



# Machine Learning for Diabetes Prediction: Performance Analysis Using Logistic Regression, Naïve Bayes, and Decision Tree Models



Rupinder Kaur<sup>1</sup>, Raman Kumar<sup>2,3\*</sup>, Swapandeep Kaur<sup>4\*</sup>, Gurneet Singh<sup>1</sup>, Arshnoor Kaur<sup>1</sup>, Sukhpal Singh<sup>4</sup>

<sup>1</sup> Department of Information Technology, Guru Nanak Dev Engineering College, 141006 Ludhiana, India

<sup>2</sup> Department of Mechanical and Production Engineering, Guru Nanak Dev Engineering College, 141006 Ludhiana, India

<sup>3</sup> Jadara Research Center, Jadara University, 21110 Irbid, Jordan

<sup>4</sup> Department of Electrical Engineering, Guru Nanak Dev Engineering College, 141006 Ludhiana, India

\* Correspondence: Raman Kumar (sehgal91@yahoo.co.in); Swapandeep Kaur (swapandeep80@gmail.com)

**Received:** 09-17-2024 **Revised:** 11-22-2024 **Accepted:** 12-10-2024

**Citation:** Kaur, R., Kumar, R., Kaur, S., Singh, G., Kaur, A., & Singh, S. (2024). Machine learning for diabetes prediction: Performance analysis using Logistic Regression, Naïve Bayes, and Decision Tree models. *Healthcraft. Front.*, 2(4), 169-187. https://doi.org/10.56578/hf020401.



@ 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Diabetes is a chronic metabolic disorder that affects millions of people worldwide, making early detection crucial for effective management. This study assesses the effectiveness of three machine learning (ML) models, Logistic Regression (LR), Naïve Bayes (NB), and Decision Tree (DT), in predicting diabetes based on data from 392 individuals, including their demographic and clinical characteristics. The dataset underwent preprocessing to maintain data integrity, was standardized for model compatibility, and analyzed through feature correlation heatmaps, feature importance assessments, and statistical significance tests. The findings revealed that LR surpassed the other models, with the highest accuracy (78%), precision (73%), and F1-score (65%) for diabetic cases. NB showed moderate performance with 75% accuracy, while DT demonstrated the lowest accuracy (71%) due to overfitting. Receiver Operating Characteristic (ROC) analysis revealed strong discriminative power across all models, although perfect Area Under the Curve (AUC) scores indicate potential overfitting needing further validation. The study emphasizes the significance of key features like Glucose, Body Mass Index (BMI), and Age, which showed notable differences between diabetic and non-diabetic individuals. By enabling early detection and proactive management, these models can contribute to reducing diabetes-related complications, enhancing patient outcomes, and lessening the burden on healthcare systems. Future research should investigate ensemble learning, deep learning, and real-time data integration from Internet of Things (IoT) devices to improve predictive accuracy and scalability.

**Keywords:** Diabetes prediction; Machine learning models; Logistic Regression; Clinical Decision Support Systems; Predictive analytics; Classification

# 1. Introduction

Diabetes is a chronic metabolic disorder impacting millions globally. It is marked by high blood glucose levels that, if untreated, can result in serious complications like cardiovascular disease (CVD), kidney failure, nerve damage, and vision problems (Garg, 2021). The rising incidence of diabetes, particularly type 2 diabetes, has become a major global health issue, increasingly straining healthcare systems. Timely detection and effective diabetes management are essential for minimizing complications and enhancing the quality of life for those affected (Al-Shanableh et al., 2024; Kumar et al., 2022). While traditional diagnostic techniques such as Fasting Blood Sugar (FBS) tests, Oral Glucose Tolerance Tests (OGTT), and Glycated Hemoglobin (HbA1c) tests are reliable, they often require laboratory settings and can be time-consuming (Samet et al., 2022). This has led to heightened interest in applying advanced computational methods like machine learning (ML) to create automated, efficient, and precise predictive models for diabetes diagnosis (Wong et al., 2022).

This study's motivation stems from the need to assess the performance of various ML algorithms in diabetes

prediction. Among the many ML models, Logistic Regression (LR), Naïve Bayes (NB), and Decision Tree (DT) classifiers are prominent in medical diagnostics for their interpretability, efficiency, and predictive strength (Marzouk et al., 2022). LR is a straightforward yet effective statistical model that estimates the likelihood of diabetes based on various input features. NB is a probabilistic framework that assumes independence among features and excels with smaller datasets. Conversely, DTs offer a structured, rule-based classification method that medical professionals can easily interpret. By evaluating these models based on accuracy, precision, recall, and computational efficiency, this study seeks to identify the best algorithm for diabetes prediction (Ebrahim & Derbew, 2023; Iparraguirre-Villanueva et al., 2023).

ML has transformed the healthcare sector by facilitating data-driven decisions, enhancing predictive analytics, and automating disease diagnoses (Kaur et al., 2023; Kiran et al., 2024). The surge in digital healthcare data, like Electronic Health Records (EHRs), wearable sensor data, and genomic information, presents a significant opportunity to utilize ML techniques that improve patient outcomes (Alenezi et al., 2023). Through the analysis of extensive datasets, ML algorithms can uncover hidden patterns and deliver precise predictions that aid healthcare professionals in diagnosing conditions, customizing treatments, and streamlining medical operations (Singh et al., 2022).

One key benefit of ML in healthcare is its capacity to handle immense volumes of complex and highdimensional data that traditional statistical techniques often struggle with (Kumar et al., 2021). Various medical applications extensively employ ML methods, including early disease detection, medical image evaluation, drug development, and patient risk assessment. Common supervised learning models for classification tasks are LR, DTs, Support Vector Machines (SVMs), and Neural Networks (Samet et al., 2022). These are effective in diagnosing conditions such as diabetes, cancer, and CVDs. Meanwhile, unsupervised learning techniques like clustering and anomaly detection help identify disease patterns, facilitate patient stratification, and detect fraudulent activity in medical billing (Jader et al., 2022).

Deep learning, a branch of ML, has significantly expanded the potential of healthcare analytics by allowing feature extraction from unstructured data sources, including medical images, text reports, and genomic data (Reddy et al., 2020; Singh et al., 2022). Convolutional Neural Networks (CNNs) excel in medical imaging tasks, particularly tumor detection (Mridul et al., 2024). At the same time, Recurrent Neural Networks (RNNs) (Mridul et al., 2024) and Transformers serve in Natural Language Processing (NLP) applications, analyzing clinical notes and medical publications (Chowdhury et al., 2021).

However, despite its promising applications, ML in healthcare confronts several challenges, including concerns about data privacy, the interpretability of complex models, biases in training datasets, and regulatory considerations (Pati et al., 2023). Still, advancements in ML methodologies and increasing collaborations between healthcare professionals and data scientists have led to significant progress in creating robust predictive models for disease diagnosis and effective patient management (Mavrogiorgou et al., 2022; Ratta & Sharma, 2024).

Over the last five years, extensive research has been conducted into diabetes prediction using ML models, with various studies assessing different algorithms to enhance diagnostic accuracy. Many researchers have concentrated on creating predictive models utilizing publicly accessible datasets, including the Pima Indians Diabetes Dataset (PIDD) from the UCI ML Repository, which features diagnostic data collected from female patients of Pima Indian descent. Numerous studies have analyzed various ML classifiers for diabetes prediction.

Shen et al. (2020) aimed to develop an AI-based mobile application for diagnosing Gestational Diabetes Mellitus (GDM), particularly in resource-limited settings. The authors trained nine ML algorithms, including SVM, Random Forest (RF), AdaBoost, k-Nearest Neighbors (kNN), NB, DT, LR, XGBoost, and Gradient Boosting Decision Tree (GBDT). Among these, SVM achieved the highest Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.780 and maintained 100% specificity, with an accuracy of 88.7% on the external validation dataset. The study concluded that SVM effectively diagnosed GDM using minimal features age and fasting blood glucose making it suitable for low-resource environments. Choubey et al. (2020) proposed an efficient diagnostic tool for early diabetes detection using feature reduction techniques. The study applied LR, kNN, ID3 DT, C4.5 DT, and NB, both with and without feature reduction through Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO). Feature reduction significantly enhanced model accuracy and reduced computation time, demonstrating the potential of this approach for real-time healthcare applications. Misra & Yadav (2020) focused on improving type 2 diabetes prediction using Recursive Feature Elimination with Cross-Validation (RFECV). The study compared the performance of LR, Artificial Neural Networks (ANNs), NB, SVM, and DT, with LR achieving the highest accuracy of 84%. The authors concluded that feature selection improved prediction accuracy while mitigating overfitting, highlighting the importance of selecting relevant features. Maniruzzaman et al. (2020) combined LR-based feature selection with ensemble classifiers such as NB, DT, AdaBoost, and RF. The dataset, derived from the National Health and Nutrition Examination Survey, was partitioned using protocols K2, K5, and K10. The combination of LR for feature selection and RF for classification yielded an accuracy of 94.25% and an Area Under the Curve (AUC) of 0.95, demonstrating the effectiveness of this hybrid approach. Ahmed et al. (2021) evaluated the performance of DT, NB, kNN, RF, Gradient Boosting (GB), LR, and SVM on multiple datasets after applying preprocessing techniques such as label encoding and

normalization. RF achieved the highest accuracy, with improvements ranging from 2.71% to 13.13%, depending on the dataset and algorithm used. The study highlighted the importance of data preprocessing and feature selection in enhancing model performance. Kumar et al. (2021) proposed an ensemble model that combined NB, RF, and LR (NB-RF-LR-SEMod). The ensemble model achieved an accuracy of 88.3%, outperforming individual classifiers. The authors emphasized that ensemble models, which leverage the strengths of multiple algorithms, were more effective in predicting diabetes than single classifiers.

Rajput et al. (2022) developed a predictive model to assist rural populations in India by comparing the performance of LR, SVM, RF, DT, NB, and kNN. SVM achieved the highest accuracy of 96%, making it the preferred algorithm for early diabetes detection in rural areas. The model facilitated improved communication between patients and healthcare providers, enhancing early diagnosis and intervention. Kumar et al. (2022) addressed the class imbalance challenge in clinical datasets by applying data-balancing techniques such as SMOTEEN, SMOTE, ADASYN, and SVM-SMOTE. The study evaluated the performance of DT, kNN, LR, ANN, SVM, and Gaussian NB on five clinical datasets, including the PIDD. The SMOTEEN technique consistently improved accuracy, precision, and recall across all classifiers and datasets, demonstrating its effectiveness in mitigating class imbalance. Lu et al. (2022) introduced a novel approach that combined patient networks with ML models to predict the risk of Type II Diabetes Mellitus (T2DM). The dataset included 1,028 T2DM patients and 1,028 non-T2DM patients. RF achieved the highest AUC of 0.91, with patient network features such as eigenvector centrality and closeness centrality significantly improving prediction accuracy. The study concluded that integrating patient networks with ML models provided valuable insights into disease progression and risk factors. Mushtaq et al. (2022) applied data-balancing techniques such as SMOTE to address dataset imbalance and compared the performance of LR, SVM, kNN, GB, NB, and RF. RF achieved an accuracy of 80.7% after applying SMOTE, while an ensemble voting algorithm combining NB, GB, and RF improved accuracy to 82.0%. The authors concluded that balancing techniques and ensemble methods were crucial for improving prediction accuracy in imbalanced datasets. Singh et al. (2022) compared the performance of SVM, kNN, LR, RF, NB, and Deep Neural Networks (DNN) for predicting chronic kidney disease (CKD). The DNN outperformed all other classifiers, achieving 100% accuracy. The study emphasized the potential of DNNs for accurate early diagnosis, especially when trained with well-preprocessed datasets.

Hossain et al. (2023) developed a model for predicting Polycystic Ovary Syndrome (PCOS) using LR, RF, DT, NB, SVM, kNN, XGBoost, AdaBoost, and Stacking Ensemble techniques. The combination of Recursive Feature Elimination (RFE) and Stacking Ensemble achieved 100% accuracy, demonstrating the effectiveness of feature selection and ensemble learning in medical diagnostics. Kangra & Singh (2023) evaluated several ML algorithms, including SVM, NB, kNN, RF, LR, and DT, using the PIDD and Germany Diabetes Dataset. SVM achieved 74% accuracy for the PIDD, while kNN and RF achieved 98.7% accuracy for the Germany dataset, highlighting the importance of dataset characteristics in algorithm performance. Acheampong et al. (2024) employed ML techniques to predict pre-metabolic syndrome (pre-MetS) and metabolic syndrome (MetS) in 919 type 2 diabetes patients. Using BORUTA feature selection, key predictors included Visceral Adiposity Index (VAI), Lipid Accumulation Product (LAP), and triglyceride-glucose index adjusted for waist-to-height ratio (TyG-WHtR). The ensemble majority voting classifier achieved AUCs of 0.79 and 0.87 for pre-MetS and MetS, respectively, supporting early detection and personalized interventions. Das et al. (2024) investigated the impact of type 2 diabetes on CVD risk in Bangladesh using data from the 2011 and 2017-2018 Bangladesh Demographic and Health Surveys. Eight ML algorithms, including RF, LR, and XGBoost, were compared, with RF achieving the highest specificity (76.96%), overall accuracy (75.21%), and AUC (80.79%). Age, wealth, and weight status significantly influenced CVD risk, highlighting the potential of ML models for early detection, targeted interventions, and improving healthcare infrastructure.

The studies reviewed showed that combining feature selection, data balancing, and ensemble learning significantly improved the accuracy and efficiency of diabetes prediction models. Advanced ML techniques have the potential to enhance early diagnosis, lower healthcare costs, and aid personalized treatment, making them valuable tools for healthcare professionals around the globe. These studies indicate that while conventional ML models like LR and C4.5 DT are still prevalent in diabetes prediction, newer approaches such as ensemble learning and deep learning promise better performance. Nonetheless, challenges like model interpretability, dataset imbalance, and computational complexity must be resolved to implement ML-based diabetes prediction systems effectively. ML models are essential in medical diagnosis, providing data-driven disease prediction approaches. LR, NB, and C4.5 DT are among the most commonly used classification algorithms, each with unique strengths and weaknesses.

LR is a straightforward but powerful method for binary classification tasks, and it is ideal for predicting diseases. It yields interpretable outcomes and operates effectively with minor to medium-sized datasets. Nevertheless, it relies on a linear relationship between features and outcomes, which can be a limitation when dealing with more complex medical data, hindering its effectiveness in non-linear scenarios (Ratta & Sharma, 2024).

NB is a probabilistic model praised for its computational efficiency and effectiveness with small datasets. It offers confidence scores for disease likelihood, making it beneficial for risk assessments. However, its assumption

of feature independence is often inaccurate in medical contexts, where physiological variables typically exhibit correlations, resulting in decreased accuracy in complicated diagnostic situations.

DTs utilize a hierarchical, rule-based classification method that captures non-linear dynamics and feature interactions. They are highly interpretable and automatically select features. However, they are susceptible to overfitting, particularly with deep trees, necessitating techniques like pruning or ensemble methods such as RF to enhance generalization.

In diabetes prediction, LR acts as a reliable baseline model, NB offers a quick yet limited probabilistic method, and DTs excel in uncovering intricate patterns but need careful adjustment. The model selection hinges on interpretability, accuracy, and the complexity of the dataset, underscoring the importance of model evaluation for achieving the best predictive performance in healthcare settings.

Diabetes is a common and potentially life-threatening metabolic condition that can result in serious complications, including CVDs, kidney failure, and nerve damage, if not diagnosed early. Conventional diagnostic approaches usually depend on laboratory tests and clinical assessments, which aren't always easily accessible, timely, or affordable. With the rising incidence of diabetes, ML methods present a valuable opportunity for automated and early prediction based on patient information. Nevertheless, the performance of various ML models differs based on parameters like accuracy, interpretability, and computational efficiency. The main challenge is pinpointing the most dependable ML model that can accurately differentiate between diabetic and non-diabetic individuals while maintaining scalability and practical usability. This study primarily aims to achieve the following objectives:

- To compare and evaluate the performance of three ML models—LR, NB, and DT—for diabetes prediction.
- To analyze key classification metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to determine the most effective model.
- To examine the computational efficiency of each model and assess its suitability for real-time healthcare applications.
- To explore the practical implications of ML-based diabetes prediction in early diagnosis, patient management, and Clinical Decision Support Systems (CDSS).
- To identify the challenges and future directions for improving ML-driven diabetes prediction models in healthcare.

This study is of great importance to medical professionals and researchers as it delivers a comprehensive analysis of ML models tailored for diabetes prediction, which aids in identifying the most efficient methods for early diagnosis. Reliable prediction models are crucial for facilitating early interventions, thereby lowering the chances of complications and enhancing patient outcomes. The research emphasizes the potential for incorporating ML models within telemedicine, wearable health devices, and hospital management systems, thereby streamlining and automating diabetes risk assessments. This study furthers the advancement of AI-driven healthcare solutions by tackling critical issues such as model interpretability, data bias, and computational efficiency, promoting improved disease management and lower healthcare costs worldwide.

### 2. Methodology

This section provides a visual summary of the methodology shown in Figure 1. The study's methods consisted of several key steps designed to develop an effective diabetes prediction model. Initially, all records with missing or null values were eliminated during preprocessing to uphold data integrity and quality. Subsequently, a Chi-square test was performed to identify features significantly correlated with diabetes. This step effectively differentiated primary features (Age, Insulin, and Glucose levels) from secondary ones, ensuring that the model's most informative variables were chosen. Next, three ML algorithms, DT, NB, and LR, were selected based on their varied approaches and strengths in classification tasks. Ultimately, these algorithms were applied to the preprocessed dataset, and their performance was assessed using accuracy, precision, recall, and F1-score metrics. This assessment revealed that LR was the most robust model, showcasing balanced and superior performance compared to the other algorithms.

### 2.1 Dataset and Preprocessing

This study used a diabetes dataset containing 392 entries, including eight input features and one target variable. Each entry corresponds to an individual, with features that encompass demographic and clinical information pertinent to diabetes prediction. To maintain data integrity, the dataset was cleaned by removing all missing or null values during preprocessing. The features include:

- Pregnancies: Number of times the patient has been pregnant (Integer).
- Glucose: Plasma glucose concentration (mg/dL) (Integer).
- **BloodPressure:** Diastolic blood pressure (mm Hg) (Integer).

- SkinThickness: Triceps skinfold thickness (mm) (Integer).
- **Insulin:** Serum insulin level (mu U/ml) (Integer).
- **BMI:** Body Mass Index (kg/m<sup>2</sup>) (Float).
- DiabetesPedigreeFunction: A score estimating diabetes risk based on family history (Float).
- Age: Patient's age (Integer).
- **Outcome:** Binary target variable indicating diabetes diagnosis (1 = diabetic, 0 = non-diabetic).

The dataset underwent preprocessing to eliminate missing values, ensuring only complete records were included for model training. Since all features were numerical, there was no need for encoding categorical variables. Furthermore, feature scaling was conducted through standardization to align the features on a comparable scale, which is vital for models like LR and NB. The dataset was randomly divided into training and testing sets in an 80-20 ratio for model validation. This division facilitated training on a broad spectrum of data points while keeping a distinct set for evaluating performance. The training set was utilized to develop and fine-tune the models, whereas the testing set offered an impartial evaluation of their generalization abilities. The dataset was carefully chosen for its significant links to diabetes prediction, incorporating features typically associated with the disease. Preprocessing aimed to enhance model accuracy and computational efficiency, establishing a solid basis for assessing the performance of LR, NB, and DT classifiers. Table 1 depicts the dataset's descriptive statistics show each feature's mean, standard deviation, minimum, and maximum values, illuminated by the data's distribution and variability.



Figure 1. Workflow of the proposed system

 Table 1. Descriptive statistics of the dataset

Feature	Count	Mean	Std	Min	25%	50%	75%	Max
Pregnancies	392	3.3	3.21	0	1	2	5	17
Glucose	392	122.63	30.86	56	99	119	143	198
BloodPressure	392	70.66	12.5	24	62	70	78	110
SkinThickness	392	29.15	10.52	7	21	29	37	63
Insulin	392	156.06	118.84	14	76.75	125.5	190	846
BMI	392	33.09	7.03	18.2	28.4	33.2	37.1	67.1
DiabetesPedigreeFunction	392	0.52	0.35	0.085	0.2697	0.4495	0.687	2.42
Age	392	30.86	10.2	21	23	27	36	81
Outcome	392	0.33	0.47	0	0	0	1	1

The dataset underwent preprocessing to eliminate any missing values, guaranteeing that only complete records were utilized for training the model. Since all features were numerical, there was no need to encode categorical variables. Furthermore, feature scaling was implemented through standardization to ensure that all features were on a comparable scale, which is vital for models like LR and NB. This dataset was specifically chosen for its significance in diabetes prediction, incorporating features typically linked to the condition. The preprocessing aimed to enhance model accuracy and computational efficiency, thereby providing a solid groundwork for evaluating the performance of LR, NB, and DT classifiers.

## 2.1.1 Feature correlation analysis

A correlation heatmap was created using Pearson correlation coefficients to explore relationships between features (in subgraph (a) of Figure 2). This visualization facilitates the identification of highly correlated features, as such correlations can influence model performance. Glucose shows the strongest positive correlation with the Outcome variable (0.47), making it a critical predictor of diabetes. BMI and Age also demonstrate moderate correlations with diabetes risk, while SkinThickness and BloodPressure reveal weaker relationships.

## 2.1.2 Class distribution of Outcome variable

The dataset exhibits a moderate imbalance, comprising 67% non-diabetic individuals and 33% diabetic ones, as illustrated in subgraph (b) of Figure 2. This imbalance may affect the performance of ML models, especially their recall and precision for the minority class. To counter this, we emphasize model evaluation metrics like recall and F1-score to guarantee accurate identification of diabetic cases.

## 2.1.3 Feature importance analysis

The DT model evaluated the significance of each feature, illustrated in subgraph (c) of Figure 2 Glucose was the most crucial predictor, trailed by BMI, Age, and Insulin. Features like Pregnancies and DiabetesPedigreeFunction also played a moderate role, while BloodPressure and SkinThickness were less significant. These findings clarify the strong performance of LR, given that its assumptions are compatible with the dataset's linear connections.



(b)



Figure 2. (a) Feature correlation heatmap; (b) Class distribution of Outcome variable; (c) Feature importance (DT)

2.1.4 Statistical tests for feature significance

To determine the statistical significance of each feature in predicting diabetes, two tests were conducted:

- *Chi-square test for categorical features:* The Chi-square test was conducted to evaluate the association between the categorical feature of Pregnancies and the target variable Outcome. The test yielded a chi-square statistic of 8.46 with a p-value < 0.00001, indicating a highly significant relationship between the number of pregnancies and the diagnosis of diabetes.
- Independent t-tests for numerical features: Independent t-tests were conducted to compare the means of diabetic and non-diabetic groups for numerical features. The results, summarized in Table 2. It shows that all features exhibit statistically significant differences between the two groups, indicated by their low p-values (all below 0.001), suggesting their importance in predicting diabetes.

Feature	t-statistic	p-value	Significance
Pregnancies	4.6	$7.62 \times 10^{-6}$	Highly significant
Glucose	11.15	$3.72 \times 10^{-23}$	Highly significant
BloodPressure	3.76	$2.13 \times 10^{-4}$	Significant
SkinThickness	5.37	$1.68 \times 10^{-7}$	Highly significant
Insulin	5.73	$3.43  imes 10^{-8}$	Highly significant
BMI	5.56	$6.80 \times 10^{-8}$	Highly significant
DiabetesPedigreeFunction	3.82	$1.75 \times 10^{-4}$	Significant
Age	6.99	$3.10 \times 10^{-11}$	Highly significant

Table 2. T-test results for numerical features

The Glucose feature stands out with the lowest p-value  $(3.72 \times 10^{-23})$ , reinforcing its critical role in diabetes prediction. Features like Age, BMI, Insulin, and SkinThickness also show significant differences, justifying their inclusion in the prediction models. Although BloodPressure and DiabetesPedigreeFunction have slightly higher p-values, their values are still below the standard threshold of 0.05, indicating they are relevant predictors. These statistical tests validate the importance of each feature, supporting their use in ML models for diabetes prediction. The results also align with the feature importance analysis (in subgraph (c) of Figure 2), where Glucose, BMI, and Age emerged as the most influential predictors.

## 2.2 Naïve Bayes Algorithm

NB is a classification approach based on Bayes' Theorem that operates under the assumption that all features are conditionally independent based on the class label (either diabetes or no diabetes). Despite this strong dependency assumption, it is well-regarded for its simplicity, computational speed, and effectiveness, particularly with high-dimensional datasets standard in medical diagnostics. The classifier calculates the posterior probability of diabetes in a specific individual by multiplying the probabilities of the observed feature values given the diabetes status and adjusting for the prior probability of diabetes. NB delivers quick and reliable predictions if the features are reasonably independent, making it a suitable choice for datasets where this assumption holds effectively (Vyas et al., 2023).

The NB algorithm, as shown in Figure 3, begins with collecting and preprocessing the dataset to ensure cleanliness and readiness for analysis. The data is then divided into training and testing sets for model evaluation. A frequency table is subsequently generated, which counts the occurrences of various features (like words in text classification) for each class. After structuring the data, the algorithm calculates each class's prior probability based on its frequency in the dataset. It then determines the likelihood of each feature belonging to a class by examining how often it appears. Employing Bayes' Theorem, the algorithm computes the probability of a specific data point belonging to each class, and the class with the highest probability is assigned as the predicted category. Finally, the model is assessed using performance metrics such as accuracy, precision, and recall; adjustments like smoothing techniques may also be applied to enhance predictions.



Figure 3. NB classification process

# 2.3 Decision Tree Classification

DTs are valuable for classification tasks because they are straightforward to interpret and can be pruned effectively to avoid overfitting. By employing feature splits to construct tree structures, these models enhance comprehension of the decision-making processes involved in distinguishing between classes, such as types of crops or the diagnosis of diabetes. When utilizing a DT for diabetes prediction, the dataset is recursively divided based on the feature values that most clearly differentiate the classes (diabetes vs. no diabetes). Each node in the tree represents a decision point identified by a specific characteristic, such as a blood glucose level. The tree structure helps clarify how decisions are made and highlights which features significantly impact the predictions (Jiny D et al., 2024; Kumar et al., 2023).

Figure 4 depicts the DT classification algorithm; it starts by analyzing the dataset to identify the optimal feature for splitting the data according to a specified criterion, such as information gain or Gini index. Based on this feature, the dataset is divided into subsets, creating branches representing various possible outcomes. Each subset's recursive process is repeated, selecting the most relevant feature to build a tree structure at each step. As the tree grows, it continues to split until it fulfils a stopping criterion, such as when all data points in a node belong to the same class or further splits do not improve classification. After constructing the tree classifies new data points by traversing from the root node to a leaf node based on feature values. Finally, the model is evaluated using performance metrics like accuracy and precision, and techniques such as pruning may be used to enhance efficiency and help prevent overfitting (Kaur et al., 2023; Kiran et al., 2024).



Figure 4. Structure of a DT for classification

## 2.4 Logistic Regression

A LR model is employed for binary classification problems where the response variable can take on two distinct outcomes. During training, it learns the coefficients for each feature, which reflect the impact of these factors on the probability of the binary outcome, such as the presence or absence of diabetes. For instance, a higher glucose coefficient signifies a stronger positive correlation between elevated blood sugar levels and the risk of developing diabetes. One of the key benefits of this model is its straightforward interpretation, allowing healthcare professionals to make informed decisions based on the insights from its coefficients. LR proves especially beneficial in medical settings, as comprehending the connections between input features and health outcomes is vital for effective patient care and decision-making. This is due to its simplicity, the ability to provide probabilistic predictions, and the ease with which it can be interpreted (Arslankaya & Yaşli, 2024).

Figure 5 illustrates how LR functions as a classification algorithm. It starts with multiple input features, each of which shapes the prediction. These features receive weights reflecting their significance in the decision-making process. In addition to the weights, a bias term is incorporated to fine-tune the model's output. All these weighted inputs are aggregated to create a linear combination. This combined value is then processed through the sigmoid function, which transforms it into a probability ranging from 0 to 1. This probability indicates the chance of an instance belonging to a specific class. Ultimately, the model employs a threshold, usually set at 0.5, to categorize the output into one of two classes. When the probability exceeds the threshold, the instance is classified as belonging to one class; otherwise, it is assigned to the other. This visualization demonstrates how LR converts raw input data into classification outcomes.



Figure 5. Flow of LR for classification

#### **2.5 Performance Metrics**

The ML models' performance for diabetes prediction was evaluated using three key metrics: precision, recall, and F1-score (Thakur et al., 2024; Yang et al., 2024). Precision measures the accuracy of optimistic predictions by calculating the proportion of correctly identified diabetic cases out of all predicted positives, helping to minimize false alarms, refer to Eq. (1). Recall, also known as sensitivity, assesses the model's ability to detect all diabetic cases, ensuring fewer instances go undiagnosed, refer to Eq. (2). The F1-score balances precision and recall by providing a single performance measure, making it particularly useful when dealing with imbalanced datasets, refer to Eq. (3). Together, these metrics comprehensively evaluate the models' effectiveness in predicting diabetes accurately (Modak & Jha, 2024).

$$Precision = \frac{True \ Positive \ (TP)}{True \ Positive \ (TP) + False \ Positive \ (FP)}$$
(1)

$$Recall = \frac{True \ Positive \ (TP)}{True \ Positive \ (TP) + False \ Negatives \ (FN)}$$
(2)

$$F1 - Score = \frac{precision * recall}{precision + recall}$$
(3)

# 3. Results and Discussion

This section thoroughly assesses the effectiveness of three ML models, LR, NB, and DT, for predicting diabetes. The evaluation is based on performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC analysis to gauge their ability to differentiate between diabetic and non-diabetic cases. The findings offer insights into each model's classification accuracy and dependability, emphasizing their strengths and weaknesses. Furthermore, the analysis addresses the computational efficiency of these models to evaluate their practicality for real-world use. The ROC Curve Analysis provides an additional perspective on the discriminative power of each model, allowing for a comprehensive comparison of their classification abilities. This analysis aims to identify the most effective ML model for diabetes prediction and explores opportunities for enhancing predictive accuracy and robustness in healthcare applications.

#### 3.1 Accuracy Comparison of ML Models

Table 3. Accuracy comparison of ML models for diabetes prediction



Figure 6. Comparison of ML models

Accuracy is a key metric for assessing the performance of classification models. It indicates the ratio of correctly classified instances to the overall number of cases in the dataset. A summary of the accuracy scores for the three models is presented in Table 3 and also shown in Figure 6.

Among the three models assessed, LR demonstrated the highest accuracy at 78%, followed by NB at 75%, and DT at 71%. The strong performance of LR indicates its ability to generalize effectively to unseen data, ensuring stable predictions for diabetes classification. The slightly lower accuracy of NB can likely be explained by its assumption of independence among features, which may not be valid for medical datasets. DT's lower accuracy may result from overfitting the training data, impairing its generalization ability.

# 3.2 Precision, Recall, and F1-Score Evaluation

In addition to accuracy, precision, recall, and F1-score offer deeper insights into the model's ability to accurately classify diabetic (positive class) and non-diabetic (negative class) patients. The class-wise performance metrics for each model are explained in Table 4.

	Class	Precision	Recall	F1-Score	Support
ID	0 (Non-diabetic)	0.81	0.88	0.84	52
LK	1 (Diabetic)	0.73	0.59	0.65	27
рт	0 (Non-diabetic)	0.75	0.83	0.79	52
DI	1 (Diabetic)	0.59	0.48	0.53	27
ND	0 (Non-diabetic)	0.8	0.83	0.81	52
ND	1 (Diabetic)	0.64	0.59	0.62	27

Table 4. Performance of LR, DT, and NB

LR showed impressive results in identifying non-diabetic cases, achieving a precision of 81% and a recall of 88%, which led to a high F1-score of 84%. In contrast, its performance for diabetic cases was weaker, with a precision of 73% and a recall of 59%, revealing that some diabetic cases were incorrectly classified as non-diabetic. The DT model lagged behind LR, recording lower precision (75%) and recall (83%) for non-diabetic cases. Regarding diabetic patients, its precision fell to 59%, and recall declined further to 48%, highlighting its difficulty in accurately detecting diabetes cases and resulting in an F1-score of 53%. NB struck a middle ground between LR and DT, achieving precision (80%) and recall (83%) for non-diabetic cases. For diabetic cases, precision (64%) and recall (59%) slightly surpassed those of the DT but remained below those of LR.

## **3.3 Comparison Summary**

LR achieved the highest F1-scores for both diabetic (65%) and non-diabetic (84%) cases, showcasing its effectiveness in balancing precision and recall. NB yielded moderate results, outperforming DT but falling short of LR in overall classification effectiveness. DT showed the lowest performance, especially with diabetic cases, where its recall was the lowest at 48%, meaning it missed nearly half of the actual positive cases. This implies that although the DT can detect intricate patterns, it may experience overfitting or high variance, rendering it less dependable for diabetes prediction than the other two models, refer to Figure 7.



Figure 7. Performance metrics comparison for diabetic class

## 3.4 ROC Curve

The Receiver Operating Characteristic (ROC) curve is essential for assessing classification model performance. It visualizes the True Positive Rate (Sensitivity/Recall) against the False Positive Rate (1 - Specificity) across various threshold levels. The AUC score quantifies a model's effectiveness in distinguishing between diabetic and non-diabetic cases. An AUC closer to 1.0 indicates superior classification ability, while 0.5 reflects random guessing (Xiong et al., 2024).



Figure 8. ML models: ROC curve

In evaluating the three ML models, LR, NB, and DT achieved an AUC score of 1.0, denoting flawless classification. The ROC curves for each model reached the top-left corner of the graph, signifying that they accurately identified diabetic and non-diabetic cases without any errors. Although these results highlight outstanding model performance, achieving such perfect scores is uncommon in real-world situations and may indicate overfitting to the training data.

Among these models, LR excelled with an AUC score of 1.0, showing its capacity to differentiate between diabetic and non-diabetic cases effectively. NB, known for its probabilistic framework, also garnered a perfect AUC score, indicating that the dataset's features aligned well with its assumptions. Though often susceptible to overfitting, the DT model also achieved an AUC of 1.0, underscoring its ability to learn from the dataset effectively. However, DTs can be sensitive to changes in training data and might need techniques such as pruning or ensemble methods to ensure stability when applied in real-world settings.

Figure 8 demonstrates robust predictive capabilities, interpreting perfect AUC scores requires caution. While these results demonstrate robust predictive capabilities, interpreting perfect AUC scores requires caution. AUC scores in practical scenarios typically range from 0.7 to 0.99, as real-world datasets often encompass noise, overlapping class distributions, and intricate feature relationships. Suppose a model consistently scores an AUC of 1.0 on training datasets yet performs poorly on unseen test data. In that case, it likely indicates overfitting and may not generalize as well to new cases. Although the ROC curve analysis suggests that all three models are highly effective for predicting diabetes, further testing on independent datasets and cross-validation is crucial for confirming their generalizability. A balanced strategy that includes regularization techniques, hyperparameter tuning, and diverse data sources would help ensure that the models remain consistent and reliable for real-world medical assessments.

### 3.5 Computational Efficiency and Complexity

Computational efficiency is vital when assessing ML models, as it affects training time, prediction speed, and resource usage. A model with high computational efficiency is especially important in real-time healthcare scenarios, where quick and precise predictions are essential for making timely decisions, refer to Figure 9.

Due to its linear approach, LR is the most computationally efficient of the three models. It demands little processing power, making it ideal for extensive healthcare applications and real-time medical diagnosis. With its reliance on basic matrix operations and absence of iterative decision-making, LR achieves rapid computations with low memory requirements.

NB is also lightweight in computation since it directly calculates probabilities from the dataset's features instead of engaging in extensive iterative calculations. This attribute makes NB especially effective for large datasets with

categorical information, allowing for swift classification without excessive processing time. However, while it operates faster than more complex models, its presumption of feature independence might reduce its accuracy in some situations.



Classification Algorithm

Figure 9. Computational efficiency of ML models

<b>Fable 5.</b>	Summary of	performance	analysis: ML	models

<b>Evaluation Metric</b>	LR	NB	DT	
Accuracy	0.78	0.75	0.71	
Precision (Diabetic class)	0.73	0.64	0.59	
Recall (Diabetic class)	0.59	0.59	0.48	
F1-score (Diabetic class)	0.65	0.62	0.53	
Computational officianay	High	Moderate	Low	
Computational efficiency	(Fastest model)	Widderate	(Slowest model)	
Interpretability	High	Moderate	Highest	

The DT is the least computationally efficient model among the three. It involves numerous splits, recursive calculations, and potential overfitting problems that require pruning to improve generalization. Consequently, DT models operate more slowly and utilize more memory than LR and NB. Nevertheless, although DTs offer substantial interpretability, their computational requirements may pose challenges in large-scale or real-time medical environments.

The overall comparative performance of these three models is detailed in Table 5, which shows that LR consistently surpasses DT and NB in accuracy, F1-score, and computational efficiency. While DT excels in interpretability, it falls short in computational efficiency and recall rates for diabetic cases. NB presents a balanced option, providing moderate accuracy and efficiency, but may struggle with correlated features. This analysis emphasizes that LR is the most practical choice for diabetes prediction. It offers an optimal blend of accuracy, efficiency, and interpretability, making it exceptionally suited for application in healthcare settings.

# 3.6 Justification and Discussion: Why One ML Model Outperforms Another

The varying performance of LR, NB, and DT models in predicting diabetes is linked to their fundamental mathematical principles, feature assumptions, and capacity for generalization. According to the evaluation metrics, LR was the top-performing model, boasting superior accuracy, recall, and F1-score relative to the other models. This advantage primarily stems from its effectiveness in modelling the relationships between features and the target variable without succumbing to overfitting.

## 3.6.1 LR vs. DT

LR surpassed DT in performance due to its linear decision boundary and regularization ability. In predicting diabetes, essential elements like Glucose levels, BMI, and Age show close to linear relationships with the disease's outcome. Because LR is tailored to model linear relationships, it adeptly captures the general trend of the dataset while avoiding overfitting. Regularization methods (like L1 and L2) ensure that LR does not learn unnecessarily complex patterns, improving new data performance.

DTs are prone to overfitting training data, especially when tree depth is not adequately managed. DTs divide data into hierarchical structures, often producing overly specific rules that may fail to generalize. This was highlighted by the model's low recall rate (48%) for diabetic cases, indicating that almost half of the actual diabetic instances were incorrectly classified as non-diabetic. The intricate nature of DTs can also result in high variance, where minor data changes cause significant variations in model performance, rendering them less reliable for practical medical applications.

# 3.6.2 LR vs. NB

LR surpasses NB primarily due to its approach to feature dependencies. Unlike NB, LR does not presume that features are independent, which allows it to effectively model intricate relationships among diabetes risk factors like Glucose levels, BMI, and Insulin levels. This adaptability contributes to its superior accuracy of 78% and a better balance between precision and recall. Moreover, LR provides probability scores that permit refined decision thresholds, enhancing the reliability of classifications.

NB operates under the assumption of feature independence, a premise that rarely holds in real-world healthcare datasets. This limitation hinders its ability to recognize interactions among interconnected medical attributes. Although NB achieved a moderate accuracy of 75%, its recall rate for diabetic cases stood at 59%, which is lower than that of LR, suggesting challenges in accurately identifying positive cases. Nonetheless, its computational efficiency and straightforwardness make it suitable for quick predictions, especially when working with large datasets.

# 3.6.3 DT vs. NB

In the comparison between DT and NB, the latter achieved a slightly higher recall of 59% for diabetic cases, while DT lagged at 48%. This suggests that NB effectively identified more diabetic patients than DT. However, DT excelled in interpretability, offering straightforward decision-making rules that medical professionals can easily comprehend. Nonetheless, its inclination to overfit resulted in unstable performance, which diminished its reliability for medical diagnosis.

# 3.6.4 Final justification: Why LR is the best choice

When assessing accuracy, interpretability, generalization capabilities, and computational efficiency, LR emerges as the leading option for predicting diabetes. It maintains a reliable balance between precision and recall, accurately classifying diabetic and non-diabetic cases while exhibiting consistent performance across various datasets.

- *Best accuracy* (78%): Outperformed both NB (75%) and DT (71%).
- *Best recall and F1-score for diabetic cases:* Shows better sensitivity in identifying diabetic cases than DT and NB.
- Better generalization: Unlike DTs, avoids overfitting and handles feature dependencies better than NB.
- *Computational efficiency:* Faster and more resource-efficient than DTs, making it practical for real-time medical applications.

Selecting a ML model relies on the dataset's characteristics, the relationships among features, and the real-world application requirements. LR is the optimal choice for predicting diabetes because of its excellent classification accuracy, robustness, and effectiveness in managing structured medical data.

# 4. Practical Implications for Diabetes Prediction

Utilizing ML models for predicting diabetes offers substantial real-world benefits for healthcare. It facilitates early detection, effective patient management, and shaping public health policies. With diabetes posing an ongoing global health crisis, incorporating ML-driven predictive models can transform diabetes diagnosis, monitoring, and management, thereby enhancing patient outcomes and alleviating the strain on healthcare systems.

- *Early detection and risk assessment:* One of the primary advantages of using ML models for diabetes prediction is their ability to identify at-risk individuals before serious symptoms develop. In contrast, traditional diagnostic methods, including blood tests and clinical evaluations, depend on healthcare facilities, trained professionals, and significant resources, which may be scarce, particularly in rural or disadvantaged areas. ML models, such as LR, NB, and DTs, can be integrated into screening tools that utilize patient data (Age, BMI, Glucose levels, Insulin levels, family history, etc.) to evaluate diabetes risk. These models enable automated, real-time predictions, assisting healthcare providers in making early interventions to prevent disease progression. Telemedicine and mobile health applications can incorporate ML algorithms to offer remote diabetes risk assessments, minimizing the necessity for frequent hospital visits. By facilitating early detection, ML-based prediction systems can help mitigate complications associated with advanced diabetes, such as CVDs, kidney failure, nerve damage, and vision impairment.
- Personalized treatment and precision medicine: ML models not only predict diabetes but also

significantly aid in tailoring treatment plans according to individual patient characteristics. Each diabetes patient reacts uniquely to medications, lifestyle adjustments, and dietary changes. Standard one-size-fitsall approaches often fall short for many individuals. To pinpoint personalized treatment possibilities, ML systems scrutinize extensive datasets gleaned from EHRs, wearable devices, and Continuous Glucose Monitoring (CGM) systems. Those with a heightened risk for diabetes can obtain customized lifestyle suggestions, such as personalized exercise routines, dietary plans, and medication schedules, enhancing adherence and overall effectiveness of treatment. Additionally, predictive analytics can assist doctors in identifying patients likely to develop complications, allowing for earlier preventative measures. By utilizing data-driven treatment methods, ML-based diabetes prediction models promote precision medicine, ensuring that patients receive optimal care tailored to their unique needs.

- *Improved CDSS:* ML models can enhance CDSS to help healthcare professionals achieve quicker and more precise diagnoses. ML-powered CDSS can process vast patient information, providing doctors with evidence-backed insights to support clinical choices. Healthcare providers can leverage ML-guided decision-making systems to confirm diagnoses and reduce misclassification risks. By evaluating patient histories and lab results, CDSS can forecast which patients need immediate medical attention, thus boosting clinical efficiency. These systems offer additional support for physicians, diminishing cognitive strain and improving decision-making precision. Furthermore, integrating ML models into hospital management systems can optimize workflow, ensuring that high-risk diabetes patients receive prompt medical care.
- **Public health benefits and cost reduction:** Utilizing ML models for predicting diabetes has significant implications for public health, allowing governments and healthcare organizations to allocate resources more effectively. Healthcare expenses related to diabetes represent a substantial burden on medical systems globally, and employing ML-driven early interventions can significantly diminish long-term healthcare costs. Authorities in healthcare can leverage ML-powered predictive analytics to pinpoint high-risk groups and craft-focused prevention initiatives. Policymakers can enact community-based programs aimed at lifestyle changes, informed by data insights derived from ML models. Insurance companies and healthcare providers can assess risk and create affordable insurance options using these models, ensuring high-risk patients access preventive care early on. By transitioning from reactive care—dealing with diabetes complications—to proactive measures aimed at prevention, ML-driven prediction models foster improved public health results and lessen the financial strain on healthcare systems.
- Integration with wearable and Internet of Things (IoT) devices: The surge of wearable health technology and IoT devices has opened up new avenues for real-time diabetes monitoring and forecasting. ML models can be embedded in smart devices, facilitating the ongoing observation of blood glucose levels, physical activity, heart rate, and dietary practices. Devices like smartwatches, fitness trackers, and glucose monitors gather live health data, which ML algorithms then analyze to uncover trends and anticipate diabetes development. Patients can receive immediate notifications if their risk level rises, allowing for prompt lifestyle changes or medical advice. Healthcare practitioners can observe high-risk patients remotely, decreasing the necessity for regular hospital visits and enhancing the management of chronic illnesses. By integrating AI-driven analytics into wearable devices, diabetes prediction becomes more attainable and forward-thinking, enabling individuals to take charge of their health.
- *Ethical considerations and challenges:* Although ML-based diabetes prediction offers several benefits, addressing ethical concerns and challenges is essential to guarantee safe and equitable usage. Data Privacy and Security: ML models depend on patient data that must be safeguarded against unauthorized access or exploitation. Implementing robust encryption and data anonymization methods is vital. Bias in Predictions: Lack of diversity in training datasets may result in biased outcomes from ML models, potentially leading to misdiagnosis in specific populations. It is essential to ensure that data is fair and representative. Model Interpretability: Both doctors and patients must comprehend how predictions are generated. ML models should explain their predictions clearly to foster trust and support informed decision-making. Regulatory Compliance: AI systems in healthcare must adhere to medical regulations and ethical standards to confirm they meet safety and reliability criteria.

Tackling these issues is crucial for accepting and integrating ML-driven diabetes prediction models in healthcare environments. ML can transform diabetes prediction and management, delivering substantial advantages in early detection, personalized therapies, clinical decision-making, and public health efforts. By incorporating ML models into screening tools, healthcare systems, wearable technology, and CDSS, the medical field can enhance patient outcomes, improve disease prevention techniques, and lower healthcare expenses. Nevertheless, to optimize the practical impact of ML-driven diabetes prediction, it is necessary to confront ethical, privacy, and regulatory obstacles while ensuring that models are interpretable, unbiased, and validated for clinical use. With ongoing innovations in artificial intelligence, IoT, and medical informatics, ML-based diabetes prediction is poised to significantly influence the future of preventive healthcare and chronic disease management.

## 5. Conclusions

This research highlighted the effectiveness of ML models LR, NB, and DT in predicting diabetes through a detailed dataset comprising demographic and clinical traits. By integrating feature correlation analysis, evaluating feature importance, and conducting statistical significance tests, the study formed a solid foundation for identifying relevant predictors. Notably, Glucose, BMI, and Age were identified as the most critical factors, reflecting their strong connections with diabetes results and substantial differences between diabetic and non-diabetic populations.

Among the three models assessed, LR achieved the best performance, boasting an accuracy of 78% and higher F1-scores, especially for diabetic instances. Its success is attributed to its linear modelling approach, which effectively captures the interplay between features and outcomes without risking overfitting. In comparison, NB attained a moderate accuracy of 75%, hindered by its premise of feature independence. The DT model, while very interpretable, recorded the lowest accuracy at 71% and encountered issues with overfitting, particularly concerning diabetic cases.

This study's visual and statistical analyses reinforced the model selection, ensuring that feature selection and preprocessing were justified by data-backed evidence. The ROC-AUC analysis further validated each model's robust discriminative ability; however, the high AUC scores hint at potential overfitting that must be addressed in future research through cross-validation and the use of external validation datasets.

From a practical standpoint, LR emerges as the most fitting model for practical healthcare applications due to its combination of accuracy, efficiency, and interpretability. Its computational simplicity makes it well-suited for inclusion in CDSS and wearable health devices, enhancing early detection, tailored treatment, and remote monitoring. This model could help mitigate complications, enhance patient outcomes, and ease the strain on healthcare systems by fostering proactive interventions. The key contributions include that the study compares three widely used ML models, enriched by feature significance analysis and visual interpretations.

LR is optimal for diabetes prediction due to its high accuracy, computational efficiency, and practical applicability. By integrating ML models into healthcare systems, early diagnosis and personalized diabetes treatment can become more accessible and valuable, ultimately improving individual and public health outcomes.

# 6. Limitations and Future Directions

Diabetes prediction using ML models encounters several obstacles, including data quality and bias stemming from insufficient diversity in datasets, which hampers generalization. Problems related to feature selection and correlation adversely affect performance; for instance, LR has difficulty with non-linearity, while NB operates under the feature independence assumption. Additionally, DTs risk overfitting, undermining their reliability in practical scenarios. It's crucial to tackle ethical issues such as data privacy, interpretability, and automation bias to ensure effective integration into healthcare.

Future studies should emphasize improving model robustness and generalizability by using more extensive, varied datasets to minimize bias. Exploring ensemble learning and deep learning techniques can boost prediction accuracy and enhance feature interactions. Developing interpretable AI methods to guarantee transparency in clinical decision-making is essential. Integrating real-time data from wearable devices with ML models can facilitate continuous monitoring and early detection. Lastly, it is crucial to establish ethical AI frameworks that address privacy, fairness, and regulatory compliance in healthcare applications.

#### **Data Availability**

The data used to support the research findings are available from the corresponding author upon request.

## **Conflicts of Interest**

The authors declare no conflict of interest.

# References

Acheampong, E., Adua, E., Obirikorang, C., Anto, E. O., Peprah-Yamoah, E., Obirikorang, Y., Asamoah, E. A., Opoku-Yamoah, V., Nyantakyi, M., Taylor, J., Buckman, T. A., Yakubu, M., & Afrifa-Yamoah, E. (2024). Predictive modelling of metabolic syndrome in Ghanaian diabetic patients: An ensemble machine learning approach. J. Diabetes Metab. Disord., 23(2), 2233-2249. https://doi.org/10.1007/s40200-024-01491-7.

- Ahmed, N., Ahammed, R., Islam, M. M., Uddin, M. A., Akhter, A., Talukder, M. A., & Paul, B. K. (2021). Machine learning based diabetes prediction and development of smart web application. *Int. J. Cogn. Comput. Eng.*, 2, 229-241. https://doi.org/10.1016/j.ijcce.2021.12.001.
- Alenezi, T. M., Sulaiman, T. H., & AbdelAziz, A. M. (2023). Applying machine learning models to electronic health records for chronic disease diagnosis in Kuwait. *Int. J. Adv. Comput. Sci. Appl.*, 14(11), 667-676. https://doi.org/10.14569/IJACSA.2023.0141168.
- Al-Shanableh, N., Alzyoud, M., Al-Husban, R. Y., Alshanableh, N. M., Al-Oun, A., Al-Batah, M. S., & Alzboono, M. S. (2024). Advanced ensemble machine learning techniques for optimizing diabetes mellitus prognostication: A detailed examination of hospital data. *Data Metadata*, *3*, 363. https://doi.org/10.56294/dm2024.363.
- Arslankaya, S. & Yaşli, G. S. (2024). Predicting diabetes disease using machine learning classification algorithms and comparison of algorithm performances. *Int. J. Comput. Exp. Sci. Eng.*, 10(2), 133-140. https://doi.org/10.22399/ijcesen.233.
- Choubey, D. K., Kumar, P., Tripathi, S., & Kumar, S. (2020). Performance evaluation of classification methods with PCA and PSO for diabetes. *Netw. Model. Anal. Health Inform. Bioinform.*, *9*, *5*. https://doi.org/10.1007/s13721-019-0210-8.
- Chowdhury, N. H., Reaz, M. B. I., Haque, F., Ahmad, S., Ali, S. H. M., Bakar, A. A. A., & Bhuiyan, M. A. S. (2021). Performance analysis of conventional machine learning algorithms for identification of chronic kidney disease in type 1 diabetes mellitus patients. *Diagnostics*, 11(12), 2267. https://doi.org/10.3390/diagnostics11122267.
- Das, S., Rahman, R., & Talukder, A. (2024). Determinants of developing cardiovascular disease risk with emphasis on type-2 diabetes and predictive modeling utilizing machine learning algorithms. *Medicine*, *103*(49), e40813. https://doi.org/10.1097/MD.00000000040813.
- Ebrahim, O. A. & Derbew, G. (2023). Application of supervised machine learning algorithms for classification and prediction of type-2 diabetes disease status in Afar regional state, Northeastern Ethiopia 2021. *Sci. Rep.*, *13*, 7779. https://doi.org/10.1038/s41598-023-34906-1.
- Garg, R. (2021). Accuracy comparison of the data mining classification techniques for the diabetic disease prediction. *Int. J. Healthc. Technol. Manag.*, *18*(3-4), 216-227. https://doi.org/10.1504/IJHTM.2021.119159.
- Hossain, M. I., Maruf, M. H., Khan, M. A. R., Prity, F. S., Fatema, S., Ejaz, M. S., & Khan, M. A. S. (2023). Heart disease prediction using distinct artificial intelligence techniques: Performance analysis and comparison. *Iran J. Comput. Sci.*, 6(4), 397-417. https://doi.org/10.1007/s42044-023-00148-7.
- Iparraguirre-Villanueva, O., Espinola-Linares, K., Flores Castañeda, R. O., & Cabanillas-Carbonell, M. (2023). Application of machine learning models for early detection and accurate classification of type 2 diabetes. *Diagnostics*, *13*(14), 2383. https://doi.org/10.3390/diagnostics13142383.
- Jader, R. F., Aminifar, S., & Abd, M. H. M. (2022). Diabetes detection system by mixing supervised and unsupervised algorithms. J. Stud. Sci. Eng., 2(3), 52-65. https://doi.org/10.53898/josse2022234.
- Jiny D, M., Navamani, G., Kumar, R., Stević, Ž., Karabašević, D., & Kumar, R. (2024). Cultivating efficiency: Streamlining food delivery logistics through graphical networks and eternal m-certified domination number. *Int. J. Ind. Eng. Prod. Res.*, 35(3), 33-43. https://doi.org/10.22068/ijiepr.35.3.2029.
- Kangra, K. & Singh, J. (2023). Comparative analysis of predictive machine learning algorithms for diabetes mellitus. *Bull. Electr. Eng. Informatics*, 12(3), 1728-1737. https://doi.org/10.11591/eei.v12i3.4412.
- Kaur, R., Kumar, R., & Aggarwal, H. (2023). Role of machine learning in sustainable manufacturing practices: An outline. In *International Conference on Intelligent Manufacturing and Energy Sustainability* (pp. 541-551). Singapore: Springer. https://doi.org/10.1007/978-981-99-6774-2\_48.

- Kiran, M. D., B R, L. Y., Babbar, A., Kumar, R., H S, S. C., Shetty, R. P., K B, S., L, S. K., Kaur, R., Alkahtani, M. Q., Islam, S., & Kumar, R. (2024). Tribological properties of CNT-filled epoxy-carbon fabric composites: Optimization and modelling by machine learning. *J. Mater. Res. Technol.*, 28, 2582-2601. https://doi.org/10.1016/j.jmrt.2023.12.175.
- Kumar, M., Bajaj, K., Sharma, B., & Narang, S. (2021). A comparative performance assessment of optimized multilevel ensemble learning model with existing classifier models. *Big Data*, 10(5), 371-387. https://doi.org/10.1089/big.2021.0257.
- Kumar, R., Channi, A. S., Kaur, R., Sharma, S., Grewal, J. S., Singh, S., Verma, A., & Haber, R. (2023). Exploring the intricacies of machine learning-based optimization of electric discharge machining on squeeze cast TiB<sub>2</sub>/AA6061 composites: Insights from morphological, and microstructural aspects in the surface structure analysis of recast layer formation and worn-out analysis. *J. Mater. Res. Technol.*, 26, 8569-8603. https://doi.org/10.1016/j.jmrt.2023.09.127.
- Kumar, V., Lalotra, G. S., Sasikala, P., Rajput, D. S., Kaluri, R., Lakshmanna, K., Shorfuzzaman, M., Alsufyani, A., & Uddin, M. (2022). Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques. *Healthcare*, 10(7), 1293. https://doi.org/10.3390/healthcare10071293.
- Lu, H., Uddin, S., Hajati, F., Moni, M. A., & Khushi, M. (2022). A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Appl. Intell.*, *52*(3), 2411-2422. https://doi.org/10.1007/s10489-021-02533-w.
- Maniruzzaman, M., Rahman, M. J., Ahammed, B., & Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. *Health Inf. Sci. Syst.*, 8(1), 7. https://doi.org/10.1007/s13755-019-0095-z.
- Marzouk, R., Alluhaidan, A. S., & El Rahman, S. A. (2022). An analytical predictive models and secure webbased personalized diabetes monitoring system. *IEEE Access*, 10, 105657-105673. https://doi.org/10.1109/ACCESS.2022.3211264.
- Mavrogiorgou, A., Kiourtis, A., Kleftakis, S., Mavrogiorgos, K., Zafeiropoulos, N., & Kyriazis, D. (2022). A catalogue of machine learning algorithms for healthcare risk predictions. *Sensors*, 22(22), 8615. https://doi.org/10.3390/s22228615.
- Misra, P. & Yadav, A. S. (2020). Improving the classification accuracy using recursive feature elimination with cross-validation. *Int. J. Emerg. Technol.*, *11*(3), 659-665.
- Modak, S. K. S. & Jha, V. K. (2024). Diabetes prediction model using machine learning techniques. *Multimed. Tools Appl.*, 83(13), 38523-38549. https://doi.org/10.1007/s11042-023-16745-4.
- Mridul, A. H., Ahsan, N., Alam, S. S., Afrose, S., Sultana, Z., & Kafi, M. T. M. (2024). Polycystic Ovary Syndrome (PCOS) disease prediction using traditional machine learning and deep learning algorithms. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, 16(3), 322-346.
- Mushtaq, Z., Ramzan, M. F., Ali, S., Baseer, S., Samad, A., & Husnain, M. (2022). Voting classification-based diabetes mellitus prediction using hypertuned machine-learning techniques. *Mob. Inf. Syst.*, 2022(1), 6521532. https://doi.org/10.1155/2022/6521532.
- Pati, A., Parhi, M., Alnabhan, M., Pattanayak, B. K., Habboush, A. K., & Al Nawayseh, M. K. (2023). An IoTfog-cloud integrated framework for real-time remote cardiovascular disease diagnosis. *Informatics*, 10(1), 21. https://doi.org/10.3390/informatics10010021.
- Rajput, D. S., Basha, S. M., Xin, Q., Gadekallu, T. R., Kaluri, R., Lakshmanna, K., & Maddikunta, P. K. R. (2022). Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India. J. Ambient Intell. Humaniz. Comput., 13(5), 2829-2840. https://doi.org/10.1007/s12652-021-03154-4.
- Ratta, P. & Sharma, S. (2024). A blockchain-machine learning ecosystem for IoT-based remote health monitoring of diabetic patients. *Healthc. Anal.*, *5*, 100338. https://doi.org/10.1016/j.health.2024.100338.

- Reddy, S. S., Sethi, N., & Rajender, R. (2020). Diabetes correlated renal fault prediction through deep learning. EAI Endorsed Trans. Pervasive Health Technol., 6(24), e4. https://doi.org/10.4108/eai.11-11-2020.166958.
- Samet, S., Laouar, M. R., Bendib, I., & Eom, S. (2022). Analysis and prediction of diabetes disease using machine learning methods. *Int. J. Decis. Support Syst. Technol.*, 14(1), 1-19. https://doi.org/10.4018/IJDSST.303943.
- Shen, J. Y., Chen, J. B., Zheng, Z. Q., Zheng, J. B., Liu, Z. R., Song, J., Wong, S. Y., & et al. (2020). An innovative artificial intelligence-based app for the diagnosis of gestational diabetes mellitus (GDM-AI): Development study. *J. Med. Internet Res.*, 22(9), e21573. https://doi.org/10.2196/21573.
- Singh, V., Asari, V. K., & Rajasekaran, R. (2022). A deep neural network for early detection and prediction of chronic kidney disease. *Diagnostics*, 12(1), 116. https://doi.org/10.3390/diagnostics12010116.
- Thakur, V., Kumar, R., Kumar, R., Singh, R., & Kumar, V. (2024). Hybrid additive manufacturing of highly sustainable Polylactic acid -Carbon Fiber-Polylactic acid sandwiched composite structures: Optimization and machine learning. *J. Thermoplast. Compos. Mater.*, 37(2), 466-492. https://doi.org/10.1177/08927057231180186.
- Vyas, A., Raman, S., Sen, S., Ramasamy, K., Rajalakshmi, R., Mohan, V., & Raman, R. (2023). Machine learningbased diagnosis and ranking of risk factors for diabetic retinopathy in population-based studies from South India. *Diagnostics*, 13(12), 2084. https://doi.org/10.3390/diagnostics13122084.
- Wong, W. K. M., Thorat, V., Joglekar, M. V., Dong, C. X., Lee, H., Chew, Y. V., Bhave, A., Hawthorne, W. J., Engin, F., Pant, A., Dalgaard, L. T., Bapat, S., & Hardikar, A. A. (2022). Analysis of half a billion datapoints across ten machine-learning algorithms identifies key elements associated with insulin transcription in human pancreatic islet cells. *Front. Endocrinol.*, 13, 853863. https://doi.org/10.3389/fendo.2022.853863.
- Xiong, Y., Liu, X. Y., Wang, Q., Zhao, L., Kong, X. D., Da, C. H., Meng, Z. H., Qu, L. L., Xia, Q. F., Liu, L. H., & Li, P. M. (2024). Machine learning-based prediction model for the efficacy and safety of statins. *Front. Pharmacol.*, 15, 1334929. https://doi.org/10.3389/fphar.2024.1334929.
- Yang, L., Kumar, R., Kaur, R., Babbar, A., Makhanshahi, G. S., Singh, A., Kumar, R., Bhowmik, A., & Alawadi,
  A. H. (2024). Exploring the role of computer vision in product design and development: A comprehensive review. *Int. J. Interact. Des. Manuf.*, 18(6), 3633-3680. https://doi.org/10.1007/s12008-024-01765-7.