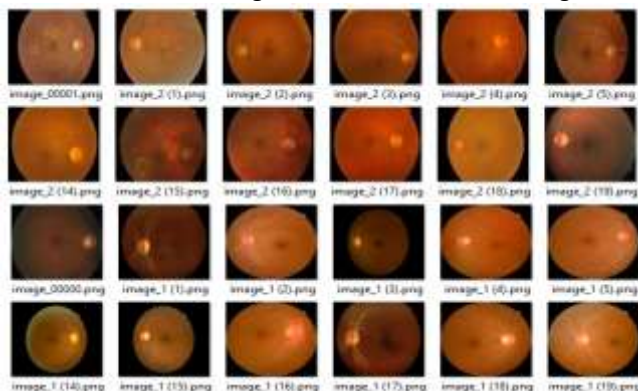


Fig.3 APTOS 2019

B) Pre-Processing

i. Data Pre-processing: Data pre-processing was done to enhance the quality of data by rectifying missing and conflicting data. In an effort to retain data continuity and minimise information loss, similarity-based imputation was used to approximate missing clinical values. Additionally, unusual and extreme observations were detected and adjusted to minimize noise and variability in the data. This helps ensure data consistency, and predictive models can more effectively learn important patterns from reliable clinical data, resulting in improved stability of the model, convergence performance, and prediction accuracy of diabetic complications.

ii. Data Visualization: Data presentation techniques were carried out to visualise the data before the model was built, and enable the understanding of the characteristics of the data and the relationships between the features. The proportion of the classification results was analysed using distribution analysis and the imbalance severity among the complication groups was discovered. Interdependencies among clinical variables were also determined and the feature relevance was assessed by correlation analysis. This stage aids in making educated decisions about data preparation, helps in finding influential medical indicators, and increases interpretability by offering insights into the links between diabetic complications and associated patient health features.

iii. Feature Scaling: All the clinical attributes were scaled to the same range of numbers and all the variables are presented in the same manner using feature scaling. In the case of medical features their magnitudes and measurement units are often different, and normalisation is useful to avoid the high valued characteristics dominating when using predictive learning. This phase is especially helpful when algorithms that are sensitive to feature magnitude and distance estimates. Scaling optimises the optimisation efficiency, enhances the stability of learning, and accurately models the relationship between health indicators of patients and diabetic complication outcomes while maintaining balance in the contribution of features.

iv. Data Splitting: The data set was split into training and testing sets, to see how well the data was predicted. Most of the data was employed to find correlations and patterns among clinical variables and the remaining data was reserved for testing new patient records. In this way we make sure that the evaluation of the model is unbiased and we avoid overfitting by testing its capacity to generalise. Effective data partitioning can help ensure accurate performance assessments and enable realistic insights into the performance of predictive models in real-world clinical screening environments.

v. Data Resampling: Data resampling techniques were used to help with the class imbalance usually observed in medical data. Oversampling was used to obtain more samples of the less frequently occurring complication outcome examples. Undersampling was also used to overcome the dominance of the majority class samples, and to balance the overall distribution of the dataset. This approach increases the predictive fairness, improves the sensitivity to rare medical diseases and decreases the categorisation bias. Balanced representation of the class helps models to learn more comprehensive pattern, leading to improved accuracy of the detection of diabetic retinopathy and kidney diseases.

C) Image processing:

i. Image Resizing: The resizing of images was done to normalise the size of retinal images to a standard input size. By doing so, it can be integrated with DL architectures, reducing the computation burden. Uniformity of

picture size allows to process the whole batch of samples with the same efficiency and to extract the same feature sets from every picture.

ii. Image Rescaling: The image rescaling is done to normalize the intensity value of the image in a fixed range. This normalisation reduces the illumination difference induced noise and enhances the numerical stability in training the model. Rescaling boosts the efficiency of gradient optimisation and allows consistent learning of visual retinal information.

iii. Shear Transformation: Retinal pictures were manipulated by shear transformation, which involved a slight change of the image geometry. This method mimics small variations in the angle of acquisition of the image and the patient's eye position. This enhances the training variability and allows the model to be more resilient to geometric distortions in real-world retinal pictures.

iv. Zooming the Image: To simulate the change of distance and focus in the camera from retina imaging, image zooming was added. This stage trains the model with magnified and decreased views to enhance the model's capabilities in detecting retinal lesions and structural abnormalities at different picture sizes.

V. Horizontal Flip: To enhance the variability of the training set, V. Horizontal Flip was performed which mirrored the retinal images along the vertical axis. This change increases the data variability and makes the model learn the spatial invariant. This improves the generalisation as retinal pattern recognition can be done independently of image orientation.

vi. Image Reshaping: The image data was reshaped in an effort to organise it in an organised way suitable for input into DL models. This stage guarantees the same channel setting and dimension matching that allow fast processing, better memory usage and accurate feature extraction in classifying retinal images.

D) Algorithms

Diabetic Retinopathy & Nephropathy Models:

Logistic Regression: LR is a probabilistic classifier which is used as a baseline and sets up linear decision boundaries between the outcome classes. It calculates class probabilities using weighted links between features and applies regularisation to avoid overfitting, resulting in stable convergence, explainability and good generalisation in various clinical prediction scenarios.

Random Forest: RF creates a number of decision tree models from random subsets of features and random samples of data, and averages the results of their predictions. This ensemble technique is capable of capturing the feature interactions that are nonlinear, minimising the variance and maximising the stability of the ensemble and so achieves outstanding generalisation and stable classification results in complex healthcare prediction scenarios.

XGBoost: XGBoost is an implementation of gradient boosting which sequentially adds better decision tree models to build a model to predict the residual errors in a series of models. It is well known for its excellent predicted accuracy, efficient convergence and consistent performance in various categorisation settings due to its regularisation algorithms, optimised training procedure and representation of nonlinear feature connections.

LightGBM: LightGBM is an optimised gradient boosting framework that uses histogram-based algorithms and leaf-wise tree growth to improve the training speed and accuracy. It is very efficient at modelling complex feature interaction, and is also computationally fast, memory efficient and generalizing well on high-dimensional and unbalanced classification problems.

CatBoost: CatBoost is a gradient boosting method with bias reduction of predictions through ordered boosting and symmetric tree construction method. Its powerful regularisation algorithms boost generalisation and stability, while successfully modelling complicated feature interactions, ensuring dependable probability estimates and consistent performance across various classification settings.

Multi-Layer Perceptron (MLP): MLP is a feedforward neural network which models non-linear relationships through successive feature transformations. It enhances the adaptability of classification, enhances prediction accuracy, and can supplement traditional ML algorithms with adaptive learning and controlled regularization methods.

Voting Classifier: VC integrates the results of many heterogeneous learning models in a majority voting or probability voting manner. The ensemble approach reduces the bias of the individual models, and helps to improve the accuracy of the prediction and the overall classification accuracy, taking advantage of the complementary strengths of the various predictive frameworks.

Stacking Classifier: Stacking Classifier is an aggregation of several base estimators with another meta-estimator. The hierarchical ensemble structure learns the relationships between the models by balancing the bias and variance of the various methods and thus improve the decision boundaries and give more robustness to the prediction.

Advanced Stacking Classifier (RF, ETC, GBC, SVC, Ridge, XGB → LR): This is an advanced stacking architecture used to combine several base classifiers such as bagging, boosting, kernel-based, logistic regression, ridge and

XGB with logistic regression as the final layer. This layered ensemble structure facilitates decision fusion, resistance to noise and generalisation in difficult high-dimensional classification problems.

Ensemble of Ensembles: Ensemble-of-Ensembles method is the extension of the ensemble of models into an ensemble of aggregation models. This hierarchical combination of various learning paradigms is useful for minimizing correlated prediction errors, for increasing the robustness of the predictions and for enhancing the stability of the classifiers and therefore highly beneficial for high-reliability medical prediction applications.

Image classification Models:

ResNet50: ResNet50 is a DNN with residual connections, which helps to train deep networks. These skip connections are helpful in improving the accuracy of classification in challenging applications such as medical image processing, whereas they are useful for gradient propagation as well as feature learning, thereby extracting the hierarchical spatial patterns.

DenseNet121: Dense connection is used in DenseNet121, that is, each layer is directly connected to all other layers (feed-forward). This not only allows for reusing features but also has the advantages of improved gradient propagation, increased representation efficiency, the ability to detect small visual patterns, and improved accuracy of classification in diagnostic imaging applications.

Xception: Xception is a CNN based on Depthwise Separable Convolutions that enable learning of channelwise and spatial features simultaneously. This architecture decreases the computational complexity with the strong representational capacity, and improves the convergence time and classification accuracy for high-resolution medical picture analysis.

NASNetLarge: NASNetLarge is a CNN architecture search based neural network which automatically identifies the most optimal structure of the network for picture categorisation. The approach is scalable and modular, and extracts rich hierarchical visual features, with high accuracy predictive performance, generalisation and good performance in challenging diagnostic imaging scenarios.

Ensemble (Xception + DenseNet121): The ensemble of Xception and DenseNet121 is the complementing feature extraction capability of efficient spatial-channel separation and dense contextual feature learning. This is because this fusion can mitigate individual model bias, improve the stability of classification and improve the accuracy of forecasts in the diversified visual representation learning.

IV. EXPERIMENTAL RESULTS

Accuracy: Accuracy of a test is the ability of the test to discriminate the patient and healthy cases in the right way. To assess the accuracy of a test we need to compute the true positive and true negative in all cases tested. This can be expressed mathematically as:

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

Precision: Precision is the proportion of correct classification of instances or samples that are labelled as positive. So the formula to compute the precision is:

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

Recall: Recall in ML is a measure of how well a model can identify all the examples of a class. It is the ratio of the predictions of positive observations to all the observations in the actual class - yes. It is the ability of a model to identify all the cases (cases in a class) that are relevant to the model in a dataset.

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

F1-Score: F1 score is ML measure of the accuracy of a model. It is a combination of the precision and recall of the model. This accuracy measure is based on the number of times the model is correct for the entire dataset.

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

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

The performance evaluation is carried out by comparing the various DL models with the standard metrics, and the results indicate that DenseNet121 outperforms all the models with respect to overall accuracy and balanced precision, recall, and F1-score, respectively, indicating more effective classification performance.

Table.2 Performance Evaluation Table – Nephropathy

ML Model	Accuracy	Precision	Recall	F1-Score	Roc-Auc	RMSE	LogLoss
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 reports several classifiers performance in both original and oversampled/undersampled cases. It demonstrates that oversampling and ensemble methods greatly improve the accuracy, stability and error metrics, with almost perfect results for the ensemble methods based on stacking.

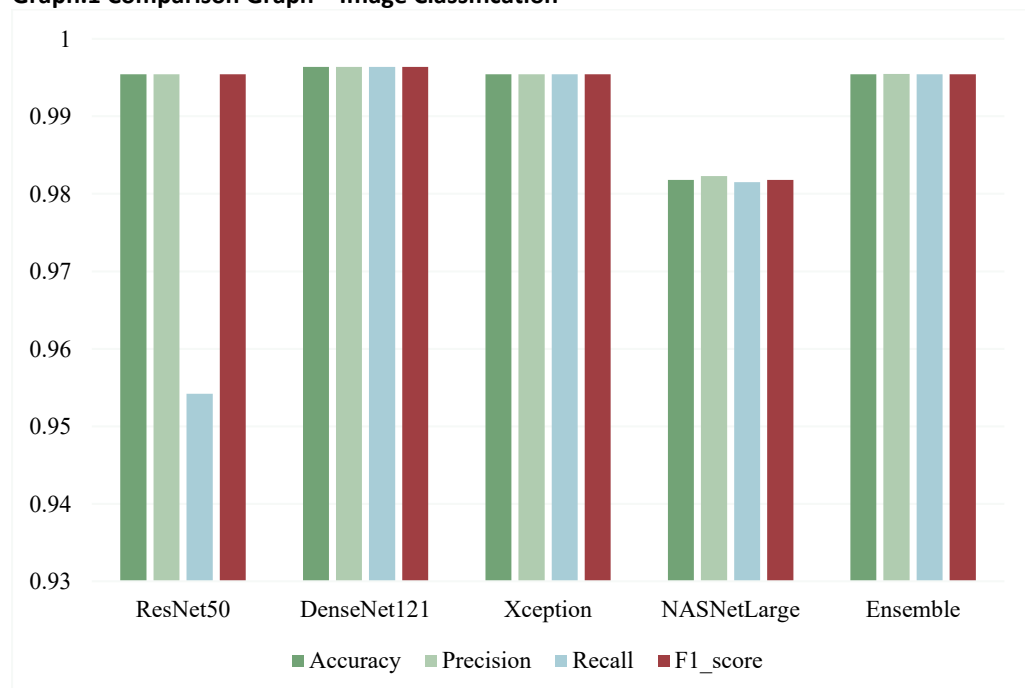
Table.3 Performance Evaluation Table – Retinopathy

ML Model	Accuracy	Precision	Recall	F1-Score	Roc-Auc	RMSE	LogLoss
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

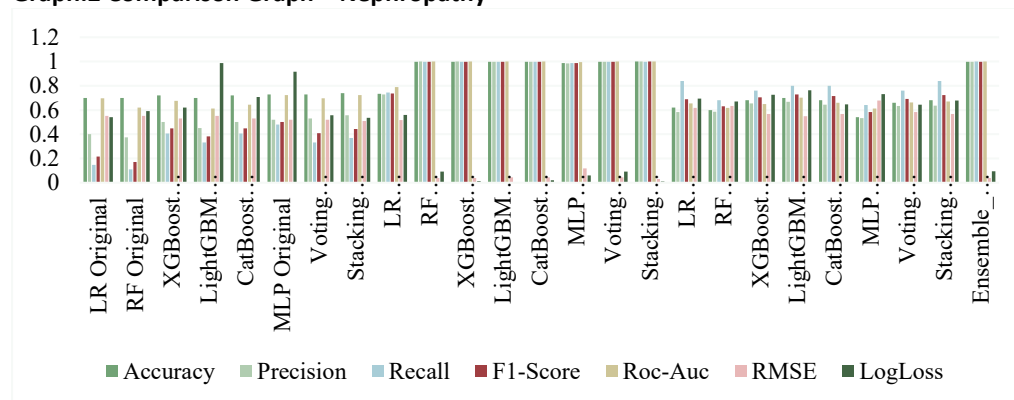
The table shows a comparison of the performance of the original setup to the oversampled and undersampled setups. Obviously, the oversampling and ensemble methods go a long way beyond the baseline models in terms of predictive accuracy, discrimination, and error reduction.

Graph.1 Comparison Graph – Image Classification



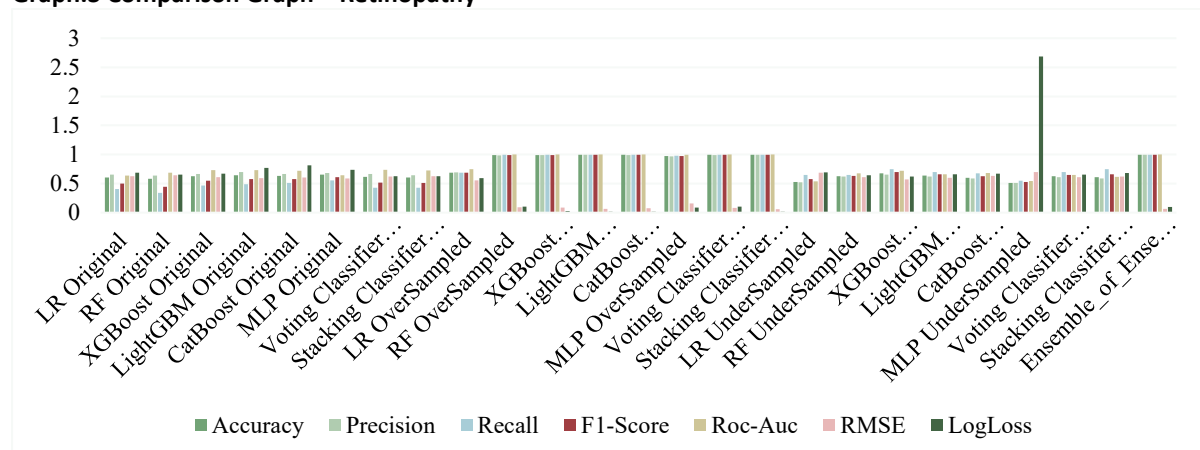
The performance indicators are across DL models in comparison graph.1. DenseNet121 and ensemble techniques consistently achieved superior accuracy, precision, recall and F1-scores compared to the other architectures.

Graph.2 Comparison Graph – Nephropathy



The comparison graph displays the differences in the performance of the classifiers and sampling techniques. The oversampling and ensemble models achieved higher accuracy, performance on the precision-recall curve and higher ROC-AUC, and lower error metrics.

Graph.3 Comparison Graph – Retinopathy



The comparison graph shows that the error and performance metrics of the different classifiers and sampling methods indicate that oversampling and ensemble techniques have better performance in terms of RMSE and log-loss, while providing consistent improvement in classification performance.



Fig. 4 Upload a input image

Uploaded retinal fundus image presented in the input interface for analysis of DR. The users can upload medical photographs and can automatically categorize using an interactive and user-friendly prediction method.



Fig. 5 Predicted Result

The classification result indicates that the model has a high level of predictive potential and that a "high-risk retinal condition" (diabetic retinopathy) is diagnosed with 100.0% certainty.

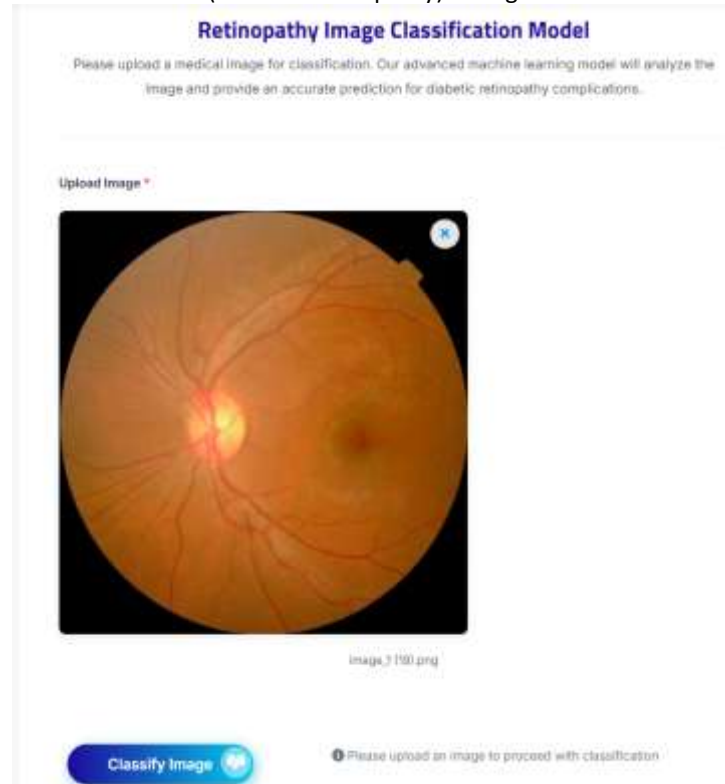


Fig. 6 Upload a input image

Retinal fundus image is displayed in the interface to assess the DR. It allows users to input ophthalmic photos and do automatic analysis with a simple and interactive classification system.



Fig. 7 Predicted Result

The classification result is no diabetic retinopathy with a 100.0% confidence. This means that the retina is healthy and that the model works reliably in providing correct support for ocular screening.

V. CONCLUSION

Diabetes and its complications, such as DR and DN remain a significant burden for health care systems worldwide and are vital to have reliable prediction models to enable early diagnosis and intervention. It uses publicly available datasets, such as organised diabetic clinical data and APTOS 2019 retinal fundus photos. Techniques used primarily are KNN imputation of missing values, outlier detection and treatment, MinMax scaling and SMOTE oversampling for class imbalance. Standard classification models were implemented for various sampling strategies and advanced models were developed further, including StackingClassifier, Ensemble-of-Ensembles, and image classification models, such as ResNet50, DenseNet121, Xception, NasNetLarge, and an ensemble of Xception + DenseNet121. Explainable AI techniques like LIME, SHAP, and Grad-CAM were employed to make the system more interpretable. Predictions were made via an easy-to-use Flask interface. The method performed better on DenseNet121 with 99.6% Accuracy for image classification, 99.9% Accuracy for nephropathy prediction in StackingClassifier Oversampled and LightGBM OverSampled with 99.6% Accuracy for retinopathy prediction. In summary, the system provides a strong, explainable and automated clinical decision support which can enhance the diagnosis accuracy, aid in informed decision making and enable scalable deployment for diabetes related comorbidities.

The approach can be further developed by incorporating multi-modal data such as genomic, lifestyle and wearable sensor data for improved prediction accuracy and personalised intervention strategies. Combining real-time streaming data with cloud-based deployment could provide an opportunity to track diabetes complications on an ongoing and scalable basis. For the image classification, more complex DL models such as vision transformers, and for tabular data, attention-based DL models could produce better results. EHRs and hospital information systems integration would make it easy to achieve clinical uptake. Furthermore, the development of Explainable AI techniques and interactive visualisation tools may lead to a further increase in the transparency of the models, the confidence of the clinicians, and the involvement of the patients in the healthcare decision-making process, which will increase the education and data-driven decision-making in the healthcare sector.

REFERENCES

- [1] Florence, S. (2025). Improving Patient Outcomes: Predicting Diabetic Retinopathy and Nephropathy with Deep Learning.
- [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 deep learning," *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, "Deep learning-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 deep learning 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.