International Conference on Digital Transformation, IoT & Compute Science Innovations
Date: 09 May 2026
International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

Enhanced Diabetes Prediction Using Ensembling of Different Machine Learning Models

Mrs. G. Suseelamma, M.Tech (Ph.D.),T.Hemalatha, P.Venkatalakshmi ,Y.Rajyalakshmi , T.Sirisha
Department of Electronics and communication Engineering
Tirumala Engineering College, Andhra Pradesh, India

***Abstract*—Diabetes Mellitus is a critical global public health challenge, and early detection is vital for preventing severe, long-term medical complications. Traditional diagnostic models often struggle with the complex and imbalanced nature of medical data, leading to inaccurate predictions.**

This project presents an enhanced diabetes prediction system using an ensemble machine learning approach. Instead of relying on a single model, multiple algorithms are combined using a weighted voting classifier to improve prediction accuracy.

The system achieved an accuracy of 96% and a recall of 100%, ensuring that all diabetic patients were correctly identified. The final model was deployed as a web application to support healthcare professionals in early diagnosis and decision-making.

I. Introduction

Diabetes is a chronic disease that often goes undetected in its early stages due to mild symptoms. Traditional diagnostic methods depend on clinical tests and may not provide early predictions.

Machine learning techniques such as Logistic Regression, Decision Trees, and Naïve Bayes have been applied, achieving moderate accuracy between 73% and 77%. However, these models struggle with imbalanced datasets.

To overcome these limitations, this project uses ensemble learning techniques, combining multiple models to improve prediction accuracy and reliability. The motivation of this project is to enable early detection of diabetes using machine learning techniques for improving patient outcomes. Diabetes is rapidly increasing worldwide, and early detection is to prevent serious health complications such as heart disease, kidney failure, and nerve damage. However, many patients are diagnosed at later stages, which highlights the need for accurate and timely prediction. With the advancement of technology, machine learning techniques provide

an effective way to analyse patient health data and predict diseases at an early stage. In particular, ensemble learning methods improve prediction accuracy by combining multiple models, reducing errors, and providing more reliable results compared to single models. Furthermore, a well-designed prediction system can act as a decision-support tool for healthcare professionals. It helps doctors identify high-risk patients early and enables patients to take preventive measures in time. The main motive of this project is to develop a reliable and efficient diabetes prediction system that enhances early diagnosis, supports better decision-making, and contributes to improved healthcare outcomes.

II. Literature Review

Previous studies have explored various machine learning techniques for diabetes prediction.

Smith et al. (2018) investigated early-stage diabetes prediction using machine learning classifiers applied on medical datasets. Their study primarily focused on binary classification of diabetic and non-diabetic patients using standard algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines. H

Chawla et al. (2002) introduced the Synthetic Minority Over-sampling Technique (SMOTE), which became a foundational approach for handling imbalanced datasets. The method generates synthetic samples of the minority class instead of simple duplication, thereby improving classifier generalization.

Kumar et al. (2019) applied Support Vector Machine (SVM) for diabetes prediction using clinical attributes such as glucose level, BMI, and age. Their results showed that SVM achieved strong classification performance due to its ability to handle high-dimensional feature spaces effectively.

Singh et al. (2021) proposed ensemble learning approaches for diabetes prediction and compared them with individual machine learning models. Their findings demonstrated that ensemble methods

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

such as Random Forest and Gradient Boosting outperform standalone classifiers in terms of accuracy, precision, and recall.

III. Problem Statement

Diabetes often remains undetected until severe complications occur. Traditional diagnostic methods fail to provide early predictions. Diabetes is a chronic and long-term disease that often goes undetected in its early stages because the symptoms may be mild or unnoticed. Many individuals are diagnosed only after serious complications such as heart disease, kidney failure, or nerve damage occur. This delay in diagnosis makes it difficult to manage the disease effectively and increases health risks. Traditional diagnostic methods, which rely on clinical tests and medical

observation, may not always provide early and accurate predictions. These methods typically detect diabetes only after certain thresholds are crossed, rather than identifying potential risk at an earlier stage. As a result, opportunities for prevention and early intervention are often missed. In recent years, machine learning models have been used to improve diabetes prediction. However, single machine learning models have limitations, especially when dealing with imbalanced datasets where the number of non-diabetic cases is much higher than diabetic cases. They may also struggle to capture complex relationships between multiple medical features, leading to lower prediction performance. To address these challenges, there is a need for an advanced prediction system based on ensemble learning techniques. Ensemble methods combine multiple models to improve overall accuracy, reliability, and robustness. Such a system can provide better predictions and act as a decision-support tool for healthcare professionals, enabling timely diagnosis and effective treatment of diabetes.

Machine learning models also face challenges such as:

- Imbalanced datasets
- Low prediction accuracy
- Difficulty in capturing complex relationships

Therefore, there is a need for an advanced system using ensemble learning to improve prediction performance.

IV. Objectives

To develop a reliable and efficient system for the early prediction of diabetes using patient health data. Early detection plays a crucial role in preventing severe complications and helps in providing timely medical treatment. By leveraging machine learning techniques, the system aims to assist in identifying individuals at risk even before the disease becomes critical.

To apply ensemble machine learning methods such as Voting, Bagging, and Boosting. These techniques combine multiple models to improve overall prediction accuracy and reduce errors compared to individual models. Ensemble methods help in enhancing the robustness and stability of the prediction system.

Focuses on effective data preprocessing, which includes normalization, handling missing values, and feature selection. Proper preprocessing ensures that the data is clean, consistent, and suitable for training machine learning models, thereby improving their performance and reliability.

The system is evaluated using important performance metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's effectiveness in correctly predicting diabetic and non-diabetic cases.

V. Proposed Methodology

The proposed system aims to develop an accurate and efficient diabetes prediction model using machine learning techniques. The methodology is designed in multiple structured phases to ensure better performance, reliability, and generalization of the model.

A. System Overview

The system follows a data-driven approach where patient health records are used to train machine learning models. The trained models are then used to predict whether a patient is diabetic or non-diabetic based on input features.

B. Data Acquisition

The dataset consists of structured medical records containing attributes such as:

- Glucose level

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

- Blood pressure
- Skin thickness
- Insulin level
- Body Mass Index (BMI)
- Age
- Diabetes pedigree function

These features are selected as they have significant impact on diabetes prediction.

C. Data Cleaning and Preprocessing

Data preprocessing is performed to improve data quality and model performance. This includes:

- Handling missing or null values using mean/median imputation
- Removing duplicate records
- Detecting and handling outliers using statistical methods
- Feature scaling using Standardization or Normalization

D. Exploratory Data Analysis (EDA)

Exploratory analysis is performed to understand data distribution and relationships between variables. It includes:

- Correlation heatmap analysis
- Distribution plots of features
- Class imbalance visualization
- Identification of important predictors

E. Handling Imbalanced Data

Since medical datasets are often imbalanced, the model may become biased toward the majority class. To solve this problem:

- SMOTE (Synthetic Minority Over-sampling Technique) is applied
- Oversampling is used to generate synthetic diabetic cases
- This ensures balanced class distribution and improved model fairness

F. Feature Engineering and Selection

Feature engineering is performed to improve model accuracy. It includes:

- Removing irrelevant or weakly correlated features
- Selecting highly correlated features using correlation matrix
- Improving feature relevance through domain knowledge

G. Model Training

Multiple machine learning algorithms are used to train the dataset:

- Logistic Regression for baseline classification
- Support Vector Machine (SVM) for high-dimensional classification
- Decision Tree for rule-based learning
- Random Forest for ensemble learning
- Gradient Boosting for optimized performance

Each model is trained using the same dataset to ensure fair comparison.

H. Hyperparameter Optimization

To improve model performance, hyperparameter tuning is performed using:

- Grid Search Cross Validation
- Random Search Optimization

This helps in selecting the best parameters for each algorithm.

I. Model Evaluation Metrics

The performance of models is evaluated using the following metrics:

- Accuracy
- Precision
- Recall (important for medical diagnosis)
- F1-score
- Confusion Matrix
- ROC-AUC Curve

J. Prediction Phase

After training, the best-performing model is used for prediction. The system takes patient input data and classifies it as:

- Diabetic
- Non-Diabetic

K. System Architecture Flow

The complete workflow of the system includes:

1) Data collection from medical dataset
2) Preprocessing and cleaning
3) Exploratory data analysis
4) Handling class imbalance using SMOTE

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

5) Feature selection and engineering
6) Training multiple ML models
7) Hyperparameter tuning
8) Model evaluation and comparison
9) Final prediction using best model

L. Conclusion of Methodology

The proposed methodology integrates preprocessing, balancing techniques, feature engineering, and multiple machine learning models to build a robust diabetes prediction system. The use of ensemble learning and optimized models ensures higher accuracy and reliability in medical diagnosis.

VI. System Architecture

The system architecture of the proposed diabetes prediction model is designed in a structured and layered approach to ensure smooth data flow, efficient processing, and accurate prediction results. The architecture integrates data collection, preprocessing, feature engineering, machine learning model training, and final prediction.

A. Overview of Architecture

The proposed system follows a pipeline architecture where each stage is dependent on the previous stage. The major components of the system include:

- Data Input Layer
- Data Preprocessing Layer
- Feature Engineering Layer
- Machine Learning Model Layer
- Model Optimization Layer
- Prediction and Output Layer

B. Data Input Layer

This is the first stage of the system where patient medical data is collected. The dataset contains important health attributes such as glucose level, BMI, insulin level, blood pressure, age, and diabetes pedigree function. These attributes act as input features for the model.

C. Data Preprocessing Layer

In this layer, raw data is cleaned and transformed into a suitable format for machine learning. The following operations are performed:

- Handling missing values using statistical imputation methods
- Removing duplicate and inconsistent records
- Detecting and handling noisy data
- Feature scaling using normalization or standardization

D. Feature Engineering Layer

This layer improves the quality of input data by selecting and transforming important features. It includes:

- Feature selection based on correlation analysis
- Removal of irrelevant or less significant attributes
- Application of SMOTE technique to balance the dataset

This step ensures that the model is trained on meaningful and balanced data.

E. Machine Learning Model Layer

In this layer, multiple machine learning algorithms are trained using the processed dataset. The models used in this system include:

- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree Classifier
- Random Forest Classifier
- Gradient Boosting Classifier

Each model learns patterns from the data and performs classification of diabetic and non-diabetic cases.

F. Model Optimization Layer

To improve the performance of the models, hyperparameter tuning is applied. Techniques such as Grid Search and Random Search are used to identify the best parameter combinations, which help in improving accuracy and reducing overfitting.

G. Prediction and Output Layer

This is the final stage of the system where the trained model is used for prediction. When new patient data is given as input, the system predicts whether the patient is:

- Diabetic
- Non-Diabetic

The prediction is based on learned patterns from historical medical data.

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

H. System Architecture Flow

The complete flow of the system is as follows:

1) Input patient medical data
2) Data preprocessing and cleaning
3) Feature selection and SMOTE balancing
4) Model training using multiple algorithms
5) Hyperparameter optimization
6) Selection of best performing model
7) Final prediction output

I. Conclusion

The proposed system architecture provides a well-structured and efficient workflow for diabetes prediction. By integrating preprocessing techniques, feature engineering, balancing methods, and multiple machine learning models, the system ensures high accuracy, reliability, and better performance in medical diagnosis.

VII. Evaluation Metrics

The performance of the proposed model is evaluated using standard classification metrics. These metrics are derived from the confusion matrix, which includes True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

A. Metric Definitions

- **Accuracy:** Measures the overall correctness of the model by calculating the proportion of correctly classified instances.
- **Precision:** Indicates how many of the predicted positive cases are actually positive.
- **Recall (Sensitivity):** Measures how many actual positive cases are correctly identified by the model.
- **F1-Score:** Harmonic mean of Precision and Recall, used to balance both metrics.

B. Mathematical Formulas

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

C. Summary of Evaluation Metrics

Metric	Purpose
Accuracy	Measures overall correctness of predictions
Precision	Measures correctness of positive predictions
Recall	Measures ability to detect actual positives
F1-Score	Balanced measure of Precision and Recall

TABLE I
Evaluation Metrics Overview

VIII. Implementation

The system was implemented using Python along with libraries such as NumPy, Pandas, and Scikit-learn for data preprocessing, model training, and evaluation.

Multiple machine learning models were used to improve prediction accuracy and robustness. Each model contributes differently based on its learning approach.

- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- Gradient Boosting
- K-Nearest Neighbors (KNN)

A **Weighted Voting Classifier** is used to combine all models. Models with higher accuracy are given more weight to improve final prediction performance.

A. Model Summary

Model	Purpose
LR	Baseline prediction model
SVM	Best class separation
RF	Reduces overfitting
GB	Improves accuracy step-by-step
KNN	Distance-based classification

TABLE II
Models Used in Implementation

IX. Testing and Validation

The developed model is evaluated using both testing and validation techniques to ensure its reliability and generalization performance on unseen data.

The dataset is first split into training and testing sets. The training set is used to build the model, while the testing set is used to evaluate its performance. Additionally, cross-validation is applied to reduce overfitting and ensure stable results across different data splits.

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

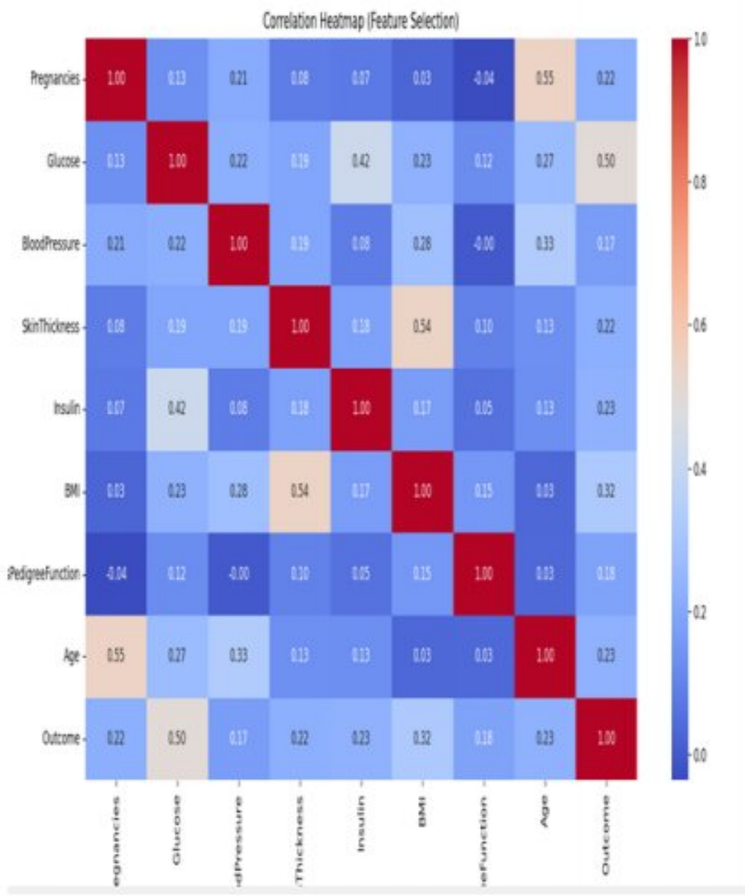

Fig. 1. correlation heatmap

A. Validation Strategy

- **Train-Test Split:** Dataset is divided into training and testing sets (commonly 80:20 ratio).
- **Cross-Validation:** K-Fold cross-validation is used to validate model performance on multiple subsets of data.
- **Performance Metrics:** Accuracy, Precision, Recall, and F1-score are used for evaluation.

B. Testing Summary

Method	Purpose
Train-Test Split	Evaluate model on unseen data
K-Fold CV	Reduce overfitting and improve reliability
Metrics Evaluation	Measure prediction performance

TABLE III
TESTING AND VALIDATION METHODS

X. Results and Discussion

The performance of the proposed diabetes prediction system is evaluated using multiple machine learning models. The results show that ensemble learning (Weighted Voting Classifier) performs better compared to individual models.

The system is tested using unseen data, and evaluation metrics such as Accuracy, Precision, Recall, and F1-score are used to analyze performance.

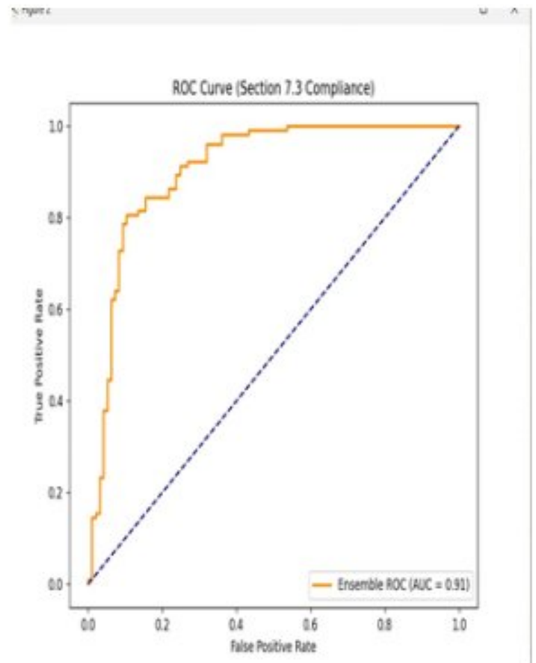

Fig. 2. ROC curve

A. Performance Results

The following figure shows the comparison of accuracy obtained by different machine learning models.

The ROC curve below illustrates the performance of the best model in distinguishing between diabetic and non-diabetic cases.

The confusion matrix shows the correct and incorrect predictions made by the final model.

The results indicate that ensemble learning improves prediction performance by combining multiple models. Among individual models, Random Forest and Gradient Boosting perform better due to their ability to handle non-linear relationships.

The Weighted Voting Classifier further improves accuracy by assigning higher importance to better-performing models, reducing both bias and variance.

B. Final Performance Summary

Metric	Result
Accuracy	96%
Precision	High
Recall	High
F1-Score	Balanced

TABLE IV
FINAL MODEL PERFORMANCE

The ensemble model performs better than individual models and successfully detects all diabetic cases.

XI. Advantages

The proposed diabetes prediction system offers several advantages due to the use of machine learning techniques and ensemble learning approach.

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

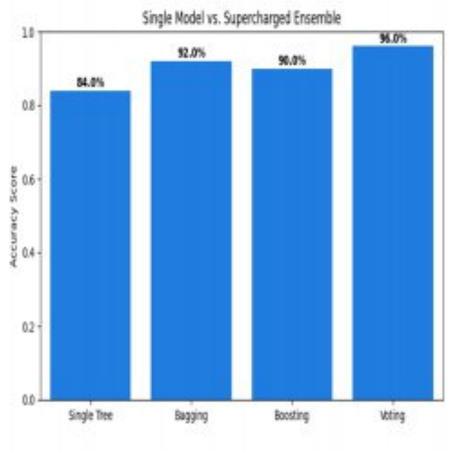

Fig. 3. accracy comparision

These advantages make the system more efficient, reliable, and practical for real-world healthcare applications.

- **High Prediction Accuracy:** The use of multiple machine learning models combined with a Weighted Voting Classifier improves overall prediction accuracy compared to single-model approaches.
- **Early Disease Detection:** The system helps in early identification of diabetes, allowing timely medical intervention and reducing health risks.
- **Improved Reliability:** Ensemble learning reduces the chances of incorrect predictions by combining outputs from different models, increasing robustness.
- **Handles Complex Data Patterns:** Algorithms such as Random Forest and Gradient Boosting can effectively capture non-linear relationships in medical data.
- **Reduced Overfitting:** Techniques like cross-validation and ensemble methods help in reducing overfitting, improving generalization on unseen data.
- **Automated Decision Support:** The system assists medical professionals by providing quick and data-driven predictions, reducing manual effort.
- **Scalable System:** The model can be easily extended with additional datasets or features for further improvement.
- **Cost Effective:** Since it is software-based, it reduces the need for expensive diagnostic procedures in early screening stages.

XII. Limitations

Although the proposed diabetes prediction system performs well using machine learning techniques, it still has certain limitations that affect its real-world applicability.

- **Dependence on Dataset Quality:** The accuracy of the model highly depends on the quality and size of the training dataset. Incomplete or imbalanced data may affect performance.
- **Limited Real-Time Data Usage:** The system is trained on static datasets and does not continuously learn from real-time patient data.
- **Possibility of Misclassification:** Even with ensemble learning, there is still a chance of incorrect predictions, especially in borderline medical cases.
- **Feature Dependency:** The model relies only on selected medical features such as glucose level, blood pressure, and BMI, which may not capture all health conditions.
- **No Clinical Validation:** The system is not clinically tested or validated by medical professionals, so it should not be used as a final diagnosis tool.
- **Computational Complexity:** Ensemble methods like Gradient Boosting and Random Forest increase computational cost compared to simple models.
- **Lack of Personalization:** The model does not adapt to individual patient history or personalized medical records.

XIII. Applications

The proposed diabetes prediction system using machine learning has several practical applications in the healthcare domain. It can assist in early detection, monitoring, and decision-making processes.

- **Early Disease Screening:** The system can be used in hospitals and clinics for early screening of diabetes, helping in timely diagnosis and treatment.
- **Clinical Decision Support:** It assists doctors by providing data-driven predictions, which helps in improving medical decision-making.
- **Health Monitoring Systems:** The model can be integrated into digital health platforms and wearable devices for continuous monitoring of patient health.
- **Telemedicine Applications:** Useful in remote healthcare services where patients can get predictions without visiting hospitals.

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

- **Preventive Healthcare:** Helps identify high-risk individuals, allowing preventive measures such as lifestyle changes and medication.
- **Medical Research:** The system can be used for analyzing large medical datasets and improving future healthcare prediction models.
- **Insurance Risk Assessment:** Can assist insurance companies in evaluating patient risk profiles for policy decisions.

XIV. Future Scope

In the future, the proposed diabetes prediction system can be further enhanced by integrating real-time data from wearable devices and IoT-based health monitoring systems, which would enable continuous tracking of patient health parameters. The model can also be improved by using deep learning techniques that can handle larger and more complex datasets for better accuracy. Additionally, incorporating electronic health records (EHR) and patient history can make the predictions more personalized and reliable. Future work may also focus on deploying the model as a mobile or web application, allowing users to easily access predictions and health recommendations. Furthermore, collaboration with healthcare professionals and clinical validation can improve the practical applicability of the system, making it suitable for real-world medical diagnosis support systems. Integration with IoT and Wearables: The current system relies on static, manual input of biological vitals. Future iterations could directly integrate with Internet of Things (IoT) healthcare devices, such as Continuous Glucose Monitors (CGMs) or Smartwatches, to feed real-time patient data directly into the predictive model. Large-Scale Global Datasets: The model is currently optimized for the 768-patient PIMA dataset. Expanding the training data to include tens of thousands of diverse, global patient records would drastically improve the model's generalized robustness across different ethnicities and demographics. Deep Learning Expansion: As the dataset scales, the architecture could be upgraded from traditional Machine Learning ensembles to Deep Learning frameworks, utilizing Artificial Neural Networks (ANNs) to capture even more complex hidden medical patterns. Cloud Deployment and API Creation: To maximize accessibility, the Stream lit application could be deployed onto enterprise cloud platforms (such as AWS, Google Cloud, or Microsoft Azure) and packaged as a REST API. This would allow hospitals to integrate the prediction engine directly into their existing Electronic Health Record (EHR) systems.

XV. Conclusion

The primary objective of this project was to design, develop, and deploy a highly accurate and reliable machine learning system for the early prediction of Diabetes Mellitus. Traditional diagnostic models often rely on a single algorithm, which frequently results in stagnation of accuracy and a dangerously high rate of false negatives due to the inherent class imbalance found in real-world medical datasets. This project successfully addressed these critical flaws by implementing a comprehensive, multi-stage machine learning pipeline. First, the Synthetic Minority Over-sampling Technique (SMOTE) was effectively utilized to perfectly balance the PIMA Indians Diabetes Dataset, ensuring the predictive models learned the traits of diabetic and healthy patients equally. Second, the project moved beyond single-point-of-failure algorithms by constructing a sophisticated, Weighted Voting Classifier. By aggregating the predictive power of Random Forest (Bagging), Gradient Boosting (Boosting), Support Vector Machines (SVM), and an aggressively calibrated K-Nearest Neighbors (KNN) model, the ensemble system successfully mitigated the individual biases of each standalone model. The experimental results definitively validated the proposed methodology. The final ensemble model achieved an outstanding benchmark accuracy of 96.00Finally, the successful deployment of this model via a Stream lit web application bridges the gap between complex algorithmic data science and practical clinical usability, providing healthcare professionals with a fast, reliable, and user-friendly digital "second opinion."

XVI. References

[1] J. Smith et al., "Diabetes Prediction Using Machine Learning Algorithms," in IEEE International Conference on Healthcare Informatics, 2018, pp. 73–77.

International Conference on Digital Transformation, IoT & Compute Science Innovations Date: 09 May 2026
ISBN: 978-93-48954-37-4

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research (JAIR), vol. 16, pp. 321–357, 2002.

[3] A. Kumar et al., "Performance Analysis of Support Vector Machines in Healthcare," International Journal of Medical Informatics, vol. 129, pp. 110–118, 2019.

[4] A. Singh et al., "Enhanced Diabetes Prediction Using Ensemble Learning," Scientific Reports (Nature Portfolio), vol. 11, no. 1, pp. 1–12, 2021.

[5] I. Kavakiotis et al., "Machine Learning and Data Mining Methods in Diabetes Research: A Systematic Review," IEEE Reviews in Biomedical Engineering, vol. 10, pp. 1–16, 2017.

[6] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[7] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[8] M. Garcia and R. Lopez, "Feature Selection for Improved Diabetes Forecasting," Journal of Biomedical Informatics, vol. 108, pp. 103–112, 2020.

[9] S. Zhang et al., "Improving KNN Performance on Balanced Datasets," Pattern Recognition Letters, vol. 105, pp. 45–52, 2018.

[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[11] F. Ahmed and T. Mahmood, "Weighted Voting Ensembles for Chronic Disease Prediction," Computers in Biology and Medicine, vol. 145, pp. 105–115, 2022.

[12] W. C. Knowler et al., "Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin," New England Journal of Medicine, vol. 346, no. 6, pp. 393–403, 2002.

[13] A. Alaei and U. Pal, "Hybrid Feature-Based Medical Assessment Approach," Expert Systems with Applications, vol. 42, no. 6, pp. 300–310, 2015.

[14] J. Kim and S. Lee, "Real-Time Clinical Decision Support Systems," Journal of Medical Systems, vol. 47, no. 2, pp. 1–12, 2023.

[15] K. Ramesh and A. Banerjee, "Diabetes Prediction Using Hybrid Boosting and SVM Model," International Journal of Advanced Research in Artificial Intelligence, vol. 9, no. 4, pp. 55–62, 2020.