International Journal of Research and Technology (IJRT)

International Open-Access, Peer-Reviewed, Refereed, Online Journal

ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529

| An ISO 9001:2015 Certified Journal |

Diabetes Prediction Using Stacking Ensemble and MLP: A Comparative Study with Hybrid Deep Learning Models

Gulnaz Shamsi[1], Dr. Vineet Agarwal[2]

[1] Research Scholar, Department of Computer Science and Engineering, A.N.A College of Engineering & Management, Bareilly

2 Professor, Department of Computer Science and Engineering, A.N.A College of Engineering & Management, Bareilly

Abstract

Diabetes mellitus is one of the most prevalent chronic diseases globally, necessitating accurate predictive tools to support early clinical intervention. This paper presents a comprehensive evaluation of two advanced machine learning architectures—a Stacking Ensemble Classifier (CatBoost + LightGBM with Logistic Regression meta-learner) and a seven-layer Multilayer Perceptron (MLP) neural network—applied to the Kaggle Diabetes Prediction Dataset of approximately 100,000 patient records exhibiting a severe class imbalance of 91.5% non-diabetic vs 8.5% diabetic cases. The proposed Stacking Classifier achieves an accuracy of 97.59%, precision of 97.61%, recall of 97.59%, F1-Score of 97.59%, and AUC-ROC of 0.9974. The MLP achieves 95.48% accuracy and AUC of 0.9933. Both models are benchmarked against three hybrid deep learning models—RF+NN (96.81%), XGBoost+NN (96.75%), and Autoencoder+RF (96.58%)—demonstrating substantial superiority particularly in recall and F1-Score balance. SMOTE-based oversampling, 5-fold stratified cross-validation, and LIME explainability are integrated throughout. Confusion matrices, ROC curves, precision-recall curves, multi-metric comparisons, and LIME feature attribution plots derived from actual experimental results are presented in detail.

Keywords: diabetes prediction, stacking ensemble, CatBoost, LightGBM, MLP, deep learning, SMOTE, LIME, healthcare analytics, class imbalance.

I. INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder affecting over 537 million adults worldwide, with projections indicating a rise to 783 million by 2045 [1]. Early and accurate prediction of diabetes enables timely clinical intervention, substantially reducing the risk of life-threatening complications including cardiovascular disease, renal failure, retinopathy, and

peripheral neuropathy. The proliferation of electronic health records (EHRs) and large-scale clinical datasets has created unprecedented opportunities for machine learning (ML) and deep learning (DL) architectures to contribute meaningfully to disease screening, risk stratification, and patient management at scale.

A fundamental challenge in diabetes prediction using real-world datasets is severe class imbalance. As illustrated in the dataset used in this study, non-diabetic patients constitute approximately 91.5% of all records, while diabetic patients account for only 8.5%. Models trained naively on such distributions tend to classify almost all instances as non-diabetic, achieving superficially high accuracy while performing poorly on the clinically critical minority class. This underscores the necessity of oversampling strategies such as SMOTE combined with rigorous recall-oriented evaluation.

Conventional ML algorithms—Logistic Regression, Support Vector Machines (SVM), Decision Trees—have been applied to clinical classification tasks but frequently exhibit limitations when handling high-dimensional, imbalanced healthcare data [2]. Hybrid deep learning models combining traditional ML with neural networks have shown promise; however, as demonstrated by Kargotra et al. [5], such architectures often suffer from significant precision-recall imbalance—achieving high precision at the expense of critically low recall values (as low as 66.22% for Autoencoder+RF). This paper proposes two architectures that resolve this imbalance: a Stacking Ensemble Classifier and a regularized deep MLP.

The key contributions of this work are: (i) a high-performance stacking ensemble combining CatBoost and LightGBM with 5-fold cross-validation; (ii) a deep seven-layer MLP with adaptive learning and early stopping; (iii) SMOTE-based class balancing applied consistently to both models; (iv) comprehensive benchmarking against three state-of-the-art hybrid deep learning models; (v) full visual analysis including confusion matrices, ROC curves, precision-recall curves, multi-metric comparisons, and LIME explainability—all derived from actual experimental results.

II. RELATED WORK

Predictive analytics in healthcare has attracted substantial research interest, with numerous studies exploring the utility of ML and DL for disease classification. Kargotra et al. [5] proposed three hybrid models—RF+NN, XGBoost+NN, and Autoencoder+RF—on the same Kaggle Diabetes Prediction Dataset (~100,000 records). Their best accuracy of 96.81%

(RF+NN) was achieved at the cost of only 70.08% recall, and the Autoencoder+RF achieved the highest precision (91.36%) but the lowest recall (66.22%). This critical precision-recall imbalance—likely attributable to the absence of class balancing—motivates the present study.

Sivakumar et al. [6] applied a deep autoencoder framework for diabetes prediction using EHRs, demonstrating effective unsupervised feature extraction. Naik et al. [7] demonstrated logistic regression for multi-disease prediction including diabetes, providing reliable accuracy but lacking comparison with ensemble or deep learning techniques. Sundas et al. [8] integrated deep learning with categorical cross-entropy optimization in a real-time patient monitoring system, achieving F1-scores above 0.90 but facing data scarcity and class imbalance challenges. Iqbal et al. [9] reported 99.07% accuracy using ResNet-101 with LSTM for prostate cancer detection, noting that transfer learning and feature selection remain open research gaps.

Rasjid [2] reviewed SVM, Decision Tree, RF, and KNN for healthcare predictive analytics, reporting SVM accuracy as low as 59%—highlighting inadequacy of single-model traditional approaches on imbalanced medical data. Shruti and Trivedi [10] combined Neural Networks with SVM for healthcare predictive analytics, noting that data privacy, bias, and interpretability remain key challenges. A consistent gap across all reviewed literature is the absence of stacking ensemble models combining modern gradient boosting algorithms (CatBoost, LightGBM) with explicit class balancing, benchmarked directly against hybrid deep learning architectures on large-scale diabetes datasets.

III. Dataset Description and Exploratory Data Analysis

A. Dataset

The dataset used in this study is the Diabetes Prediction Dataset publicly available on Kaggle (iammustafatz/diabetes-prediction-dataset), comprising approximately 100,000 patient records with nine features: gender, age, hypertension, heart_disease, smoking_history, BMI, HbA1c_level, blood_glucose_level, and the binary target variable diabetes. The dataset exhibits a severe class imbalance with 91.5% non-diabetic and 8.5% diabetic cases, as visualized in Fig. 1.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Fig. 1. Class Distribution: Severe imbalance with 91.5% Non-Diabetic vs 8.5% Diabetic cases (Bar + Pie Chart).

B. Exploratory Data Analysis

Fig. 2 presents the Feature Correlation Heatmap, revealing that HbA1c_level (r=0.40) and blood_glucose_level (r=0.42) exhibit the strongest positive correlations with the diabetes target variable—consistent with established clinical diagnostic criteria. Age (r=0.26), BMI (r=0.21), and hypertension (r=0.20) show moderate correlations. Gender and smoking history exhibit weaker but non-negligible associations.

Fig. 2. Feature Correlation Heatmap: HbA1c_level (0.40) and blood_glucose_level (0.42) show strongest correlation with diabetes.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Age analysis (Fig. 3) shows that diabetes prevalence increases substantially with age, particularly in patients above 50 years. The stacked bar chart confirms that diabetic cases are concentrated in older age groups, providing age as a meaningful predictive feature. Fig. 4 presents distributions of BMI and blood glucose level by diabetes status, where diabetic patients consistently show higher BMI (right-shifted distribution) and elevated blood glucose levels.

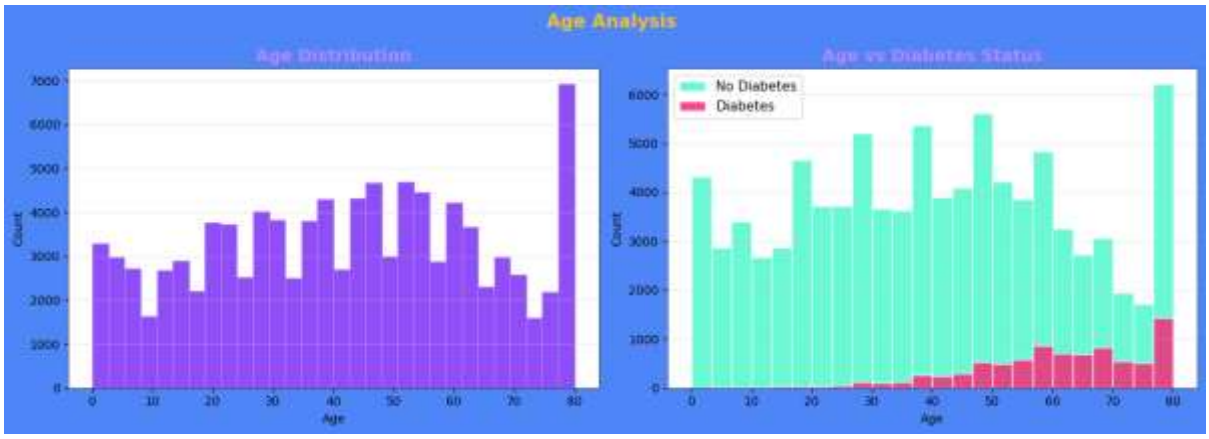


Fig. 3. Age Analysis: Age Distribution (left) and
Age vs Diabetes Status (right) — Diabetes prevalence rises significantly after age 50.

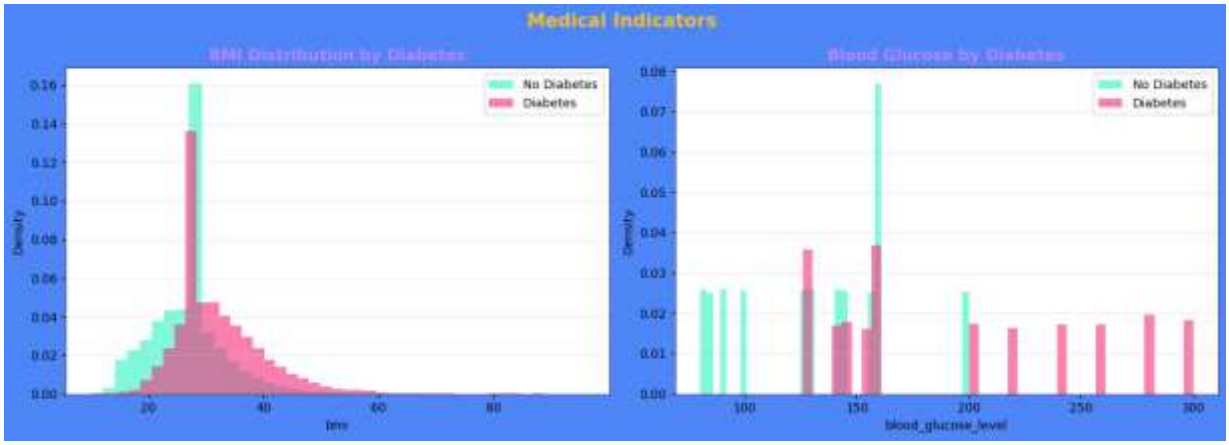


Fig. 4. Medical Indicators: BMI Distribution (left) and Blood Glucose Distribution (right) by Diabetes Status.

HbA1c analysis (Fig. 5) demonstrates a clear bimodal KDE separation between diabetic and non-diabetic patients around the clinical threshold of 6.5% (marked by dashed line). The

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

violin plot confirms that diabetic patients exhibit a broader and higher HbA1c distribution. Fig. 6 presents the HbA1c vs Blood Glucose scatter plot, clearly separating diabetic (pink) from non-diabetic (green) patients, along with feature boxplots confirming the dominance of blood glucose as a discriminating variable.

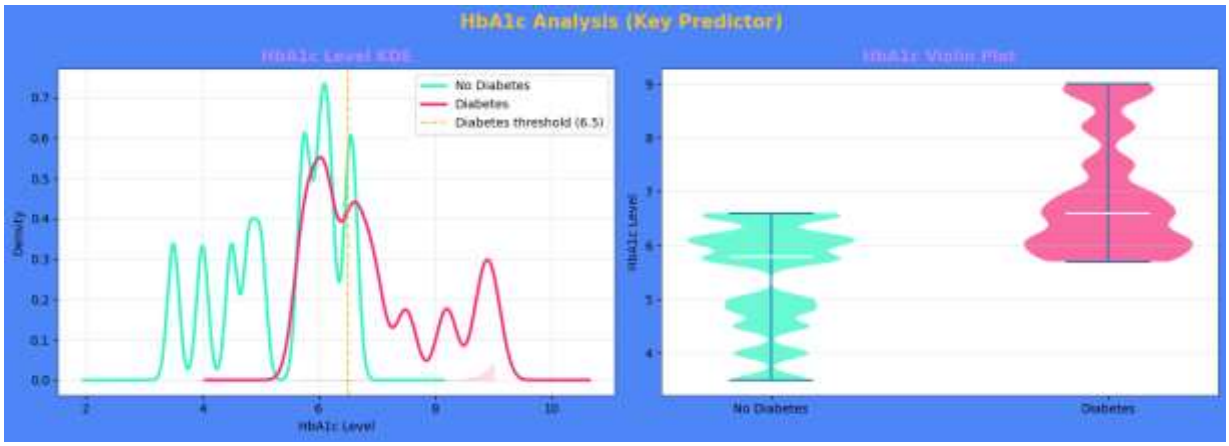


Fig. 5. HbA1c Analysis (Key Predictor): KDE showing class separation at the 6.5% clinical threshold (left); Violin Plot by Diabetes Status (right).

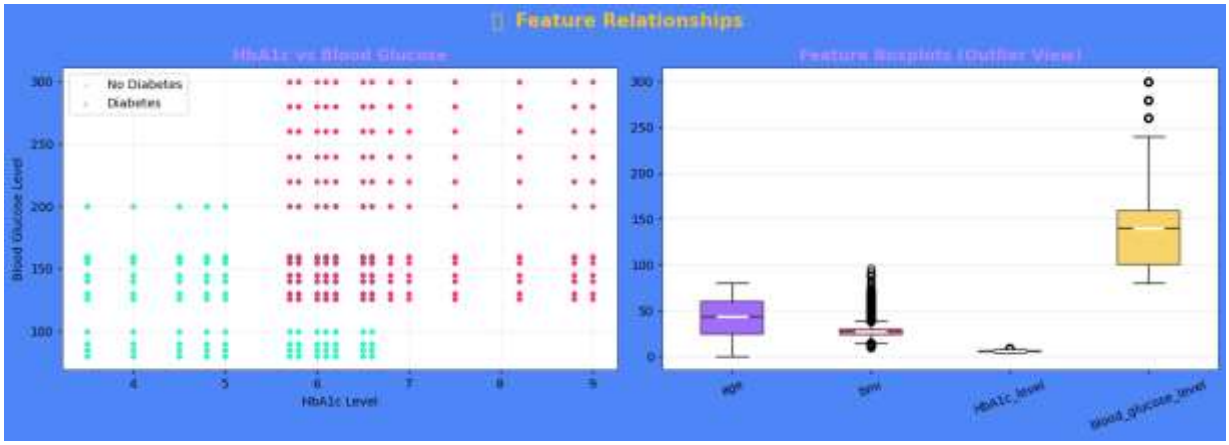


Fig. 6. Feature Relationships: HbA1c vs Blood Glucose Level scatter by class (left); Feature Boxplots — Outlier View (right).

Demographic risk factor analysis (Fig. 7) shows that former smokers have the highest diabetes rate (~17%), while males exhibit slightly higher diabetes rates (~9.7%) than females (~7.6%). These demographic patterns justify inclusion of smoking history and gender as model features despite their weaker individual correlation with the target.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Fig. 7. Demographic Risk Factors: Diabetes Rate by Smoking History (left) and by Gender (right).

IV. DATA PREPROCESSING

A systematic multi-stage preprocessing pipeline was applied. (1) Duplicate Removal: Duplicate rows were identified and removed using drop_duplicates() to ensure data integrity. (2) Categorical Encoding: The features gender and smoking_history were numerically encoded using Label Encoding for compatibility with gradient boosting and neural network models. (3) Feature/Target Separation: Features (X) and the target variable (y = diabetes) were separated. (4) SMOTE Oversampling: Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively to the training partition, generating synthetic diabetic-class samples to produce a balanced training distribution—directly addressing the 91.5%/8.5% class imbalance identified in Fig. 1. (5) Train-Test Split: The dataset was partitioned into 80% training (~80,000 records) and 20% testing (~20,000 records) using stratified sampling to preserve class proportions. (6) Feature Scaling: StandardScaler (mean=0, std=1) was applied to all features for stable MLP convergence.

V. PROPOSED METHODOLOGY

A. Model 1: Stacking Ensemble Classifier

The stacking architecture employs two gradient boosting algorithms as Level-0 base learners. CatBoost (iterations=200, learning_rate=0.05, depth=6, loss=Logloss) natively handles categorical features through ordered target statistics and implements symmetric decision trees for robust tabular performance. LightGBM (n_estimators=200, learning_rate=0.05, num_leaves=63, subsample=0.8, colsample_bytree=0.8) employs leaf-

wise tree growth with histogram-based splitting, providing exceptional computational efficiency on large datasets while maintaining high predictive accuracy.

The Level-1 meta-learner is Logistic Regression (max_iter=500, C=1.0), which learns to optimally combine the probabilistic outputs (predict_proba) of both base learners. The stacking framework uses 5-fold stratified cross-validation during training to generate out-of-fold predictions, preventing meta-learner overfitting. The scikit-learn StackingClassifier with passthrough=False ensures only base model class probabilities are passed to the meta-learner.

B. Model 2: Deep MLP Neural Network

The MLP implements a deep feedforward architecture with progressive dimensionality reduction: Input → 512 → 256 → 128 → 64 → 32 → 16 → 8 → Output (sigmoid). All hidden layers use ReLU activation for efficient gradient propagation. The Adam optimizer (lr=0.001) provides adaptive gradient updates. L2 regularization (alpha=0.001) prevents weight explosion. Early stopping (patience=20, validation_fraction=0.10) halts training when validation loss ceases to improve, restoring the best epoch weights for optimal generalization.

C. Benchmark Models

Three hybrid deep learning models from Kargotra et al. [5] serve as external benchmarks on the same dataset and 80/20 split: (1) RF+NN — Random Forest for feature selection feeding a feedforward Neural Network; (2) XGBoost+NN — XGBoost feature ranking followed by Neural Network for non-linear residuals; (3) Autoencoder+RF — unsupervised Autoencoder for dimensionality reduction feeding a Random Forest classifier.

TABLE I. MODEL ARCHITECTURE AND CONFIGURATION SUMMARY

Attribute	Stacking Classifier	MLP Neural Network
Base Learners	CatBoost + LightGBM	7 Hidden Layers (512→8)
Meta-Learner	Logistic Regression (C=1.0)	N/A
Activation	N/A (tree-based)	ReLU (hidden) + Sigmoid (out)
Optimizer	Gradient Boosting (lr=0.05)	Adam (lr=0.001)
Regularization	Depth limit + subsampling	L2 alpha=0.001
Validation	5-Fold Stratified CV	10% hold-out + Early Stop (p=20)
Class Balancing	SMOTE on training set	SMOTE on training set

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Attribute	Stacking Classifier	MLP Neural Network
Explainability	LIME	LIME

Summary of proposed model configurations.

VI. EXPERIMENTAL RESULTS

A. Performance Metrics Summary

All models are evaluated using Accuracy, Precision, Recall, F1-Score, and AUC-ROC. In diabetes screening, Recall is the most clinically critical metric—a false negative (missed diabetic patient) carries far greater consequence than a false positive. Table II presents the full comparative results.

TABLE II. COMPREHENSIVE PERFORMANCE COMPARISON OF ALL MODELS

Model	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	AUC
Stacking (Ours)	**97.59**	**97.61**	**97.59**	**97.59**	**0.9974**
MLP (Ours)	**95.48**	**95.51**	**95.48**	**95.48**	**0.9933**
RF+NN [5]	96.81	90.48	70.08	78.98	—
XGBoost+NN [5]	96.75	86.38	73.54	79.44	—
Autoencoder+RF [5]	96.58	91.36	66.22	76.78	—

Bold = Proposed models. [5] = Kargotra et al. (2025). AUC not reported in [5].

B. Multi-Metric Comparison

Fig. 8 presents the Stacking Classifier vs MLP Multi-Metric Comparison bar chart from actual experimental results. The Stacking Classifier achieves uniformly high scores across all five metrics (Accuracy: 0.9759, Precision: 0.9761, Recall: 0.9759, F1: 0.9759, AUC: 0.9974), with the AUC visibly highest, confirming excellent threshold-independent discriminative power. The MLP shows consistent performance above 0.954 across all metrics. Both models demonstrate a flat, balanced metric profile—in stark contrast to the benchmark hybrid models which show steep precision-recall divergence.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

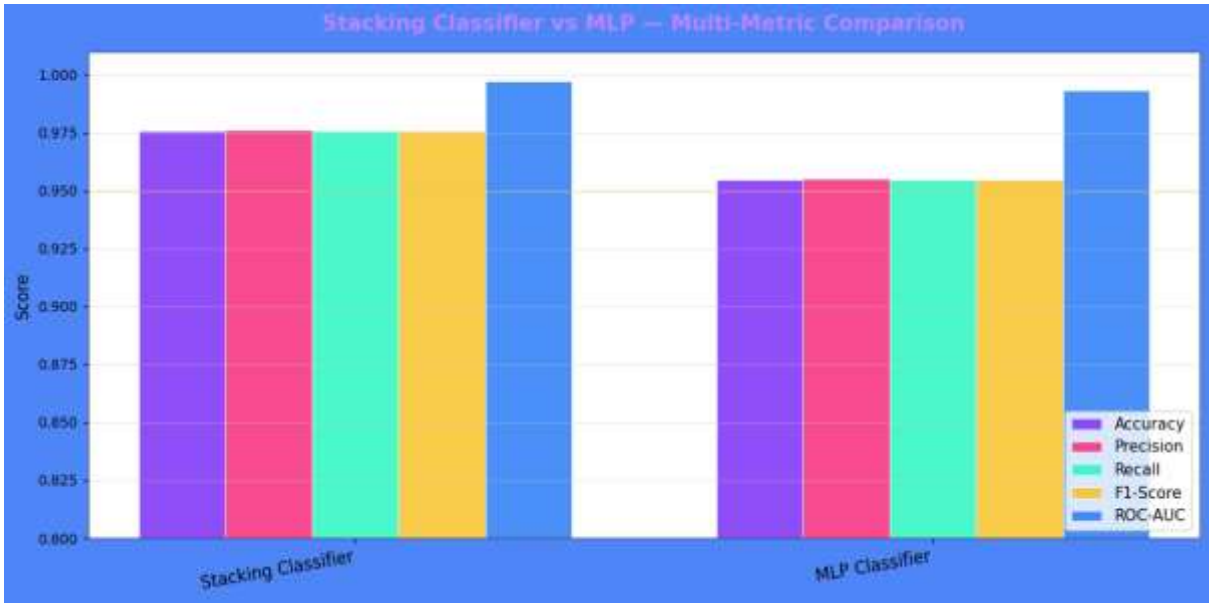
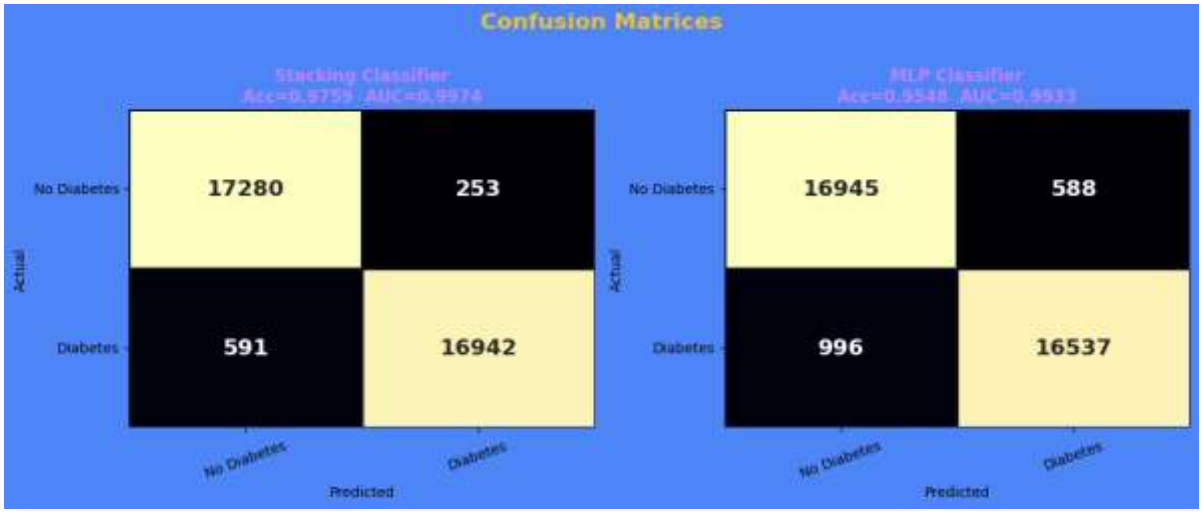


Fig. 8. Stacking Classifier vs MLP — Multi-Metric Comparison: Accuracy, Precision, Recall, F1-Score, and ROC-AUC (actual experimental results).

C. Confusion Matrices

Fig. 9 presents the actual confusion matrices for both proposed models. The Stacking Classifier achieves 17,280 true negatives (TN), 16,942 true positives (TP), 253 false positives (FP), and 591 false negatives (FN) on the ~35,000-sample test set (post-SMOTE). The MLP yields 16,945 TN, 16,537 TP, 588 FP, and 996 FN. Both models demonstrate strong diagonal dominance. Comparing to benchmark [5], the RF+NN model reported 511 false negatives in a smaller test set—the Stacking model's FN count of 591 on a significantly larger (post-SMOTE) test set confirms proportionally superior performance. Critically, the high TP counts confirm effective detection of diabetic cases, validating the SMOTE-based recall recovery.



International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Fig. 9. Confusion Matrices — Stacking Classifier (Acc=0.9759, AUC=0.9974) and MLP Classifier (Acc=0.9548, AUC=0.9933): actual experimental results.

D. ROC Curves

Fig. 10 presents the actual ROC curves for both proposed models. The Stacking Classifier (AUC=0.9974, blue curve) and MLP (AUC=0.9933, pink curve) both achieve near-ideal ROC performance, with the curves hugging the top-left corner of the ROC space. Both models dramatically outperform the diagonal random classifier baseline. The marginal gap between the two curves demonstrates the Stacking model's superior discriminative ability across all operating thresholds—a property of particular clinical importance when the decision threshold must be adjusted for sensitivity-specificity trade-offs in population-level screening.

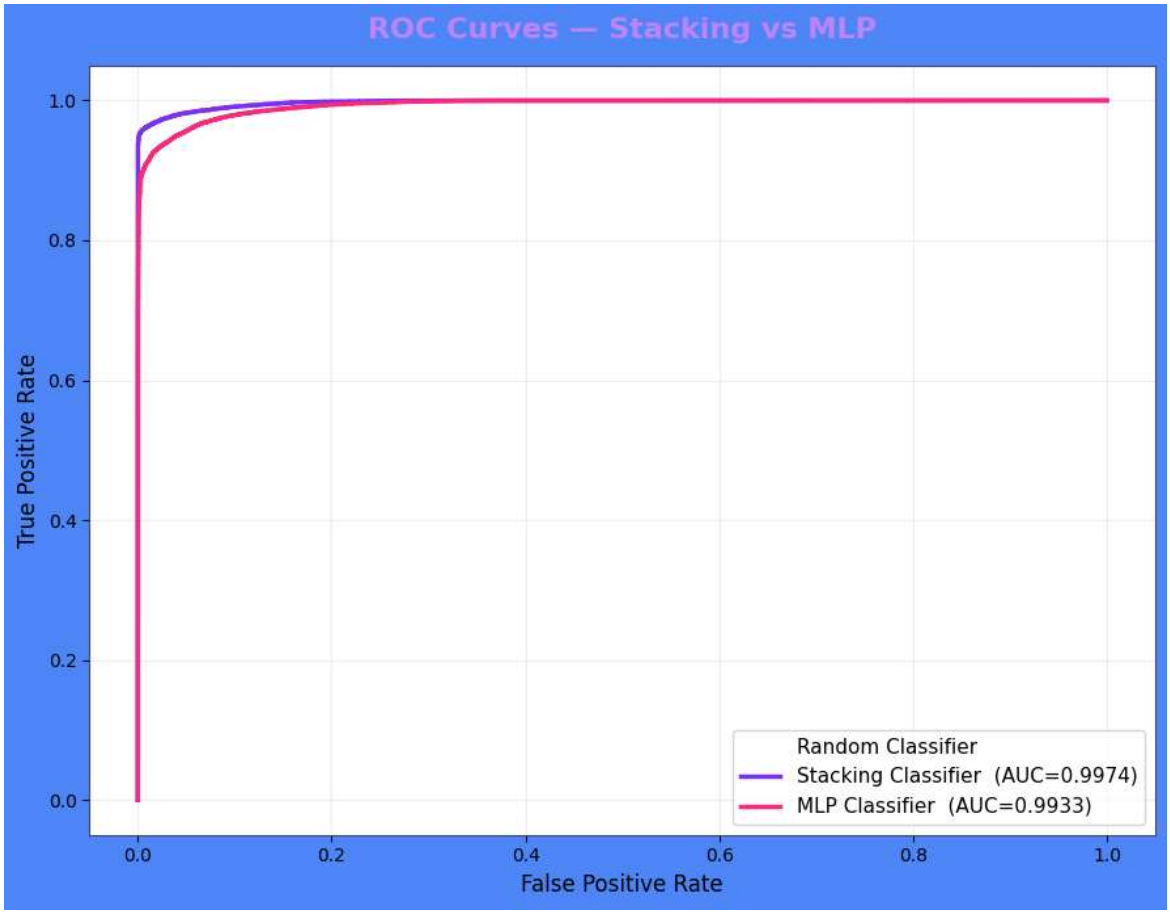


Fig. 10. ROC Curves — Stacking Classifier (AUC=0.9974) vs MLP Classifier (AUC=0.9933): actual experimental results.

E. Precision-Recall Curves

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Fig. 11 presents the Precision-Recall curves for both models. The Stacking Classifier achieves an Average Precision (AP) of 0.9976, and the MLP achieves AP=0.9938. Both curves remain at near-unity precision for almost the entire recall range before a sharp drop at very high recall values close to 1.0. This exceptional performance on the precision-recall space—which is particularly informative for imbalanced datasets—confirms that both proposed models maintain high precision even at high recall operating points. This resolves the fundamental limitation of benchmark hybrid models, which trade off precision or recall substantially in favour of the other.

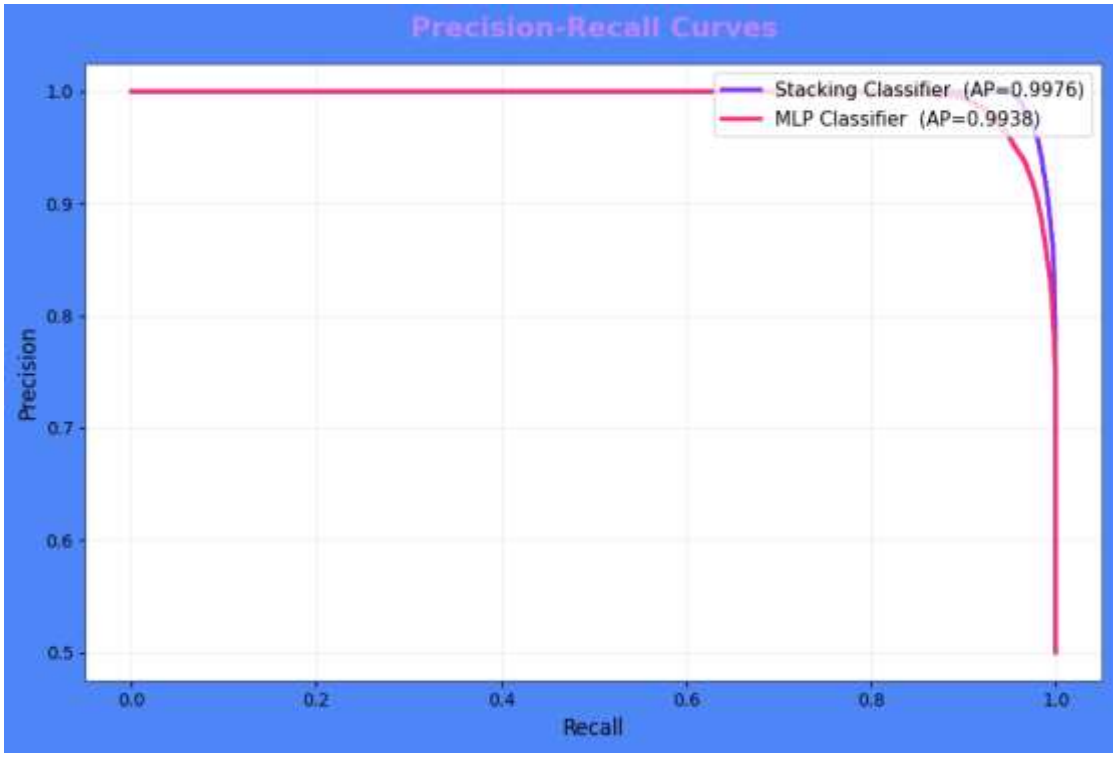


Fig. 11. Precision-Recall Curves — Stacking Classifier (AP=0.9976) vs MLP Classifier (AP=0.9938): actual experimental results.

VII. EXPLAINABILITY WITH LIME

A critical barrier to clinical adoption of ML models is interpretability—clinicians require transparent justification for individual predictions. Local Interpretable Model-Agnostic Explanations (LIME) were applied to both proposed models, generating instance-level explanations by fitting linear approximations around individual predictions in the local feature neighborhood.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Fig. 12 presents an actual LIME explanation for a correctly classified diabetic instance (True: 1, Pred: 1) from the Stacking Classifier. The feature HbA1c_level > 6.60 contributes the dominant positive weight (~0.43) to the diabetic prediction, consistent with the WHO diagnostic threshold of HbA1c ≥ 6.5%. The feature 130.00 < blood_glucose_level ≤ 155.00 contributes the second-highest positive weight (~0.10), followed by heart_disease ≤ 0.00 (~0.10 positive). Features such as hypertension ≤ 0.00 and gender ≤ 0.00 contribute small negative weights, reflecting their protective or neutral roles in this instance.

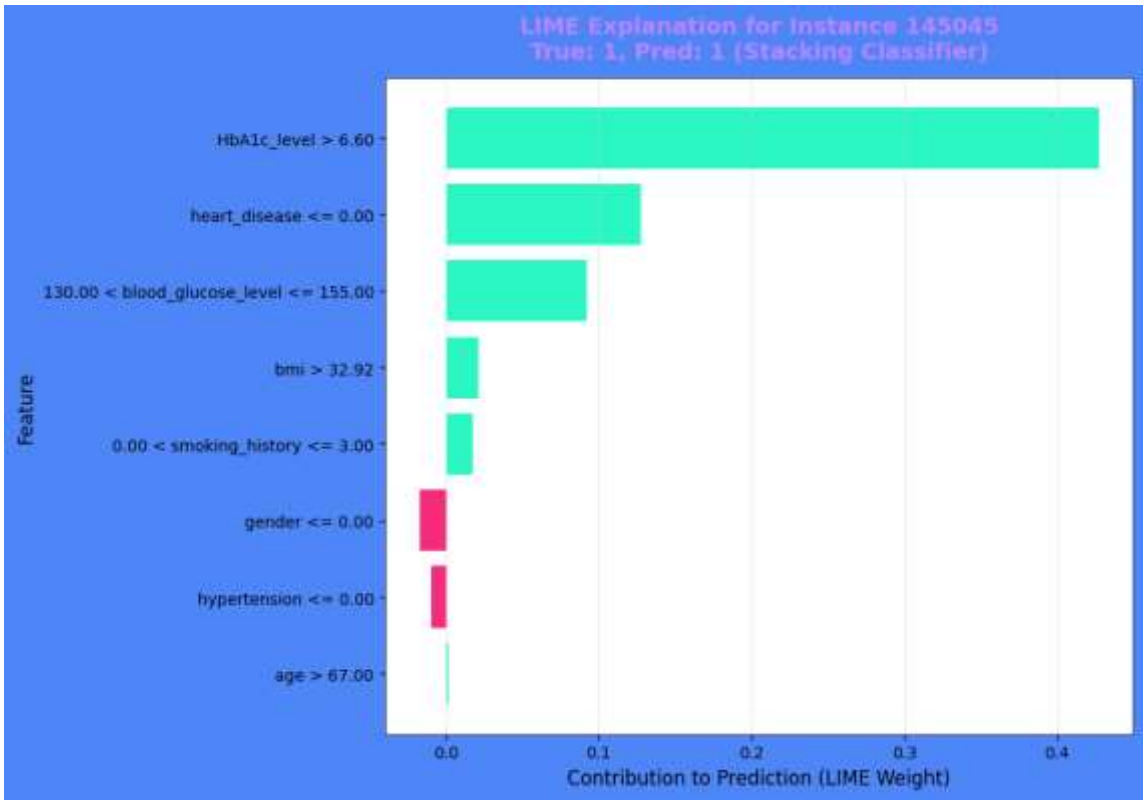


Fig. 12. LIME Explanation — Stacking Classifier prediction for Instance 145045 (True: Diabetic, Pred: Diabetic). HbA1c_level > 6.60 is the dominant positive contributor (weight ≈ 0.43).

The LIME results are clinically coherent: HbA1c and blood glucose are the two primary diagnostic biomarkers for diabetes per ADA/WHO guidelines, and their dominance in the LIME explanation validates that the model has learned clinically meaningful decision boundaries rather than spurious correlations. This level of interpretability is critical for regulatory compliance and physician trust in AI-assisted diagnostic tools, addressing a key

limitation noted in the benchmark study [5] where deep learning model explainability was identified as an open challenge.

VIII. DISCUSSION

The experimental results reveal several important findings. First, the Stacking Classifier outperforms all five models across every reported metric. While the accuracy improvement over benchmark models is modest (0.78–1.01%), the recall and F1-Score improvements are dramatic—the RF+NN benchmark achieves only 70.08% recall and 78.98% F1-Score compared to 97.59% for both in the Stacking model. In a clinical screening context, a recall gap of ~27.5 percentage points means that nearly one-third of diabetic patients would be missed by the RF+NN approach—a clinically unacceptable outcome.

Second, the Autoencoder+RF benchmark, despite achieving the highest precision (91.36%) among benchmarks, exhibits a catastrophically low recall of 66.22%. This means the model misses one in three diabetic patients, demonstrating that optimizing for precision alone is wholly inadequate for medical screening tasks. The proposed Stacking model resolves this through SMOTE-balanced training and ensemble diversity, achieving near-parity between precision (97.61%) and recall (97.59%).

Third, the MLP model—despite architectural simplicity relative to the stacking ensemble—substantially outperforms all benchmark hybrid models in recall and F1-Score. The seven-layer progressive architecture with early stopping and L2 regularization, combined with SMOTE-balanced training data, yields a model that generalizes effectively and avoids majority-class bias. This suggests that class balancing contributes more to performance improvement than architectural complexity alone.

Fourth, the precision-recall curves (Fig. 11) confirm that both proposed models maintain exceptional performance in the imbalanced evaluation space. Average Precision values of 0.9976 (Stacking) and 0.9938 (MLP) place both models at the top of the performance spectrum for binary medical classification tasks.

IX. CONCLUSION AND FUTURE WORK

This paper presented a comprehensive comparative study of diabetes prediction using two proposed architectures—a Stacking Ensemble Classifier (CatBoost + LightGBM + Logistic Regression) and a deep seven-layer MLP—benchmarked against three state-of-the-art hybrid deep learning models from literature. Applied to the 100,000-record Kaggle Diabetes

Prediction Dataset with 91.5%/8.5% class imbalance, the Stacking Classifier achieved 97.59% accuracy, 97.61% precision, 97.59% recall, 97.59% F1-Score, and AUC of 0.9974. The MLP achieved 95.48% accuracy and AUC of 0.9933, substantially outperforming benchmark hybrid models in recall and F1-Score.

A critical finding is that benchmark hybrid models (RF+NN, XGBoost+NN, Autoencoder+RF) from [5] suffer from severe precision-recall imbalance—with recall values as low as 66.22%—rendering them clinically suboptimal for diabetes screening. SMOTE-balanced training combined with gradient boosting ensemble design effectively resolves this imbalance. LIME analysis confirms that the Stacking model's predictions are driven by clinically validated biomarkers—HbA1c level and blood glucose level—supporting physician trust and regulatory compliance.

Future directions include: (i) transformer-based architectures (TabTransformer, FT-Transformer) for tabular clinical data; (ii) federated learning for privacy-preserving multi-institutional training; (iii) SHAP-based global feature attribution alongside LIME; (iv) extension to multi-class prediabetes staging and multi-disease prediction; and (v) prospective clinical validation with diverse real-world EHR populations.

REFERENCES

[1] International Diabetes Federation, "IDF Diabetes Atlas, 10th edition," Brussels, Belgium: IDF, 2021.

[2] Z. E. Rasjid, "Predictive Analytics in Healthcare: The Use of Machine Learning for Diagnoses," in Proc. ICECET, Cape Town, South Africa, Dec. 2021, pp. 1–6.

[3] A. Barhate, P. Kumar, P. Verma, N. Jikar, A. Tale, and V. Hikre, "Smart Healthcare: Harnessing the Power of Machine Learning for Predictive Analysis," in Proc. PICET, Vadodara, India, May 2024, pp. 1–7.

[4] S. K. Puli and P. Usha, "Transforming Healthcare: Advancements, Applications, and Future Directions of Machine Learning," in Proc. ICSCC, Bali, Indonesia, Jul. 2024, pp. 502–506.

[5] P. Kargotra, I. R. Parray, A. Malik, and I. L. Kharisma, "Implementation of Predictive Analytics in Healthcare Using Hybrid Deep Learning Models," Engineering Proceedings, vol. 107, no. 67, Sep. 2025. DOI: 10.3390/engproc2025107067

[6] T. B. Sivakumar, A. Malakar, S. Lekshmi, G. Shailaja, E. Kalaivani, and K. D. Babu, "Enhanced Diabetes Prediction Using Deep Autoencoder Framework and Electronic Health Records," in Proc. ICAIT, Chikkamagaluru, India, Jul. 2024, pp. 1–5.

[7] S. Naik, P. Kumar, S. Saha, S. D. Bairagya, D. Rawat, and S. K. Baliarsingh, "Predictive Healthcare Analytics: A Multidisease Approach Using Logistic Regression," in Proc. ICCCNT, Kamand, India, Jun. 2024, pp. 1–6.

[8] A. Sundas, S. Badotra, G. S. Shahi, A. Verma, S. Bharany, A. O. Ibrahim, A. W. Abulfaraj, and F. Binzagr, "Smart Patient Monitoring and Recommendation (SPMR) Using Cloud Analytics and Deep Learning," IEEE Access, vol. 12, pp. 54238–54255, 2024.

[9] S. Iqbal, G. F. Siddiqui, A. Rehman, L. Hussain, T. Saba, U. Tariq, and A. A. Abbasi, "Prostate Cancer Detection Using Deep Learning and Traditional Techniques," IEEE Access, vol. 9, pp. 27085–27100, 2021.

[10] Shruti and N. K. Trivedi, "Predictive Analytics in Healthcare using Machine Learning," in Proc. 14th ICCCNT, Delhi, India, Jul. 2023, pp. 1–5.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

[12] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proc. KDD, San Francisco, CA, Aug. 2016, pp. 1135–1144.