Predicting Diabetic Retinopathy and Nephropathy Complications Using Machine Learning Techniques

M. Thangavel,
Department of Computer Science,
Government Arts College
(Autonomous), Salem, India

A.M. Kalpana
Department of Computer Science and
Engineering, Government College of
Engineering, Dharmapuri, India

S. Ananth,
Department of Artificial Intelligence &
Data Science, Mahendra Engineering
College (Autonomous), Namakkal,
India

***Abstract*— Diabetes and its complications, particularly Diabetic Retinopathy (DR) and Diabetic Nephropathy (DN), pose significant challenges to global healthcare systems, requiring accurate predictive models for early diagnosis and intervention. Traditional approaches often underperform due to imbalanced datasets and complex feature interactions. Publicly available datasets, including structured diabetic clinical data and the APTOS 2019 retinal fundus images, were utilized. Preprocessing involved KNN imputer for missing values, outlier detection and handling, MinMax scaling, and SMOTE oversampling to balance the data. Multiple machine learning models were implemented for classification, including variants of Logistic Regression, Random Forest, XGBoost, LightGBM, CatBoost, Multi-Layer Perceptron, as well as hybrid ensemble approaches. Advanced models were further developed, comprising StackingClassifier, Ensemble-of-Ensembles, and image classification architectures including ResNet50, DenseNet121, Xception, NasNetLarge, and an ensemble of Xception + DenseNet121. Evaluation metrics included Accuracy, Precision, Recall, F1-Score, ROC-AUC, RMSE, and LogLoss. DenseNet121 achieved the highest classification performance with 99.6% Accuracy, while StackingClassifier Oversampled reached 99.9% Accuracy for nephropathy, and LightGBM OverSampled attained 99.6% Accuracy for retinopathy. Explainable AI techniques, including LIME, SHAP, and Grad-CAM, provided model interpretability, and a Flask framework-based interface enabled user-friendly prediction deployment, ensuring accurate, transparent, and actionable clinical decision support.**

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

I. Introduction

Diabetes mellitus has emerged as a major public health challenge worldwide, characterized by persistent hyperglycemia resulting from metabolic dysfunctions that affect multiple organ systems. The global prevalence of diabetes has increased dramatically over recent decades, with the number of affected individuals rising from approximately 108 million in the early 1980s to over 422 million adults, and projections indicating further growth in the coming decades [1]. The condition is associated with severe long-term complications, including visual impairment, kidney dysfunction, cardiovascular diseases, and neurological disorders, which significantly contribute to morbidity and mortality [2]. The increasing disease burden affects both developed and developing nations, creating substantial healthcare expenditures and social consequences that demand improved clinical monitoring and early diagnostic approaches [3].

Among the most severe complications of diabetes are diabetic retinopathy and diabetic nephropathy, which can lead to irreversible vision loss and chronic kidney failure if not identified at early stages. Traditional diagnostic strategies rely heavily on clinical examinations, laboratory measurements, and specialist interpretation, which often fail to capture subtle disease progression patterns [4]. These approaches are further limited by the growing volume of heterogeneous patient data and the complex interrelationships between clinical, demographic, and physiological factors [5]. Although various analytical techniques have been explored, existing predictive methods frequently struggle to generalize across diverse populations and may not adequately address the challenges posed by class imbalance and evolving disease manifestations [6].

The primary objective of this investigation is to analyze predictive methodologies capable of improving the early detection of diabetic retinopathy and nephropathy by leveraging comprehensive healthcare datasets. This includes examining the ability of advanced analytical frameworks to identify underlying relationships among clinical indicators, patient history, and diagnostic imaging information. The investigation aims to evaluate predictive consistency, reliability, and adaptability across multiple clinical scenarios while supporting interpretability and practical usability within healthcare environments [7]. Additionally, the effort seeks to provide comparative insights into model effectiveness under diverse patient conditions, thereby contributing to improved clinical assessment strategies and facilitating better patient management and monitoring practices [8].

The significance of this investigation lies in its potential to enhance early diagnostic accuracy and support proactive clinical decision-making, ultimately reducing the progression of diabetes-related complications. Improved predictive frameworks can assist healthcare professionals in identifying high-risk individuals, enabling timely intervention and personalized treatment planning [9]. Furthermore, the integration of data-driven predictive systems within healthcare infrastructures may reduce medical costs, improve patient outcomes, and support large-scale disease management initiatives. As healthcare systems increasingly adopt intelligent analytical technologies, such advancements are expected to contribute to more efficient, transparent, and patient-centered diabetes management strategies worldwide [10].

II. Related Work

The application of intelligent analytical techniques has significantly advanced predictive healthcare, particularly in chronic disease diagnosis. Janiesch et al. presented a comprehensive overview of machine learning and deep learning applications, highlighting their ability to process large-scale healthcare datasets and extract complex relationships among variables [11]. Their work emphasized the transformative potential of automated learning frameworks in clinical decision-making but noted challenges related to interpretability and data quality. Hasan et al. proposed an ensemble-based predictive framework combining multiple classifiers to improve diabetes prediction accuracy, demonstrating enhanced classification stability and performance compared to individual models [12]. However, their approach required extensive model tuning and showed limitations in addressing highly imbalanced medical datasets.

Several investigations have explored predictive modeling techniques to identify diabetes risk using structured healthcare data. Birjais et al. developed a machine learning-based diagnostic system capable of forecasting future diabetes risk through patient clinical indicators, achieving promising prediction outcomes while highlighting concerns related to data generalization across diverse populations [13]. Similarly, Sadhu and Jadli conducted a comparative evaluation of classification methods for early-stage diabetes detection, demonstrating improved diagnostic accuracy through model comparison while identifying challenges in handling high-dimensional clinical features [14]. Xue et al. further contributed by designing predictive methodologies capable of identifying diabetes progression patterns using healthcare datasets, emphasizing improved diagnostic precision but reporting increased computational complexity and limited adaptability to dynamic clinical environments [15].

Feature optimization and early disease detection have also been widely investigated. Le et al. introduced a wrapper-based feature selection strategy enhanced by metaheuristic optimization techniques, significantly improving prediction accuracy and reducing redundant clinical attributes [16]. Although the approach enhanced model efficiency, it required extensive computational resources and large training datasets. Shafi and Ansari focused on early diabetes detection by integrating predictive modeling techniques within healthcare decision systems, demonstrating improved disease screening capabilities but highlighting concerns regarding scalability and real-time deployment in clinical settings [17]. These investigations collectively demonstrate the potential of advanced analytical frameworks in improving predictive healthcare, while also emphasizing persistent challenges related to data heterogeneity and computational cost.

Further studies have evaluated traditional classification techniques for diabetes prediction and comparative model analysis. Sisodia and Sisodia analyzed multiple classification algorithms using patient clinical data and reported improved prediction accuracy across various models, though their evaluation lacked integration of diverse healthcare data sources [18]. Hassan et al. explored classification-based prediction systems and demonstrated effective disease detection performance but identified limitations in capturing complex relationships among multi-dimensional patient features [19]. Kandhasamy and Balamurali conducted performance analysis across several classifier models and highlighted the importance of selecting optimal predictive frameworks for healthcare diagnostics while noting inconsistencies in model performance across different datasets [20].

Despite substantial advancements, existing investigations primarily focus on single-dataset predictive modeling, limited feature integration, and restricted evaluation across multiple complication types. Additionally, many frameworks lack interpretability and real-time usability within clinical decision-support environments. The present investigation addresses these challenges by incorporating diverse clinical and imaging datasets, evaluating predictive consistency across multiple diabetes complications, and integrating interpretability and deployment-oriented frameworks to support reliable and transparent healthcare decision-making.

III. Materials And Methods

The proposed system focuses on accurate prediction of diabetic retinopathy and nephropathy complications using publicly available diabetic clinical and retinal datasets. The framework begins with data ingestion and quality enhancement, followed by standardized preparation processes to ensure reliable model development and evaluation. To improve predictive robustness, advanced balancing strategies were integrated to address skewed class distributions, while feature optimization techniques were employed to retain the most informative clinical attributes and reduce redundancy. A hybrid Ensemble-of-Ensembles architecture was developed, combining multiple complementary learning models through a hierarchical stacking framework to enhance generalization and reduce prediction variance. In addition, deep learning-based image classification architectures were incorporated to extract discriminative features from retinal fundus images, enabling multimodal complication assessment. Explainable Artificial Intelligence techniques were integrated to provide transparent interpretation of model predictions and highlight influential clinical and imaging features. Furthermore, a lightweight web-based deployment interface was implemented to enable accessible and real-time prediction support. The integrated framework is expected to improve diagnostic accuracy, reliability, interpretability, and clinical usability for early identification of diabetic complications.

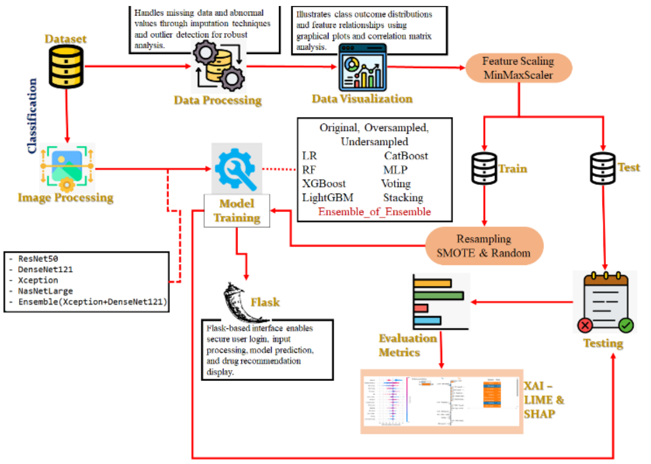


Fig. 1. System Architecture

The system architecture illustrates an end-to-end intelligent classification framework integrating data preprocessing, visualization, model training, evaluation, and deployment. Raw datasets undergo preprocessing, feature scaling, and resampling to handle noise and class imbalance. Multiple deep learning and machine learning models are trained and validated using cross-validation strategies.

Journal of Science Engineering Technology and Management Science *ISSN:3049-0952*
Volume 03, Issue 08, August 2026 *www.jsetms.com*

Explainability is incorporated through XAI techniques to interpret predictions. Finally, a Flask-based interface enables real-time input processing, prediction delivery, and result visualization, ensuring usability, transparency, and scalability.

A) Dataset Collection

The Diabetic Retinopathy–Nephropathy dataset was collected from the Kaggle repository and contains 767 patient clinical records with 22 structured attributes. The dataset includes demographic, lifestyle, physiological, and biochemical features such as age, diabetes duration, blood pressure, HbA1c, lipid profiles, insulin usage, and medication details. It provides binary class labels for Diabetic Retinopathy (DR) and Diabetic Nephropathy (DN). The dataset contains minor missing values and moderate class distribution variations, reflecting real clinical conditions. Its comprehensive multi-parameter representation makes it highly suitable for predictive modeling 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

The APTOS 2019 dataset was obtained from the Kaggle repository and contains retinal fundus images collected for automated diabetic retinopathy screening. The dataset includes high-resolution color retinal images labeled into two classes, Diabetic Retinopathy (DR) and No DR, enabling supervised image classification. It reflects real-world screening conditions with variations in image quality, illumination, and disease severity. The dataset's diversity and clinically validated annotations make it suitable for deep learning-based retinal analysis and effective for developing scalable, automated diabetic retinopathy detection systems.

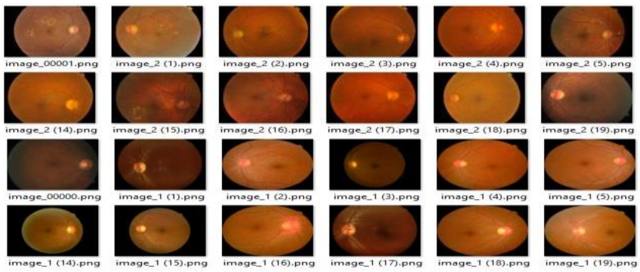


Fig.3 APTOS 2019

B) Pre-Processing

i. Data Pre-processing: Data pre-processing was conducted to enhance dataset quality by addressing incomplete and inconsistent records. Missing clinical values were estimated using similarity-based imputation to preserve data continuity and reduce information loss. Additionally, abnormal and extreme observations were identified and adjusted to minimize noise and variability within the dataset. This step improves overall data consistency and ensures that predictive models learn meaningful patterns from reliable clinical information, thereby enhancing model stability, convergence performance, and prediction reliability for diabetic complication classification.

ii. Data Visualization: Data visualization techniques were employed to explore dataset characteristics and understand feature relationships prior to model development. Distribution analysis was performed to examine the proportion of classification outcomes and identify imbalance severity among complication classes. Correlation analysis was also conducted to evaluate interdependencies between clinical variables and determine feature relevance. This step supports informed data preparation decisions, assists in identifying influential medical indicators, and improves interpretability by providing insights into relationships between diabetic complications and associated patient health attributes.

iii. Feature Scaling: Feature scaling was applied to normalize clinical attributes into a consistent numerical range, ensuring uniform representation across all variables. Since medical features often vary in magnitude and measurement units, normalization prevents dominance of high-valued attributes during predictive learning. This step is particularly beneficial for algorithms sensitive to feature magnitude and distance calculations. By maintaining balanced feature contributions, scaling improves optimization efficiency, enhances learning stability, and supports accurate modeling of relationships between patient health indicators and diabetic complication outcomes.

iv. Data Splitting: The dataset was divided into separate training and testing subsets to evaluate predictive performance objectively. The majority portion was utilized for learning patterns and relationships among clinical attributes, while the remaining portion was reserved for validation using unseen patient records. This separation ensures unbiased model assessment and prevents overfitting by verifying generalization capability. Proper data partitioning improves reliability of performance evaluation and provides realistic insights into how predictive models perform in practical clinical screening scenarios.

v. Data Resampling: Data resampling techniques were implemented to address class imbalance commonly observed in medical datasets. Oversampling was used to generate additional minority class instances, improving representation of less frequent complication outcomes. Undersampling was also applied to reduce dominance of majority class samples, balancing overall dataset distribution. This step enhances predictive fairness, improves sensitivity toward rare medical conditions, and reduces classification bias. Balanced class representation enables models to learn more comprehensive patterns, resulting in improved detection accuracy for diabetic retinopathy and nephropathy complications.

C) Image processing:

i. Image Resizing: Image resizing was performed to standardize retinal image dimensions to a uniform input size. This step ensures compatibility with deep learning architectures and reduces computational complexity. Standardized image dimensions also improve batch processing efficiency and enable consistent feature extraction across all retinal samples.

ii. Image Rescaling: Image rescaling was applied to normalize pixel intensity values within a fixed numerical range. This normalization reduces variation caused by illumination differences and improves numerical stability during model training. Rescaling enhances gradient optimization efficiency and supports consistent learning of visual retinal features.

iii. Shear Transformation: Shear transformation was applied to slightly distort retinal images by shifting image geometry. This technique simulates minor variations in image acquisition angles and patient eye positioning. It increases training diversity and helps the model become more robust to geometric distortions in real-world retinal images.

iv. Zooming the Image: Image zooming was implemented to simulate variations in camera distance and focus levels during retinal imaging. By exposing the model to both magnified and reduced views, this step improves the model's ability to detect retinal lesions and structural abnormalities across different image scales.

V. Horizontal Flip: Horizontal flipping was applied to mirror retinal images across the vertical axis, creating additional training variations. This transformation increases dataset diversity and helps the model learn spatial invariance. It improves generalization by enabling recognition of retinal patterns regardless of image orientation.

vi. Image Reshaping: Image reshaping was performed to organize image data into a structured format compatible with deep learning model input requirements. This step ensures consistent channel configuration and dimensional alignment, enabling efficient processing, improved memory utilization, and accurate feature extraction during retinal image classification.

D) Algorithms

Diabetic Retinopathy & Nephropathy Models:

Logistic Regression: Logistic Regression functions as a baseline probabilistic classifier that establishes linear decision boundaries between outcome classes. It estimates class probabilities using weighted feature relationships and incorporates regularization to prevent overfitting, ensuring stable convergence, interpretability, and reliable generalization across diverse clinical prediction scenarios.

Random Forest: Random Forest enhances predictive robustness by aggregating multiple decision trees constructed using randomized feature subsets and bootstrapped samples. This ensemble mechanism captures nonlinear feature interactions, reduces variance, and improves stability, enabling strong generalization and reliable classification performance in complex healthcare prediction environments.

XGBoost: XGBoost is a gradient boosting framework that sequentially refines predictions by minimizing residual errors across multiple decision trees. Its regularization strategies, optimized training process, and ability to model nonlinear feature relationships contribute to high predictive accuracy, efficient convergence, and consistent performance across varied classification conditions.

LightGBM: LightGBM is an optimized gradient boosting algorithm utilizing histogram-based learning and leaf-wise tree growth to enhance training efficiency and predictive performance. It effectively captures complex feature interactions while maintaining computational speed, memory efficiency, and strong generalization capabilities across high-dimensional and imbalanced classification tasks.

CatBoost: CatBoost is a gradient boosting algorithm designed to reduce prediction bias through ordered boosting and symmetric tree structures. Its advanced regularization techniques improve generalization and stability while effectively modeling complex feature relationships, ensuring reliable probability estimation and consistent performance across diverse classification environments.

Multi-Layer Perceptron (MLP): Multi-Layer Perceptron is a feedforward neural network that captures nonlinear relationships through hierarchical feature transformations. By learning complex dependencies among input variables, it enhances classification flexibility, improves predictive accuracy, and complements traditional machine learning models through adaptive learning and controlled regularization mechanisms.

Voting Classifier: Voting Classifier integrates multiple heterogeneous learning models by aggregating their predictions through majority or probability-based voting. This ensemble strategy reduces individual model bias, enhances predictive stability, and improves overall classification accuracy by leveraging complementary strengths across diverse predictive frameworks.

Stacking Classifier: Stacking Classifier combines multiple base learning models using a meta-level classifier that optimally integrates their predictions. This hierarchical ensemble structure captures inter-model dependencies, refines decision boundaries, and enhances predictive robustness by balancing bias and variance across heterogeneous classification methods.

Advanced Stacking Classifier (RF, ETC, GBC, SVC, Ridge, XGB → LR): The advanced stacking architecture integrates diverse base learners, including bagging, boosting, kernel-based, and linear models, with a logistic regression meta-classifier. This layered ensemble structure optimizes decision fusion, enhances robustness against noise, and improves generalization across complex, high-dimensional classification scenarios.

Ensemble of Ensembles: The Ensemble-of-Ensembles strategy combines multiple ensemble models into a unified higher-order aggregation framework. By integrating diverse learning paradigms, this hierarchical fusion minimizes correlated prediction errors, improves robustness, and enhances classification stability, making it highly effective for high-reliability medical prediction applications.

Image classification Models:

ResNet50: ResNet50 is a deep convolutional neural network utilizing residual connections to enable efficient training of deep architectures. These skip connections improve gradient propagation and feature learning, allowing extraction of hierarchical spatial patterns and enhancing classification accuracy in complex medical image analysis tasks.

DenseNet121: DenseNet121 employs dense connectivity, allowing each layer to receive feature information from preceding layers. This design promotes feature reuse, improves gradient flow, and enhances representational efficiency, enabling accurate identification of subtle visual patterns and improving classification reliability in diagnostic imaging environments.

Xception: Xception is a convolutional neural network based on depthwise separable convolutions, enabling independent spatial and channel-wise feature learning. This architecture reduces computational complexity while maintaining strong representational power, improving convergence speed and classification accuracy in high-resolution medical image analysis tasks.

NASNetLarge: NASNetLarge is a neural architecture search–based convolutional network that automatically identifies optimal structural configurations for image classification. Its scalable modular design captures rich hierarchical visual features, enabling high predictive accuracy, improved generalization, and reliable performance across complex diagnostic imaging conditions.

Ensemble (Xception + DenseNet121): The Xception and DenseNet121 ensemble integrates complementary feature extraction capabilities by combining efficient spatial–channel separation with dense contextual feature learning. This fusion reduces individual model bias, enhances classification stability, and improves predictive accuracy through diverse visual representation learning.

IV. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

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

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

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

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

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

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 compares multiple deep learning models using standard metrics, showing that DenseNet121 achieves the highest overall accuracy and balanced precision, recall, and F1-score, demonstrating superior classification effectiveness.

Table.2 Performance Evaluation Table – Nephropathy

ML Model	Accuracy	Precision	Recall	F1-Score	Roc-Auc	RMSE	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 compares multiple classifiers under original, oversampled, and undersampled settings, demonstrating that oversampling and ensemble strategies substantially improve accuracy, stability, and error metrics, with stacking-based ensembles achieving near-perfect performance.

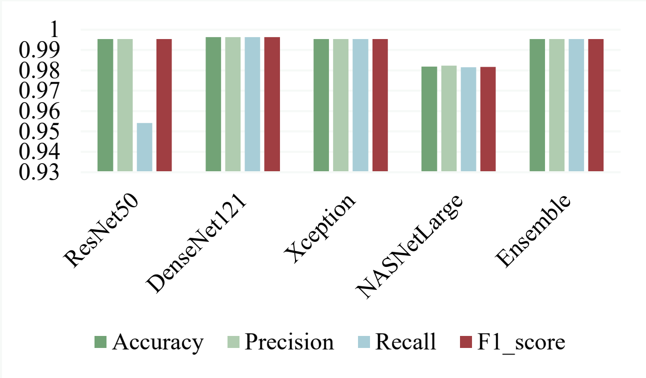
Table.3 Performance Evaluation Table – Retinopathy

ML Model	Accuracy	Precision	Recall	F1-Score	Roc-Auc	RMSE	Log Loss
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

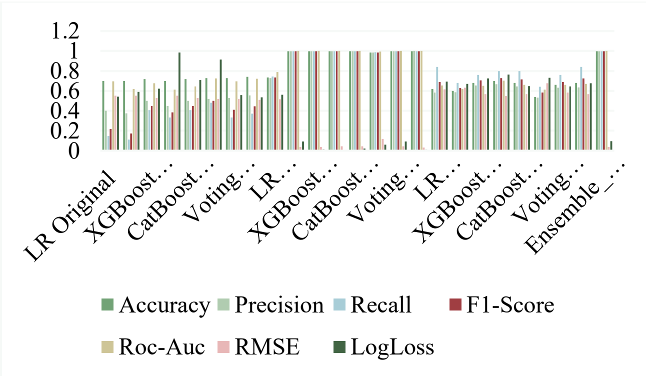
The table presents comparative performance across original, oversampled, and undersampled settings, showing that oversampling and ensemble-based methods significantly enhance predictive accuracy, discrimination, and error reduction compared to baseline models.

Graph.1 Comparison Graph – Image Classification



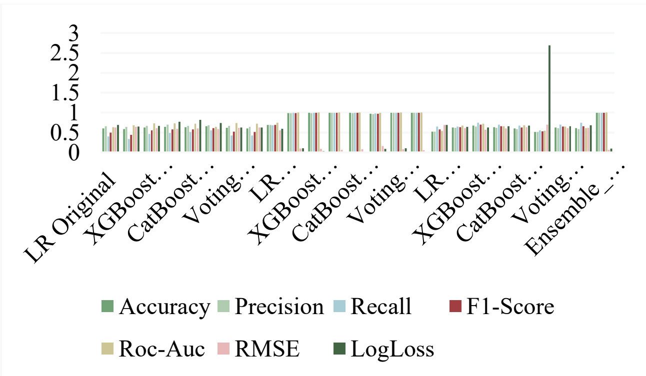
The comparison graph.1 illustrates performance metrics across deep learning models, showing DenseNet121 and ensemble approaches achieving consistently higher accuracy, precision, recall, and F1-scores compared to other architectures.

Graph.2 Comparison Graph – Nephropathy



The comparison graph illustrates performance variations across classifiers and sampling strategies, demonstrating that oversampling and ensemble-based models achieve superior accuracy, balanced precision–recall, higher ROC-AUC, and reduced error metrics.

Graph.3 Comparison Graph – Retinopathy



The comparison graph presents error and performance metrics across classifiers and sampling methods, indicating that oversampled and ensemble approaches yield lower RMSE and log-loss with consistently improved classification performance.

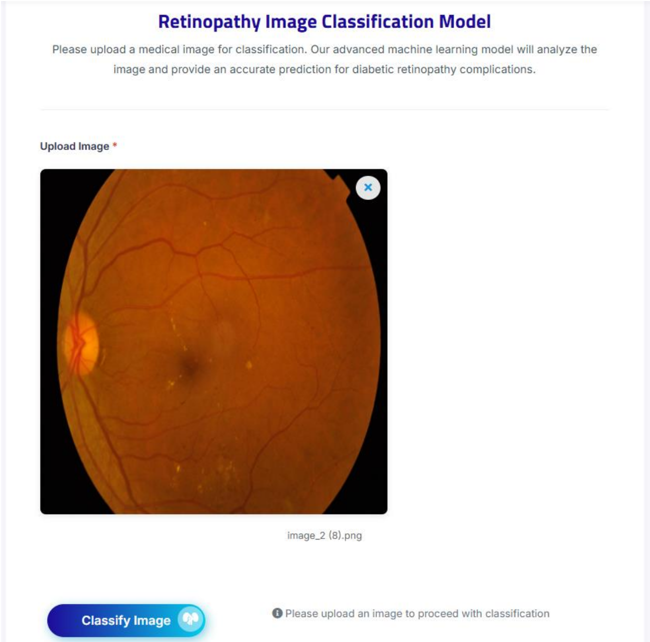


Fig. 4 Upload a input image

The input interface displays an uploaded retinal fundus image for diabetic retinopathy analysis, allowing users to submit medical images and initiate automated classification through an interactive, user-friendly prediction system.

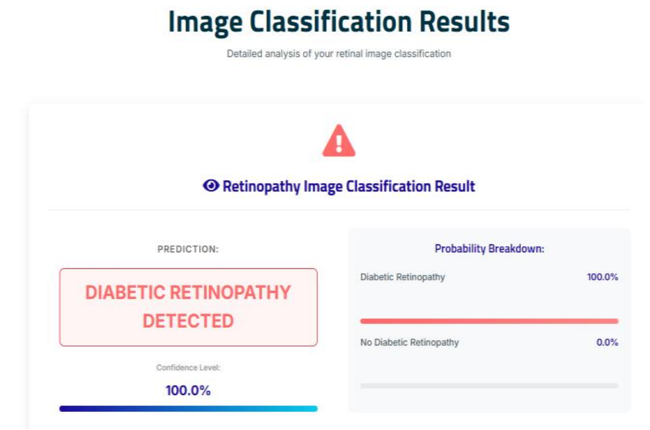


Fig. 5 Predicted Result

The classification output indicates diabetic retinopathy detection with a confidence level of 100.0%, confirming a high-risk retinal condition and demonstrating the model's strong predictive capability for clinical decision support.

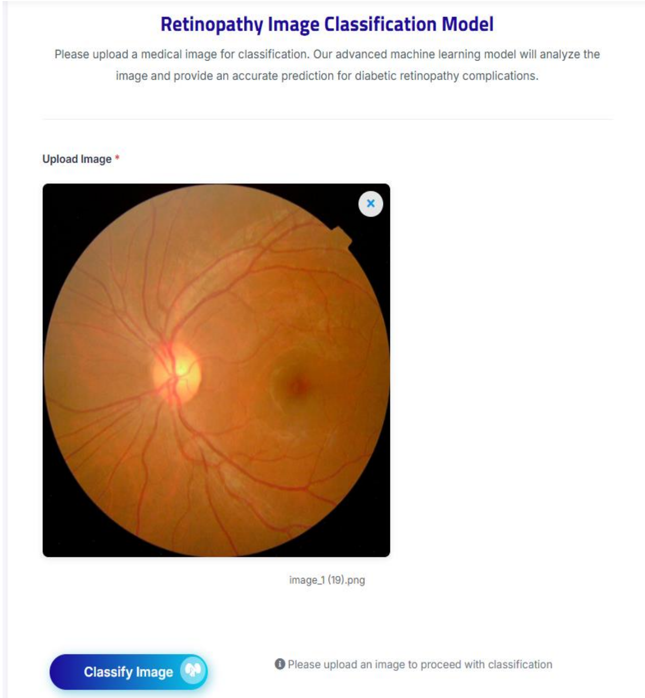


Fig. 6 Upload a input image

The interface displays an uploaded retinal fundus image for diabetic retinopathy assessment, allowing users to submit ophthalmic images and initiate automated analysis through an intuitive and interactive classification system.

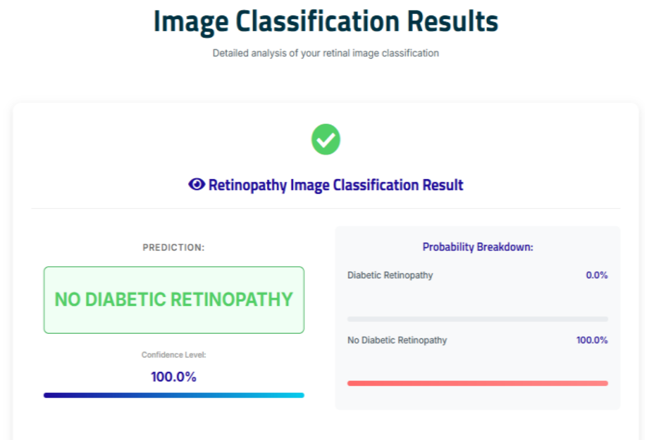


Fig. 7 Predicted Result

The classification output confirms no diabetic retinopathy with a confidence level of 100.0%, indicating a healthy retinal condition and demonstrating reliable model performance for accurate ophthalmic screening support.

V. CONCLUSION

Diabetes and its complications, particularly Diabetic Retinopathy (DR) and Diabetic Nephropathy (DN), present critical challenges to global healthcare systems, emphasizing the need for accurate predictive models to support timely diagnosis and intervention. The system leverages publicly available datasets, including structured diabetic clinical data and the APTOS 2019 retinal fundus images. Key methodologies employed include KNN imputation for missing values, outlier detection and handling, MinMax scaling, and SMOTE oversampling to address class imbalance. While standard classification models were implemented across various sampling strategies, advanced models were further developed, comprising StackingClassifier, Ensemble-of-Ensembles, and image classification architectures including ResNet50, DenseNet121, Xception, NasNetLarge, and an ensemble of Xception + DenseNet121. Explainable AI techniques, including LIME, SHAP, and Grad-CAM, were integrated to enhance interpretability, and a Flask-based interface enabled user-friendly prediction deployment. The system demonstrated superior effectiveness, with DenseNet121 achieving 99.6% Accuracy for image classification, StackingClassifier Oversampled attaining 99.9% Accuracy for nephropathy prediction, and LightGBM OverSampled reaching 99.6% Accuracy for retinopathy prediction. Overall, the system provides reliable, interpretable, and automated clinical decision support, enhancing diagnostic precision, enabling informed decision-making, and supporting scalable implementation for diabetes-related complications.

The system can be further enhanced by integrating multi-modal data, including genomic, lifestyle, and wearable sensor information, to improve predictive accuracy and personalized intervention strategies. Incorporating real-time streaming data and cloud-based deployment could enable scalable, continuous monitoring of diabetic complications. Advanced deep learning architectures, such as vision transformers for image classification and attention-based models for tabular data, may further enhance performance. Integration with electronic health records and hospital information systems would facilitate seamless clinical adoption. Additionally, expanding Explainable AI techniques and interactive visualization tools can improve model transparency, clinician trust, and patient engagement, supporting more informed, data-driven healthcare decision-making.

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.