

# Machine Learning Based Diabetes Mellitus Prediction for Precision Healthcare

Angeline Christobel Y<sup>1</sup>, Priya N<sup>2\*</sup>, V.Karthikeyan<sup>3</sup>

<sup>1</sup>Associate Professor& Head, Department of Computer Science, SRM Institute of Science and Technology, Faculty of Science and Humanities, Ramapuram, Chennai, India  
[angelinechristobel5@gmail.com](mailto:angelinechristobel5@gmail.com)

<sup>2</sup> Assistant Professor, Department of Computer Science, SRM Institute of Science and Technology, Faculty of Science and Humanities, Ramapuram, Chennai, India

Corresponding author: [priyanarayanasami.91@gmail.com](mailto:priyanarayanasami.91@gmail.com)

<sup>3</sup>Assistant Professor & Head, Department of Computer Applications, Hindustan College of Arts & Science, Chennai, India [vanikarthik1@gmail.com](mailto:vanikarthik1@gmail.com)

## Abstract

Early and precise detection of Type-2 Diabetes Mellitus (T2DM) can significantly improve patient outcomes and lower long-term healthcare costs. Machine learning (ML) classifiers have become useful decision-support tools for clinical prediction tasks. This study provides a systematic and reproducible comparison of three commonly used supervised classifiers such as Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM). These classifiers were evaluated on the Pima Indians Diabetes Dataset (PIMA). A thorough preprocessing pipeline is applied, which included median imputation of invalid zero values, z-score normalization, and a stratified 80/20 train-test split. Hyperparameter optimization using GridSearchCV with 5-fold stratified cross-validation was performed. The performance was assessed using six metrics like accuracy, precision, recall, F1-score, ROC-AUC, and training time. The experimental results show that Random Forest achieves the highest predictive accuracy at 81% and ROC-AUC at 86%. This is due to its ensemble variance-reduction mechanism. SVM with an RBF kernel achieved the highest recall at 75% and a comparable F1-score at 76%. Meanwhile, Decision Tree remained the fastest and most interpretable model. These findings provide practical guidance on which algorithms healthcare practitioners should choose and highlight the need to balance predictive performance with clinical interpretability and computational efficiency.

**Keywords:** Machine Learning, Diabetes Mellitus, Decision Tree, Random Forest, Support Vector Machine.

**How to cite this article:** Angeline Christobel Y, Priya N, Karthikeyan V. Machine Learning Based Diabetes Mellitus Prediction for Precision Healthcare. *Int J Drug Deliv Technol.* 2026;16(52s): 172-178. DOI: 10.25258/ijddt.16.52s.19

## 1. Introduction

Diabetes mellitus is a long-term, progressive metabolic disorder marked by persistent high blood sugar due to problems with insulin secretion, insulin action, or both. Its complications, including heart disease, kidney disease, eye problems, and nerve damage, account for a large share of global illness and early death. The International Diabetes Federation (IDF) states that over 530 million adults worldwide had diabetes in 2023, and this number is expected to rise sharply in the coming decades. Type-2 Diabetes Mellitus (T2DM), which makes up more than 90% of all diabetes cases, is largely preventable and responds well to early treatment, making timely screening a priority for public health. Traditional diagnostic methods, such as fasting plasma glucose, glycated hemoglobin (HbA1c), and oral glucose tolerance tests, are proven but resource-heavy, dependent on the operator, and often impractical in low-resource or high-volume screening situations. The growth of electronic health records and clinical data repositories offers a chance for data-driven predictive models to support these traditional methods. Machine learning (ML) algorithms can find complex patterns in multidimensional clinical data, allowing risk assessment without additional invasive tests. Among various ML classifiers, Decision Trees (DT), Random Forests (RF), and Support Vector Machines (SVM) are especially notable in medical

informatics. Decision Trees are preferred for their clarity and understandable rules, which is crucial in clinical contexts where accountability and transparency matter. Random Forests, as ensemble learners, reduce the risk of overfitting found in single trees by combining predictions from multiple independent estimators. Support Vector Machines, based on statistical learning theory, achieve reliable performance by maximizing the gap between different classes and work well in moderate-dimensional spaces with few training samples. The Pima Indians Diabetes Dataset (PIMA), created by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), is the most widely used benchmark for testing diabetes prediction algorithms. Its mix of body measurements, biochemical data, and demographic information, along with a clear binary outcome, makes it a great resource for comparing classifiers systematically. However, many studies often report inconsistent data processing methods, different validation protocols, and incomplete performance metrics, making it hard to compare results directly and reproduce findings.

This study aims to fill these gaps by thoroughly comparing DT, RF, and SVM on the PIMA dataset using a consistent experimental approach. Key contributions include: (i) a common preprocessing method with health-driven data filling; (ii) hyperparameter tuning through nested cross-validation;

\*Author for Correspondence: [vanikarthik1@gmail.com](mailto:vanikarthik1@gmail.com)

(iii) a detailed multi-metric assessment of discrimination ability (ROC-AUC), sensitivity to class imbalance (F1-score, Recall), and efficiency (training time); and (iv) an analysis of the trade-offs of each algorithm relevant to clinical use. This work is intended to provide a reproducible reference for undergraduates, clinical data scientists, and practitioners looking for evidence-based recommendations on choosing algorithms for diabetes screening.

## 2. Related work

A lot of research has been done over the past decade on the intersection of artificial intelligence and diabetes prediction. Many algorithms, including deep neural networks and traditional statistical models, have been studied in connection with different datasets and clinical settings. This review presents key contributions in a clear way, focusing on methodological developments, dataset features, and performance standards relevant to this inquiry. On the PIMA dataset, preprocessing has repeatedly shown to be a crucial factor in how well models perform. Both SVM and RF classification accuracy were found to be much improved by systematic normalization and feature scaling, as shown by Patro and Sahu [1]. This highlights how sensitive these algorithms are to the distribution and size of the input features. By directly comparing RF to individual classifiers on PIMA, Shaukat et al. [2] validated that RF performed better, with the authors attributing the improvement to the ensemble's inherent variance reduction.

Reza [3] investigated how kernel selection and hyperparameter optimization affected SVM performance. He found that the predictive results across medical datasets were much improved with fine-tuned regularization parameters. A concentrated comparison between RF and SVM was presented by

Priyatma [4] on PIMA. The results showed that RF was more resilient across many situations, but SVM needed rigorous calibration to reach competitive results. Results like this back up the general idea that data features and tuning technique determine performance, rather than a particular method having a monopoly. Considerations of fairness, interpretability, and clinical reproducibility are particularly important when deploying ML models for patient risk stratification; these issues were highlighted in a macro-level review of AI applications in healthcare by Rajkomar et al. [5]. A recurring theme in this investigation is the need to strike a balance between medical AI's predictive accuracy and its interpretability, as pointed out by Waughfa[6]. Advanced feature selection and ensemble optimization strategies have been utilized in recent investigations. To optimize RF hyperparameters on PIMA, Ali et al. [7] opted for evolutionary search algorithms, which include, DE, PSO and GA, resulting in ROC-AUC of 0.956 and an accuracy around 89.7%. RF accuracy barely reaches 76.3%, with neural networks slightly outperforming classic approaches (78.57% accuracy) in a global comparison of multiple algorithms by Alzboon et al. [8]. Explainability of deep learning system, by integration of a post-hoc interpretability framework, is achieved by Iftikhar et al. [9] revealing that interpretability and high accuracy do not necessarily need to be opposing forces. An interesting combination between feature selection and classifier selection that yield 91.5% accuracy for SVM has been reported by the author(s) [10], thus highlighting the benefit of performing both types of selection simultaneously.

This investigation defines an understandable, directly comparable, threshold on which subsequent investigations must base their comparison, focusing on experimental repeatability, uniform preprocessing and comprehensive multi-metric examination.

**Table 1. Summary of Key Related Work**

Author(s), Year	Journal	Dataset / Methods	Key Findings
Patro & Sahu (2023)	Journal of Medical Systems	PIMA; normalization, scaling, ML models	Preprocessing improved SVM and RF classification performance significantly
Shaukat et al. (2023)	Healthcare Analytics	PIMA; DT, RF, SVM	RF achieved superior accuracy compared to individual classifiers
Reza (2023)	Expert Systems with Applications	Medical datasets; SVM with kernel optimization	Hyperparameter tuning and kernel choice enhanced predictive outcomes
Priyatma (2023)	Future Generation Computer Systems	PIMA; RF vs. SVM	RF provided robust results; SVM competitive with careful tuning
Rajkomar et al. (2023)	Nature Medicine	Review; AI in healthcare	Highlighted fairness, interpretability, and reproducibility challenges

Author(s), Year	Journal	Dataset / Methods	Key Findings
Waghfa (2023)	Applied Intelligence	Review of ML for diabetes	Balance between accuracy and interpretability is essential in medical AI
Ali et al. (2025)	Int. J. of Environmental Sciences	PIMA; ANOVA F-test, PCA, GA, PSO, DE	RF with Differential Evolution: ~89.7% accuracy, ~0.956 ROC-AUC
Alzboon et al. (2025)	arXiv	PIMA; LR, DT, RF, KNN, NB, GB, NN	Neural Network outperformed others (78.57%); RF achieved ~76.3%
Iftikhar et al. (2025)	Applied Sciences	Deep learning with explainable AI	Interpretable DL model for diabetes with high accuracy
Author(s) (2025)	Frontiers in Medicine	PIMA; SVM, RF, KNN, NB with feature selection	SVM with feature selection achieved highest accuracy

### 3. Dataset and Problem Formulation

In an effort to better understand the factors that put Pima Native Americans of Arizona at risk for type 2 diabetes, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) first created the Pima Indians Diabetes Dataset (PIMA). Each of the 768 records in the dataset pertains to a female patient who is 21 years of age or older. If the patient does not acquire a diabetes diagnosis within five years after the clinical measurements, then the prediction target is a binary outcome variable. Several features of the dataset make it a realistic and challenging benchmark: (i) a moderate class imbalance, with about 35% of cases being positive (diabetic); (ii) attributes like blood pressure, skin thickness, and insulin have missing physiological values encoded as biologically implausible zeros; and (iii) there is a relatively compact feature space of eight clinical predictors. Dataset features, clinical importance, and measurement units are listed in Table 2.

**Table 2. Pima Indians Diabetes Dataset-Attributes and Descriptions**

Attribute	Description	Unit	Notes
Pregnancies	Number of times pregnant	Count	Indicator of reproductive history; relevant to diabetes risk
Glucose	Plasma glucose concentration at 2-hour oral glucose tolerance test	mg/dL	Strongest single predictor of diabetes onset
BloodPressure	Diastolic blood pressure	mm Hg	Cardiovascular risk factor; missing values encoded as zero
SkinThickness	Triceps skinfold thickness	mm	Indirect measure of body fat percentage; contains zero-imputed missing values
Insulin	2-hour serum insulin concentration	$\mu$ U/mL	Related to insulin resistance; high proportion of missing values
BMI	Body Mass Index ( $\text{kg}/\text{m}^2$ )	$\text{kg}/\text{m}^2$	Obesity indicator; critical diabetes risk factor
DiabetesPedigreeFunction	Genetic diabetes likelihood score based on family history	Dimensionless	Captures hereditary predisposition
Age	Age of the patient	Years	Risk increases with advancing age

Attribute	Description	Unit	Notes
Outcome	Binary diabetes diagnosis label	0 = No; 1 = Yes	Target variable for classification

Get the best discriminative accuracy and generalizability by learning a function  $f: X \rightarrow \{0, 1\}$  that maps each patient record to a diabetes status label (0 = non-diabetic; 1 = diabetic) from an eight-dimensional clinical feature vector  $x = (\text{Pregnancies}, \text{Glucose}, \text{BloodPressure}, \text{SkinThickness}, \text{Insulin}, \text{BMI}, \text{DiabetesPedigreeFunction}, \text{Age})$ .

#### 4. Methodology

The experimental pipeline was built to ensure repeatability and methodological soundness at all stages of development: data preparation, feature engineering, model training, hyperparameter tuning and evaluation.

##### 4.1 Data Preprocessing

Blood Pressure, Skin Thickness, Insulin, Body Mass Index (BMI), and Glucose characteristics with zero values were considered missing data due to their physiological implausibility. In order to stop data from leaking out, the column-wise median, which is calculated only on the training partition, was used to replace these entries. Because Insulin and Skin Thickness had skewed distributions, median imputation was used over mean imputation.

After imputation, with the use of z-score normalization (standardization), all continuous features were adjusted to have a zero mean and unit variance:  $x' = (x - \mu) / \sigma$ . Algorithms that are sensitive to feature scale, like SVM, won't give too much weight to qualities with big absolute magnitudes (like glucose and insulin) because to this change.

Using stratified random sampling, we were able to maintain the same class ratio in the training and test sets, with 80% of the data going into the training set and 20% into the test set. All cross-validation folds kept the stratification limitation.

##### 4.2 Hyperparameter Optimization

Tuning the hyperparameters was done on the training set using GridSearchCV, which is an exhaustive grid search method that incorporates 5-fold stratified cross-validation. Parameter grids that were investigated were as follows:

- Decision Tree (DT): criterion  $\in \{\text{gini}, \text{entropy}\}$ ; min\_samples\_leaf  $\in \{1, 2, 4, 8\}$ ; max\_depth  $\in \{3, 5, 7, 10, \text{None}\}$ .
- RF: n\_estimators  $\in \{100, 150, 200\}$ ; max\_depth  $\in \{5, 10, \text{None}\}$ ; max\_features  $\in \{\text{sqrt}, \text{log2}\}$ ; min\_samples\_leaf  $\in \{1, 2, 4\}$ .
- SVM: RBF kernel, with  $C \in \{0.1, 1, 10, 100\}$  and  $\gamma \in \{\text{scale}, \text{auto}, 0.01, \text{and } 0.001\}$ .

The ROC curve (ROC-AUC) was employed as the main scoring metric for model selection during cross-validation because it is robust, threshold-independent and handles class imbalance well, which is especially important for clinical risk prediction models.

##### 4.3 Machine Learning Algorithms

The selection of three classifiers that were representative of these different paradigms of learning, with a balance of advantages and disadvantages, was then made. These were:

###### 4.3.1 Decision Tree (DT)

A non-parametric supervised classifier known as a decision tree (DT) uses a greedy, recursive splitting strategy to divide the feature space. The feature and threshold that minimizes impurity, as evaluated by Gini Impurity ( $G = \sum p_i(1-p_i)$ ) or Information Gain ( $IG = H(\text{parent}) - \sum_k w_k H(\text{child}_k)$ ), are chosen by the algorithm at each internal node. The tree continues to develop until it reaches a final stopping point, which could be a maximum depth, a minimum leaf size, or the achievement of zero impurity gain.

The main benefit of decision trees (DTs) is that they are easy to understand and communicate with clinical stakeholders because each root-to-leaf path is a clear and comprehensible decision rule. But overfitting happens when unpruned trees pick up on random noise in the training set. Important regularization mechanisms include depth constraints and minimum leaf-sample thresholds. Decision trees are unstable as well; even little changes to the training data might cause the trees to take on very diverse shapes. Because of this, they are used as basis learners in ensembles that aim to reduce variance, such Random Forests.

###### 4.3.2 Random Forest (RF)

Random Forests are an ensemble learning method that constructs  $T$  decorrelated decision trees by combining two sources of randomization such as bootstrap aggregation (bagging) and random feature subspace sampling. Each tree  $t_i$  is trained on a bootstrap replicate  $D_i \subset D$ , and at each split, only a random subset of  $\sqrt{p}$  features (for  $p$ -dimensional data) is considered. The final prediction is obtained by majority voting across all  $T$  trees for classification:

$$\hat{y} = \arg \max_c \sum_{t=1}^T \mathbf{1}[t_t(x) = c]$$

The dual randomization strategy reduces variance without commensurately increasing bias, yielding more stable and accurate predictions than a single tree. Random Forests also provide a natural measure of feature importance via mean decrease in impurity or mean decrease in accuracy enabling identification of the most discriminative clinical variables. The primary limitations are increased computational cost ( $O(T)$  training complexity) and reduced single-model interpretability, since the ensemble decision cannot be reduced to a single transparent rule pathway.

### 4.3.3 Support Vector Machine (SVM)

One kind of classifier with its origins in statistical learning theory is the Support Vector Machine (SVM). The goal of training is to identify, given the regularization parameter  $C$ 's soft-margin constraint, the maximum-margin hyperplane  $H: \mathbf{w} \cdot \mathbf{x} + b = 0$  that divides the two classes with the highest geometric margin  $2/|\mathbf{w}|$ . It is easier to attain linear separability in a high-dimensional reproducing kernel Hilbert space, which the kernel trick implicitly maps inputs into, for non-linearly separable data. The ability to model non-linear decision boundaries and its proven effectiveness on tabular medical datasets led to the selection of the Radial Basis Function (RBF) kernel, which is defined as

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

SVMs minimize structural risk instead of just empirical risk, which allows them to achieve high generalization on small-to-medium datasets. With appropriately tuned  $C$  and  $\gamma$ , they are less likely to be overfit than decision trees. Nevertheless, the training complexity increases at a super-linear rate as the sample size increases, going from  $O(n^2)$  to  $O(n^3)$ , and the clinical interpretability of the final decision boundary, which is determined by a sparse set of support vectors in a high-dimensional kernel space, is limited.

### 4.4 Evaluation Metrics

Model performance was evaluated using six complementary metrics derived from the confusion matrix (TP: true positives; TN: true negatives; FP: false positives; FN: false negatives):

- Accuracy =  $(TP + TN) / (TP + TN + FP + FN)$  → overall correctness.
- Precision =  $TP / (TP + FP)$  → positive predictive value; minimizes false alarms.
- Recall (Sensitivity) =  $TP / (TP + FN)$  → true positive rate; minimizes missed diagnoses.
- F1-Score =  $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$  → harmonic mean balancing precision and recall.
- ROC-AUC → area under the Receiver Operating Characteristic curve; threshold-independent discriminative power.
- Training Time (seconds) → wall-clock time for model fitting; proxy for computational efficiency.

### 5. Results and Discussion

Table 3 summarizes the comparative performance of DT, RF, and SVM evaluated on the held-out test set (20% of 768 records) following hyperparameter optimization via 5-fold stratified cross-validation.

**Table 3. Comparative Performance of Classifiers on the PIMA Dataset**

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	ROC-AUC (%)	Train Time (s)
Decision Tree (CART)	77	73	69	71	77	0.03
Random Forest	81	79	74	76	86	2.14
SVM (RBF kernel)	80	78	75	76	85	0.87

#### 5.1 Overall Performance

Random Forest outperformed all other ensemble methods in terms of accuracy (81%), ROC-AUC (86%), and the ability to aggregate decorrelated trees to reduce prediction variance. This study confirms the results of previous PIMA investigations[2,7] and quantifies the advantage of ensemble learning over a single classifier by comparing the ROC-AUC of RF (86% accuracy) and DT (77% accuracy).

The clinical significance of SVM's (RBF kernel) top recall of 75% among the three models lies in the fact that recall, defined as the proportion of actual diabetics appropriately identified, is the most safety-critical metric in the context of diabetes screening, where missed diagnoses, or false negatives, have more serious clinical implications than false positives. When looking at overall discriminative power (ROC-AUC), RF has the upper hand, although both ensemble and kernel-based techniques balance recall and accuracy similarly, as shown by the near-parity in F1-score between 76% and 76% for RF and SVM, respectively.

#### 5.2 Interpretability vs. Performance Trade-Off

Even though DT was the worst performing (77% accuracy) compared to other models, its training time (2.14s) was only twice that of RF (around 0.03 s). Moreover than anything, DT has put forward a clear and verifiable process which medical experts could use for decision-making purposes and be presented to them. Even though DT is behind in terms of its accuracy by a little gap, the advantage of interpretability would balance out when dealing with lack of computational power, or regulatory transparency requirements (e.g. For EU AI Act).

The Random Forest uses aggregation of 100-200 trees to achieve its decision surface, and this cannot be reduced to a single rule that can be interpreted by humans. Given that age, BMI and glucose are also key clinical features; feature importance scores indicated age, BMI and glucose as three most discriminative variables (not shown); these provided an indirect form of interpretability and align well with the existing clinical knowledge. However, the interpretability is not straight-forward for SVM. SVM has its decision boundary set in a kernel induced feature space and therefore require post-hoc explainability methods like

SHAP and LIME to be applied before it can be used for clinical decision making.

### 5.3 Clinical and Deployment Implications

The three recommendations resulting from the experiment are (i) RF is the choice when both increase of prediction accuracy and robustness is desirable, (ii) SVM is the choice when high recall (sensitivity) is required and especially to be the best choice for population screening like the one related to disease diagnosis since it has no false negatives to avoid under-diagnosing cases, and (iii) DT is the best choice when interpretability of the model, regulatory compliance and computation power are important. Note that the three classifiers were trained on similar demographic information, i.e. Pima women living in Arizona, U.S.A. Whose ages were 21 and over. Hence, it is warranted that an external validation should be performed using subjects with different clinical features and ethnicities before the actual implementation in clinical practice. Also, because the statistical power is restricted by the relatively small size of the data set (768 records), the stated performance estimation values should be understood in conjunction with appropriate confidence intervals.

### 5.4 Limitations

Some limitations need to be discussed before we generalize the results obtained so far. Firstly, since only data from a cohort of only one ethnic female group has been used (PIMA data set), generalizability to other groups cannot be inferred. Secondly, median imputation, a method used to remedy measurement error, cannot account for all such issues; a zero imputation for missing Insulin and Skin Thickness values is one example of this. Third, there has been no external validation study included and as such our future works will also have to involve future clinical data obtained from external institutes. Furthermore, advanced ensemble methods such as gradient boosting or XGBoost have not been explored as well as deep learning architectures. Lastly, demographic bias and fairness have not been examined.

### 6. Conclusion

In this work, the benchmark Pima Indians Diabetes Dataset is used and systematically and reproducibly compared Decision Tree, Random Forest, and Support Vector Machine for the early prediction of Type 2 Diabetes Mellitus. The Random Forest classifier demonstrated better performance than any other algorithm under a unified preprocessing and evaluation setting, yielding 81% accuracy and 86% ROC-AUC. Based on its recall value of 75%, SVM (RBF kernel) can be used as a sensitive screening classifier to promote early detection and treatment. The Decision Tree classifier shows favorable train speed of 0.03s, and its interpretability can compensate for the mediocre accuracy of 77% in a clinical and controlled environment. This research systematically compares the performance of Decision Tree, Random Forest, and Support Vector Machine classifiers for the early

prediction of Type 2 Diabetes Mellitus using the standard Pima Indians Diabetes Dataset in the context of healthcare and pharmaceutical analysis. The Random Forest classifier, with its ensemble learning approach to variance reduction, achieved the best prediction performance in this comparison, with a resultant accuracy of 81% and a ROC-AUC of 86%, making it the ideal choice for clinical decision support systems in drug management and patient monitoring applications. The Support Vector Machine (RBF kernel) model, despite its lower accuracy, has a recall of 75%, thus it can serve as an effective screening tool to facilitate early diagnosis and treatment intervention in patient management. The Decision Tree model's high interpretability and remarkably fast training time of 0.03s are significant advantages in clinical decision-making when transparency is required. The results indicate that an algorithm's selection for specific healthcare and drug-delivery-related applications should be based on the required trade-off between accuracy, interpretability, computational speed and screening sensitivity, and that further investigation can include larger and heterogeneous clinical datasets, sophisticated feature engineering methods, deep learning architectures with explanations, intelligent drug-delivery support systems tailored for individual patient care in diabetes treatment, bias-aware and fairness-constrained predictive frameworks, and containerized, scalable machine learning pipelines.\

### References

- [1] R. Patro and A. Sahu, "Impact of preprocessing techniques on machine learning algorithms for diabetes prediction," *Journal of Medical Systems*, vol. 47, no. 2, pp. 112–120, 2023, doi: 10.1007/s10916-023-01987-3.
- [2] K. Shaukat, T. M. Alam, and M. U. Sarwar, "Comparative analysis of machine learning algorithms for diabetes prediction using PIMA dataset," *Healthcare Analytics*, vol. 3, no. 1, p. 100145, 2023, doi: 10.1016/j.health.2023.100145.
- [3] M. Reza, "Enhancements in support vector machine models for medical prediction," *Expert Systems with Applications*, vol. 221, p. 119749, 2023, doi: 10.1016/j.eswa.2023.119749.
- [4] Y. Priyatma, "Random Forest versus SVM: Comparative study for diabetes diagnosis," *Future Generation Computer Systems*, vol. 146, pp. 432–441, 2023, doi: 10.1016/j.future.2023.01.010.
- [5] A. Rajkomar, E. Chen, and D. Ouyang, "Artificial intelligence in healthcare: Opportunities and challenges," *Nature Medicine*, vol. 29, no. 5, pp. 1015–1025, 2023, doi: 10.1038/s41591-023-02360-7.
- [6] J. Waughfa, "Machine learning techniques for diabetes detection: A review," *Applied Intelligence*, vol. 53, no. 12, pp. 14567–14581, 2023, doi: 10.1007/s10489-023-04521-4.
- [7] A. A. Ali, G. R. Galal, and H. S. Hassan, "Diabetes prediction on Pima Indian dataset using machine learning techniques," *International Journal of*

- Environmental Sciences*, vol. 11, no. 7, pp. 529–550, 2025, doi: 10.64252/3a8wqx36.
- [8] M. S. Alzboon, M. Al-Batah, M. Alqaraleh, A. Abuashour, and A. F. Bader, “A comparative study of machine learning techniques for early prediction of diabetes,” *arXiv preprint arXiv:2506.10180*, 2025. [Online]. Available: <https://arxiv.org/abs/2506.10180>
- [9] K. Iftikhar, N. Javaid, I. Ahmed, and N. Alrajeh, “A novel explainable deep learning framework for accurate diabetes mellitus prediction,” *Applied Sciences*, vol. 15, no. 16, p. 9162, 2025, doi: 10.3390/app15169162.
- [10] “Advanced supervised machine learning methods for precise diabetes mellitus prediction using feature selection,” *Frontiers in Medicine*, vol. 12, p. 1620268, 2025, doi: 10.3389/fmed.2025.1620268.
- [11] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed., 2023. [Online].