International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 02 February 2026
DOI: 10.47191/ijcsrr/V9-i2-28, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

# Multiclass Diabetes Classification using Multimodal Artificial Intelligence

**Mohammed M. Al-Mhadawi[1], Ayman Hammoodi[2], Younis Kadthem Hameed[3], Qahtan M. Yas[4*]**

[1]General Directorate for Education of Diyala, Ministry of Education, Diyala, Baqubah, 32001, Iraq
[2]Computer Science Department, College of Veterinary Medicine, University of Diyala, Baqubah, 32001, Iraq
[3]Computer Science Department, Al-Muqdad of Education College, University of Diyala, Baqubah, 32001, Iraq
[4]Computer Science Department, College of Engineering for Artificial Intelligence Technology, University of Diyala, Baqubah, 32001, Iraq

**ABSTRACT:** Diabetes mellitus is a prevalent metabolic disorder globally. Its primary etiologies encompass socioeconomic determinants, behavioral risk factors, and underlying comorbidities. Numerous epidemiological studies have investigated various diabetes phenotypes, impacting both sexes across the entire age spectrum. This study utilizes a dataset containing clinical profiles of 1,000 subjects assessed on multiple biometric and sociodemographic variables. The objective is to classify diabetes into type 1, type 2, and prediabetes using an array of deep learning and machine learning algorithms. Currently, artificial intelligence-driven diagnostic methods represent a state-of-the-art approach for disease stratification. This research evaluates the performance of six classification algorithms for determining glycemic status: random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost), convolutional neural network (CNN), recurrent neural network (RNN), and long short-term memory (LSTM) network. Results demonstrate that the XGBoost classifier attained the highest predictive accuracy of 91% with a training duration of 20 seconds, surpassing the other models. These findings underscore the potential of advanced computational algorithms for precise diabetes phenotyping and risk assessment, offering significant implications for disease management and public health interventions.

**KEYWORDS:** Artificial Intelligence Models, Classification, Detection, Diagnosis, Multiclass Diabetes.

## 1. INTRODUCTION

Diabetes mellitus is a widespread disease caused by a chronic metabolic disorder, leading to a disruption in blood sugar regulation in those affected. This disease includes the first type, which is caused by the body's inability to produce insulin, while the second type is caused by the body's incapacity to react to insulin[1]. The percentage of people affected by this disease is very high, and it is considered a serious and fatal disease among humans. Morbidity and mortality rates have become higher among diabetics than in people without diabetes. Official statistics indicate that one in eight adults is expected to have diabetes, reaching 853 million, a 45% increase by 2050 [2][3].

Diabetes poses a global health challenge due to the rising number of affected individuals. The International Diabetes Federation (IDF) estimates that 589 million individuals between the ages of 20 and 79 suffer from diabetes [2][4]. People with diabetes typically suffer from various complications and an increased mortality rate. Rapid intervention to provide immediate healthcare for the prevention and early detection of this disease has become essential. Health reports indicate that 1 in 2 adults with diabetes is undiagnosed, which in turn poses a significant challenge to healthcare providers. However, almost 10% of the $760 billion spent on healthcare worldwide is related to diabetes [5][6]. Diabetes is classified into two primary types: type 1 and type 2. Type 1 diabetes, which is insulin-dependent, is a significant burden, especially for children and their families. In countries with weak economies, those with type 1 diabetes often struggle to access insulin or self-care tools, leading to acute and chronic complications, and sometimes death or severe disability [7] [8]. In contrast, type 2 diabetes accounts for over 90% of cases globally and is the most prevalent form. Type 2 diabetes, which typically develops later in life, is characterized by the body's cells not reacting fully to insulin, known as insulin resistance. This type is often less severe at onset and may lack symptoms, making diagnosis difficult. Extended undiagnosed type 2 diabetes can result in increasing complications over time [9][10].

Computer science uses advanced techniques to develop tools or systems for evaluating data and managing complexities in many applications. AI techniques can now be used for diabetes to digest data efficiently and create tools to manage the illness. Advances in AI are crucial in providing safer technology through secure and user-friendly designs. These designs also take into account

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 02 February 2026
DOI: 10.47191/ijcsrr/V9-i2-28, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

uncertainties in technical systems [11]. Six machine learning and deep learning algorithms as a RF [12], SVM [13], XGBoost [14], CNN [15][16], RNN [17], and LSTM [18][19] were used in this study to detect and classify diabetes. These models contributed to providing physicians and healthcare providers with the necessary information through intelligent programs that effectively manage diabetes. In contrast, the AI approaches offer opportunities for diabetics to care for themselves and make decisions in emergencies. These tools also give providers flexible monitoring options and resource use within healthcare systems [20].

## 2. RELATED WORKS

Diabetes primarily impairs blood sugar regulation and is a widespread global disease. Many studies have addressed its types using various methods, including artificial intelligence, to aid diagnosis and support physicians and patients.

Thaiyalnayaki, K. [21] applied MLP and SVM algorithms to classify diabetes mellitus. This approach provided optimal treatment management for a diabetic patient dataset. Nine key features were identified and analyzed. The MLP classifier attained a 77.474% accuracy for 595 instances, correctly classifying 77.474% and misclassifying 22.526%. In comparison, the SVM classifier achieved 65.1042% accuracy for 500 cases and misclassified 34.8958% of them. These results indicate that the MLP classifier outperformed the SVM in terms of accuracy.

Islam, Md Merajul, et al. [22] used various statistical measures to identify diabetes risk factors. Specifically, the values of independent variables were examined using the t-test, while the chi-square measure was applied to test categorical variables for the diabetes dataset. In their analysis, 1,569 people participated in the 2011 Population and Medical Survey of England, and it was found that 127 of them had diabetes. The authors then classified and predicted outcomes in the diabetes dataset by applying a set of machine learning algorithms. According to their findings, just eleven of the fifteen attributes were substantially related to diabetes. Notably, the RT classifier based on bagged CART was the most accurate, achieving a 94.3% accuracy rate with a value of 0.600 under the curve.

Kumari, S., Kumar, D., & Mittal, M.[23] developed a soft voting approach using several machine learning models, such as LR, RF, and NB. The study applied multiple ML algorithms to predict diabetes and aimed for the best accuracy. Researchers used the PIMA Diabetes Dataset, which includes data from people both with and without diabetes. They assessed baseline classifiers with criteria such as accuracy, precision, recall, and F1_score. Findings showed 73.48% accuracy, 79.04% precision, 71.45% recall, and an F1-score of 80.6%.

Gollapalli, Mohammed, et al. [24] evaluated diabetes types (Type 1, Type 2, and prediabetes) using machine learning models in classification and prediction. A Saudi Arabian hospital dataset was utilized to manage the illness and stop its development. Several algorithms were used in four different experiments to get the best outcomes. The significance of permuted qualities highlighted five important features: gender, education, insulin usage, nutrition status, and use of antidiuretic medication. For scenarios 2, 3, and 4, the data collection was balanced using the synthetic minority oversampling technique. The novel stacking model demonstrated encouraging empirical findings, achieving the highest accuracy of 94.48% in classifying diabetes.

Shams, M. Y., Tarek, Z., & Elshewey, A. M. [25] developed a novel approach for early-stage detection of diabetes. They utilized a variety of machine learning techniques and called it the Recurrent Feature Elimination Unit (RFE-GRU). The researchers used the PIMA Indian Diabetes Mellitus (PIDD) dataset, which contains 768 patients and 9 parameters. RFE is crucial for selecting features from the training dataset. GRU addresses the challenge of feature gradient disappearance and amplification. The results demonstrated that the RFE-GRU model delivered promising prediction results. It outperformed other classifiers. The evaluation scores, using accuracy, precision, recall, F1 score, and area under the curve (AUC), were 90.70%, 90.50%, 90.70%, and 0.9278, respectively.

## 3. RESEARCH METHODOLOGY

This section outlines the suggested approach implemented in this study. An open-source Kaggle dataset on diabetes was used [26]. Types of diabetes were identified and classified using an integration of DL and ML models. Figure 1 depicts the research process from data gathering to results acquisition.

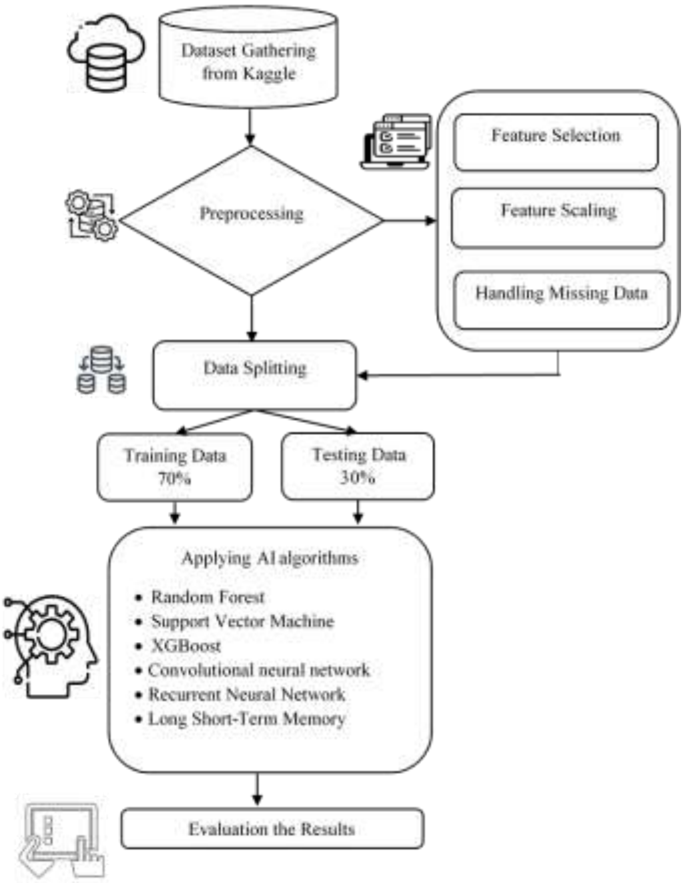


**Fig 1. Research tracking from data gathering to results acquisition**

### 3.1 Collocation of the Dataset

An invaluable resource for diabetes study and analysis, Kaggle provided the dataset. [26]. This website provides a variety of open-source clinical and demographic data, supporting the development of clinical projects and providing a deeper understanding of the disease. Complete medical records for 1,000 individuals with diabetes, separated into three groups (diabetic, non-diabetic, and pre-diabetic), are included in the dataset. The dataset includes various diabetes symptoms such as gender, age, urea, creatinine, anxiety, glycated hemoglobin, cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein, very low-density lipoprotein, body mass index, and class. The parameters were carefully distributed to include a wide range of symptoms, ensuring diversity and accuracy in the resulting models. As a result, this diversity has significantly contributed to the development of scientific methods, not only for identifying diabetes symptoms but also for the development of healthcare technologies based on ML and DL algorithms. Table 1, shows the description of the parameters of diabetes mellitus.

**Table 1. Description of the parameters of diabetes mellitus**

| Feature | Description |
|---|---|
| **Gender** | The biological sex of the individual. Usually encoded as: 0 = Female, 1 = Male. Gender can impact diabetes risk due to hormonal and lifestyle differences. |
| **AGE** | The age of the subject in years. Age is a critical risk factor as diabetes risk increases with age, especially after 45. |
| **Urea** | A measurement of urea (in mg/dL) in the blood. High levels may indicate kidney issues, which are common complications of diabetes. Normal range: ~7-20 mg/dL. |

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 02 February 2026
DOI: 10.47191/ijcsrr/V9-i2-28, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

| | |
|---|---|
| **Cr (Creatinine)** | Measures the level of creatinine in the blood (mg/dL). Also a marker of kidney function. Elevated Cr levels may suggest impaired kidney function, often linked with diabetes. The standard range:~0.6- 1.3 mg/dL. |
| **HbA1c (Glycated Hemoglobin)** | An important measure of typical levels of blood sugar for the previous two to three months. Shown as a percent. - Diabetic: >6.5%, Pre-diabetic: 5.7-6.4% , Normal: <5.7%. |
| **Chol (Cholesterol)** | Total cholesterol in the blood (mg/dL). A common complication of diabetes is high fat levels, which increases the likelihood of heart disease. Normal: <200 mg/dL. |
| **TG (Triglycerides)** | Measures the amount of fat in the blood (mg/dL). Hyperglycemia and resistance to insulin are associated with elevated levels. Normal: <150 mg/dL. |
| **HDL (High-Density Lipoprotein)** | The "good" cholesterol (mg/dL). Higher levels are better. Helps remove excess cholesterol from the bloodstream. - Ideal: >40 mg/dL (men), >50 mg/dL (Ladies). |
| **LDL (Low-Density Lipoprotein)** | The "bad" lipid level (mg/dL). High levels contribute to plaque buildup in arteries. - Optimal: <100 mg/dL. |
| **VLDL (Very Low-Density Lipoprotein)** | Another type of "bad" cholesterol (mg/dL), carries triglycerides. Often estimated from TG/5. High VLDL is associated with increased diabetes risk. Normal: 2-30 mg/dL. |
| **BMI (Body Mass Index)** | A body fat measurement determined by height and weight (kg/m2). Overweight (BMI ≥30) significantly increases the chance of getting diabetes with type 2: <18.5 - Normal: 18.5-24.9 - Overweight: 25-29.9 - Obese: >30 |
| **Class** | Target label: Indicates diabetes status. Typically coded as: 0 = Non-Diabetic, 1 = Diabetic, 2 = Predict-Diabetic. This is the outcome you're trying to predict or classify. |

### 3.2 Data Preprocessing

The data processing and cleaning phase is a fundamental and important step in the data analysis process. During this phase, the dataset is converted into a format suitable for analysis, modeling, and interpretation. This section describes all the specific steps followed based on the code applied to pre-process and clean the dataset, taking into account data type, handling zero values, transforming categorical variables, and preparing the data for modeling.

### 3.3 Feature Selection

At this stage, the process of selecting attributes (variables) for the dataset is carried out in two steps. In the first step, the dataset is initialized and loaded, and then the categorical attributes are converted into numerical attributes. This step enables the selected data to be properly processed using ML and DL algorithms. In the next step, the features are prepared, and the targeted characteristics and parameters are separated.

### 3.4 Data Splitting

This step of data processing is essential. The data is divided into two primary categories, training data and test data, during the data splitting process. The dataset used in this study was split into 70% training data and 30% test data.

### 3.5 Identify Artificial Intelligence Algorithms

A briefly introduces the AI algorithms with it's models as ML and DL used in this study. The selected traditional models were applied for classification and evaluation using training and test data [27][28]. Primary classifiers will be explained in detail.

### 3.6 Evaluation Metrics

The effectiveness of models is typically assessed using several significant measures. The confusion matrix is one of the most essential measures for assessing the performance of the classification models used in this study. It involves four key criteria (TP, TN, FP, and FN). The confusion matrix is typically sized 2 × 2 and includes axes for both the true and expected values. The parameters of the confusion matrix can be described in detail as follows [29]:

- True Positives (TP): positive elements that are correctly identified as positive,
- False Positives (FP): negative elements that are incorrectly identified as positive,

• False Negatives (FN): positive elements that are incorrectly identified as negative, and
• True Negatives (TN): negative elements that are correctly identified as negative.

These parameters can be used to calculate several measures, such as accuracy, precision, recall, and F-measure, which are extensively employed in data mining and machine learning to assess models. Table 1, shows a confusion matrix diagram [30].

**Table 2. Confusion matrix**

| | | Predicated Measures Positive =1 | Negative =0 |
|---|---|---|---|
| Actual Measures | Positive = 1 | TP | FP |
| | Negative = 0 | FN | TN |

- **Accuracy measure**

The accuracy measure indicates the success of a binary classification test in either ruling out or confirming a condition. Accuracy can be defined as the proximity to the target, or the average proximity to it. Typically, the weighted arithmetic mean of precision and inverse precision is shown as bias-weighted. It can also be represented as a prevalence-weighted scale of recall and inverse recall.[31]. The calculation of the accuracy measure can be represented as follows:

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

- **Precision measure**

One popular measure is the precision of test results, which indicates how consistently or efficiently information has been aggregated. A crucial criterion in the evaluation matrix is the precision metric, which can capture data regarding the kinds and frequencies of errors that could happen during training [32]. As a result, the positive instances and forecasts are the main emphasis of this measure's contribution. Therefore, the precision metric could be defined as the ratio of correctly categorized positive instances to all positive examples identified by the system:

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

- **Recall measure**

The recall measure is the most widely used metric for evaluating sensitivity. In addition, this metric represents the average total recall of the TPR measure. Thus, the ratio of positive true values to values that were categorized as positive is known as the recall measure. This metric and its group solely highlight predictions and positive examples. Consequently, the recall metric is crucial in the assessment matrix since it may capture data regarding the kinds and frequencies of errors that could happen upon training. As a result, the recall measure can be defined as the number of positive examples divided by the total positive example values [33]. The recall measure is represented by the following formula:

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

- **F-measure**

The F-measure is essential when evaluating examples that are imbalanced compared to others. This measure is often used to evaluate and improve information performance when the number of negatives in the sample increases compared to the number of positives. This measure also allows for the calculation of differential weighting between the recall and precision measures, although these measures usually have equal weight. To ensure a value close to the ideal, it is necessary to reflect the true positives, as this measure typically corresponds to the arithmetic mean of the expected true positive values [34]. The F-score is represented as follows:

$$F - measure = \frac{2*Prec*Rec}{Prec+Rec} \quad (4)$$

### 3.6.1 ROC curve

AUC, often known as the ROC curve, is a crucial statistic for figuring out how true and false positive rates relate to one another in binary classifiers. Many studies have adopted this metric because it is superior to absolute accuracy in classifier evaluation and is particularly popular for stationary, imbalanced data. However, the dataset must be sorted repeatedly for each example when measuring AUC [35]. However, when processing big datasets, AUC cannot be calculated directly because it requires a full scan of the streams after each iteration. Therefore, the AUC metric will be limited to use with data sources for full streams and periodic stop sets that skew its performance and are computationally infeasible, especially for real-world applications [36].

## 4. RESULTS AND DISCUSSIONS

The findings from the experiments carried out for this study will be examined. Figure 2, shows the outcomes of all techniques that were used, based on the ROC-AUC curve and confusion matrix. These results are essential for calculating the remaining metrics in the next stage for all classifiers, which differ according to the rules each algorithm uses to classify and detect diabetes types in patients.

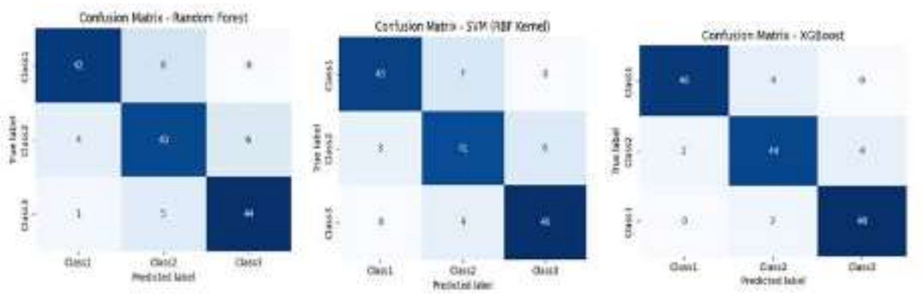


**Fig 1. Confusion matrix results for ML models**

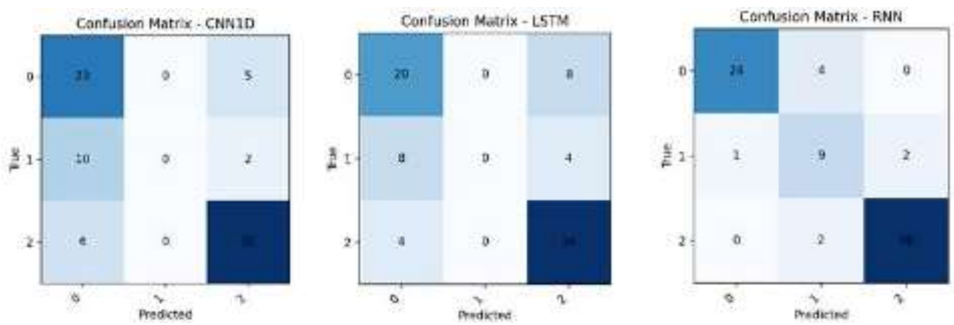


**Fig 2. Confusion matrix results for ML and DL models**

The baseline parameters (TP, FP, TN, and FN) form the basis for calculating the remaining key metrics by which algorithms are evaluated. A confusion matrix based on the actual and expected values was used to determine the baseline parameter values. The confusion matrix provided the final values for calculating the TPR and FPR values for each metric.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 02 February 2026
DOI: 10.47191/ijcsrr/V9-i2-28, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

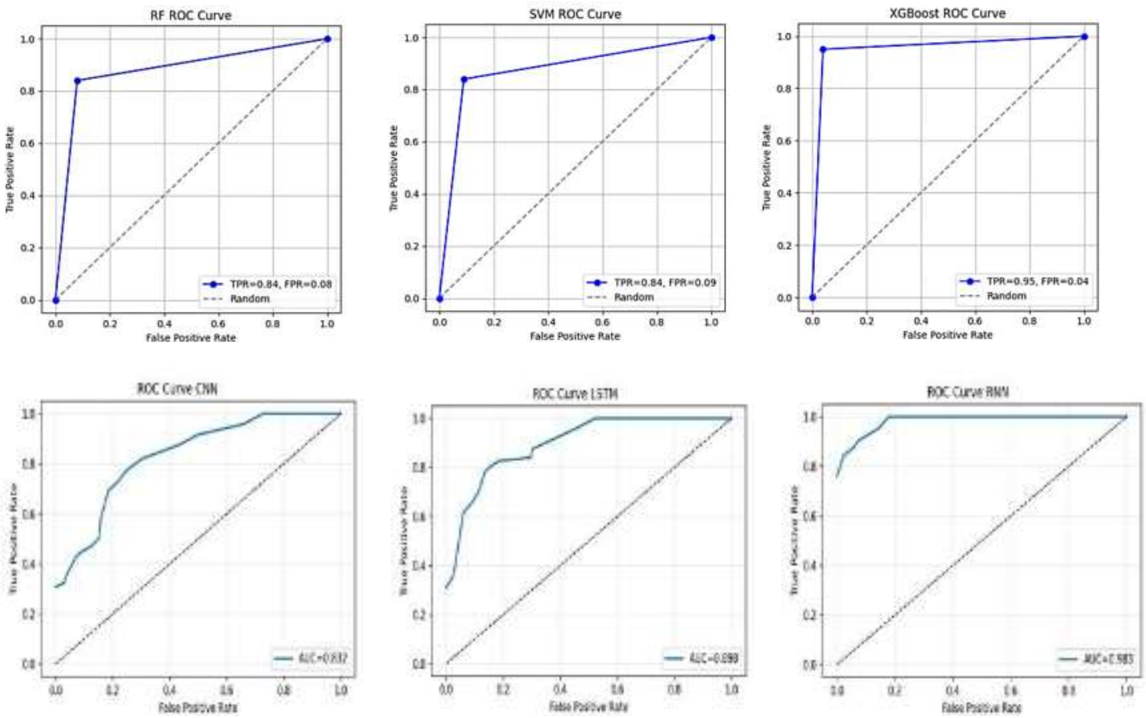


**Fig 3. ROC–AUC curve for ML and DL classifier**

Figure 3, shows the ROC–AUC curves for each classifier based on the baseline parameters derived from the confusion matrix. This figure displays the separate curves for all algorithms used in this study. This metric is considered better than overall accuracy for evaluating classifiers, especially with imbalanced data. The curves indicate that the top-performing algorithms, as XGBoost, are closer to the upper-left corner (with an AUC near 1). This indicates that these models may maintain a high True Positive Rate (TPR) while maintaining a low False Positive Rate (FPR), as evidenced by their high ability to distinguish between classes. Although the other curves perform well, they have smaller Area Under the Curve (AUC) values, showing less discrimination compared to the leading models. The results display noticeable variation in the metric values, as small changes in TPR and FPR can significantly impact the outcomes. Table 3, shows the values of the ROC–AUC curve with TPR and FPR for each classifier.

**Table 3. ROC–AUC curve with TPR and FPR values**

| Model | TPR | FPR | Precision | Recall | F1_score |
|---|---|---|---|---|---|
| **RF** | 0.84 | 0.08 | 0.87 | 0.85 | 0.86 |
| **SVM** | 0.86 | 0.07 | 0.88 | 0.86 | 0.87 |
| **XGBoost** | 0.92 | 0.04 | 0.92 | 0.9 | 0.91 |
| CNN | 0.554 | 0.165 | 0.56 | 0.65 | 0.57 |
| RNN | 0.851 | 0.054 | 0.89 | 0.89 | 0.89 |
| LSTM | 0.536 | 0.180 | 0.74 | 0.79 | 0.76 |

Figure 11 depicts the evaluation results of these models and highlights the most prominent differences between them based on various metrics. Three primary metrics, precision, recall, and F1-score, were computed to assess and contrast the models used. The findings demonstrated that the XGBoost algorithm outperforms the remaining algorithms.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 02 February 2026
DOI: 10.47191/ijcsrr/V9-i2-28, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

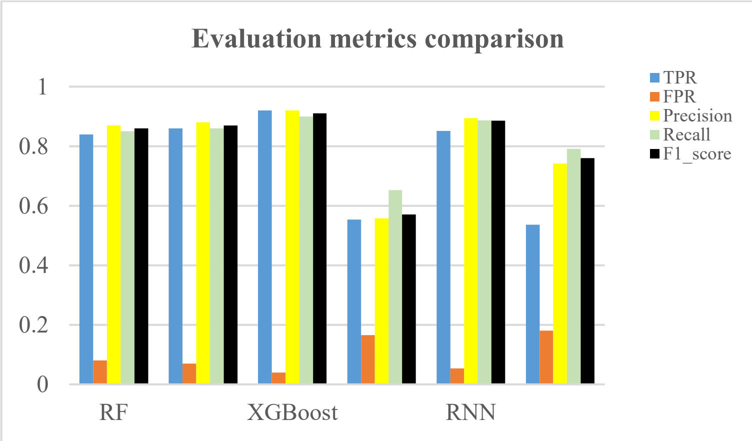


**Figure 4. Evaluation metrics comparison**

A thorough evaluation of the cost-based algorithms' performance is provided in Table 4, which clearly illustrates the conflict between computing efficiency and accuracy. The XGBoost model attained the best accuracy of 91%, demonstrating its outstanding ability to identify intricate patterns. However, this performance came with the longest training time of 20 seconds, indicating a significant computational load. In contrast, SVM and RF algorithms performed poorly, with training times of 8 and 10 seconds and accuracies of 88% and 85%, respectively. While deep learning algorithms in general demonstrated poor performance, notable differences were observed between models. The CNN model achieved the lowest accuracy of 65% and was trained in 5 seconds. In contrast, the RNN model reached 89% accuracy in 5 seconds, while the LSTM model achieved 79% accuracy with a 7-second training time. Overall, deep learning models performed worse with short training periods compared to machine learning models, which achieved higher performance with longer training times. These results may stem from the small dataset size and over-processing, which disproportionately impacted deep learning models.

In the end, the XGBoost algorithm had the highest accuracy in classifying its results. The XGBoost algorithm showed the best accuracy-to-highest computational cost ratio, providing robust diagnostic reliability without requiring an excessive amount of training time. This balance makes XGBoost particularly suitable for practical applications requiring both high precision and operational efficiency. Figure 5 depicts the accuracy values relative to the training time recorded for each classifier.

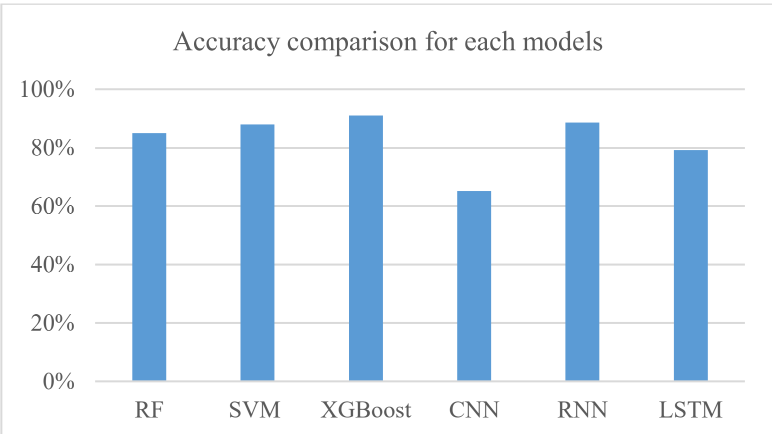


**Fig 5. Accuracy comparison for each model**

Table 4 provides a detailed evaluation of the advantages and disadvantages of all the classifiers used. It highlights the performance of each classifier employed in this study.

**Table 4. Pros and Cons of classifiers based on the accuracy**

| Model | Accuracy (%) | Pros | Cons |
|---|---|---|---|
| **RF** | 85% | Handles non-linear data, reduces overfitting, robust to noise | Slower training, less interpretable |
| **SVM** | 88% | Effective in high-dimensional spaces, good generalization | Requires careful tuning, slow with large datasets |
| **XGBoost** | 91% | High accuracy, handles missing data, regularization built-in | Complex to tune, longer training time |
| **CNN** | 65% | Excellent for spatial data, automatic feature extraction | Not ideal for tabular data, computationally intensive |
| **RNN** | 89% | Previous memory, dealing with long sequences and weight sharing over time periods | Slow, suffering from fading gradient and explosive gradient |
| **LSTM** | 79% | Flexible and strong, a solution for fading gradients | High cost, over-equipping, interpretation issue |

## 5. CONCLUSION

For decades, the prevalence of diabetes has posed a significant challenge to healthcare providers worldwide. The most common risk factors for the disease are social, behavioral, and medical in nature. Numerous studies have explored different types of diabetes, affecting people of all ages and both genders. This study utilized a Kaggle dataset that included three types of diabetes, categorized based on several factors. The main challenge was classifying these types using six machine learning and deep learning algorithms, which are among the most promising methods for disease identification. According to the results, the XGBoost model achieved the highest accuracy of 91% with a training time of 20 seconds. This outcome represents a positive step toward improved diabetes classification and detection, offering valuable opportunities for both healthcare professionals and patients. In future work, these findings could benefit classification, detection, and prediction in other fields.

### 5.1 Limitations of the study

This study's limitations included the following classification of diabetes mellitus using artificial intelligence models based on types of diabetes:

- Data Quality: Model performance is impacted by data imbalance, feature quality, and any other factors.
- Tuning parameters: One of the most important tools for achieving optimal performance is tuning parameters for each model (the settings that control the learning process).
- Interpretability: Understanding how the model functions in decision-making, particularly in healthcare applications, can be challenging when six machine learning and deep learning models are used.
- Computational Cost: The process of training and validating six models is computationally expensive, especially for deep learning models.
- Performance Metrics: Relying solely on accuracy factors is an insufficient indicator of clinical data, which is often imbalanced

**Acknowledgment**

The authors express their gratitude to the University of Diyala's Scientific Research Committee for supporting this significant effort. They also acknowledge the colleagues who helped to improve the quality of the research by offering guidance and recommendations.

**Competing Interests**

No conflict of interest.

**Funding Information**

No funding.

**REFERENCES**

1. A. Z. Woldaregay *et al.*, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," *Artif. Intell. Med.*, vol. 98, no. August 2018, pp. 109–134, 2019, doi: 10.1016/j.artmed.2019.07.007.
2. A. M. Carracher, P. H. Marathe, and K. L. Close, *International Diabetes Federation 2017*, vol. 10, no. 5. 2018. doi: 10.1111/1753-0407.12644.
3. M. B. L. and B. R. Qahtan M. Yas, A. A. Zadain*, B. B. Zaidan, "Towards on Develop a Framework for the Evaluation and Benchmarking of Skin Detectors Based on Arti¯cial Intelligent Models using Multi-Criteria Decision-Making Techniques Qahtan," *Int. J. Pattern Recognit. Arti¯cial Intell.*, vol. 31, no. 2, p. 1759002, 2016, doi: 10.1142/S0218001417590029.
4. Q. M. Yas, B. N. Adday, and A. S. Abed, "Evaluation Multi Diabetes Mellitus Symptoms by Integrated Fuzzy-based MCDM Approach," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 13, pp. 4069–4082, 2021.
5. R. Williams *et al.*, "Global and regional estimates and projections of diabetes-related health expenditure: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Res. Clin. Pract.*, vol. 162, 2020, doi: 10.1016/j.diabres.2020.108072.
6. A. A. Zaidan, B. B. Zaidan, M. A. Alsalem, O. S. Albahri, A. S. Albahri, and M. Y. Qahtan, *Multi-agent learning neural network and Bayesian model for real-time IoT skin detectors: a new evaluation and benchmarking methodology*, vol. 32, no. 12. Springer London, 2020. doi: 10.1007/s00521-019-04325-3.
7. D. Care and S. S. Suppl, "2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2025," *Diabetes Care*, vol. 48, no. January, pp. S27–S49, 2025, doi: 10.2337/dc25-S002.
8. J. Villar *et al.*, "Association between fetal abdominal growth trajectories, maternal metabolite signatures early in pregnancy, and childhood growth and adiposity: prospective observational multinational INTERBIO-21st fetal study," *Lancet Diabetes Endocrinol.*, vol. 10, no. 10, pp. 710–719, 2022, doi: 10.1016/S2213-8587(22)00215-7.
9. V. Bellou, L. Belbasis, I. Tzoulaki, and E. Evangelou, "Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses," *PLoS One*, vol. 13, no. 3, pp. 1–27, 2018, doi: 10.1371/journal.pone.0194127.
10. H. Wu *et al.*, "Worldwide estimates of incidence of type 2 diabetes in children and adolescents in 2021," *Diabetes Res. Clin. Pract.*, vol. 185, pp. 0–28, 2022, doi: 10.1016/j.diabres.2022.109785.
11. I. Contreras and J. Vehi, "Artificial intelligence for diabetes management and decision support: Literature review," *J. Med. Internet Res.*, vol. 20, no. 5, pp. 1–21, 2018, doi: 10.2196/10775.
12. H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024, doi: 10.58496/bjml/2024/007.
13. Ü. Ağbulut, A. E. Gürel, and Y. Biçen, "Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison," *Renew. Sustain. Energy Rev.*, vol. 135, p. 110114, 2021, doi: 10.1016/j.rser.2020.110114.
14. X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Mach. Learn. with Appl.*, vol. 6, no. September, p. 100154, 2021, doi: 10.1016/j.mlwa.2021.100154.
15. Y. Ayat *et al.*, "NOVEL DIABETES CLASSIFICATION APPROACH BASED ON CNN-LSTM : ENHANCED PERFORMANCE AND ACCURACY," vol. 25, no. 1, 2024, doi: 10.29354/diag/183633.
16. B. Al-Bander, Q. M. Yas, H. Mahdi, and R. K. H. S. Al-Hamd, "Benchmarking of deep learning algorithms for skin cancer detection based on a hybrid framework of entropy and VIKOR techniques," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 29, no. 8, pp. 2634–2648, 2021, doi: 10.3906/elk-2103-65.
17. L. Dénes-Fazakas *et al.*, "Physical Activity Detection for Diabetes Mellitus Patients Using Recurrent Neural Networks," *Sensors*, vol. 24, no. 8, pp. 1–18, 2024, doi: 10.3390/s24082412.
18. A. O. Eboka *et al.*, "Journal of Future Artificial Intelligence Resolving Data Imbalance Using a Bi-Directional Long-Short Term Memory for Enhanced Diabetes Mellitus Detection," 2025.
19. Q. Yas and Y. Kadthem, "Rainfall rate prediction using recurrent neural network with long short-term memory algorithm : Iraq case study Rainfall rate prediction using recurrent neural network with long short-term memory algorithm : Iraq case study Qahtan M . Yas * Younis Kadthem," *Int. J. Comput. Appl. Technol.*, vol. 74, no. 1–2, pp. 125–135, 2024, doi: 10.1504/IJCAT.2024.141352.

20. S. Ellahham, N. Ellahham, and M. C. E. Simsekler, "Application of Artificial Intelligence in the Health Care Safety Context: Opportunities and Challenges," *Am. J. Med. Qual.*, vol. 35, no. 4, pp. 341–348, 2020, doi: 10.1177/1062860619878515.
21. K. Thaiyalnayaki, "Classification of diabetes using deep learning and svm techniques," *Int. J. Curr. Res. Rev.*, vol. 13, no. 1, pp. 146–149, 2021, doi: 10.31782/IJCRR.2021.13127.
22. M. M. Islam, M. J. Rahman, D. Chandra Roy, and M. Maniruzzaman, "Automated detection and classification of diabetes disease based on Bangladesh demographic and health survey data, 2011 using machine learning approach," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 3, pp. 217–219, 2020, doi: 10.1016/j.dsx.2020.03.004.
23. S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *Int. J. Cogn. Comput. Eng.*, vol. 2, no. November 2020, pp. 40–46, 2021, doi: 10.1016/j.ijcce.2021.01.001.
24. M. Gollapalli *et al.*, "A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM," *Comput. Biol. Med.*, vol. 147, no. April, p. 105757, 2022, doi: 10.1016/j.compbiomed.2022.105757.
25. M. Y. Shams, Z. Tarek, and A. M. Elshewey, "A novel RFE-GRU model for diabetes classification using PIMA Indian dataset," *Sci. Rep.*, vol. 15, no. 1, pp. 1–22, 2025, doi: 10.1038/s41598-024-82420-9.
26. "Multiclass Diabetes Dataset." [Online]. Available: https://www.kaggle.com/datasets/yasserhessein/multiclass-diabetes-dataset
27. G. Dharmarathne, T. N. Jayasinghe, M. Bogahawaththa, D. P. P. Meddage, and U. Rathnayake, "A novel machine learning approach for diagnosing diabetes with a self-explainable interface," *Healthc. Anal.*, vol. 5, no. December 2023, p. 100301, 2024, doi: 10.1016/j.health.2024.100301.
28. A. Alazzawi, Q. M. Yas, and B. Rahmatullah, "A Comprehensive Review of Software Development Life Cycle methodologies : Pros , Cons , and Future Directions," *Iraqi J. Comput. Sci. Math.*, vol. 4, no. 4, pp. 173–190, 2023.
29. Q. M. Yas, A. A. Zaidan, and B. B. Zaidan, "Comprehensive insights into evaluation and benchmarking of real-time skin detectors: Review, open issues & challenges, and recommended solutions," *Meas. J. Int. Meas. Confed.*, vol. 114, pp. 243–260, 2018, doi: 10.1016/j.measurement.2017.09.027.
30. S. Muawanah, U. Muzayanah, M. G. R. Pandin, M. D. S. Alam, and J. P. N. Trisnaningtyas, "Stress and Coping Strategies of Madrasah's Teachers on Applying Distance Learning During COVID-19 Pandemic in Indonesia," *Qubahan Acad. J.*, vol. 3, no. 4, pp. 206–218, 2023, doi: 10.48161/Issn.2709-8206.
31. M. M. Mijwil and M. Aljanabi, "A Comparative Analysis of Machine Learning Algorithms for Classification of Diabetes Utilizing Confusion Matrix Analysis," *Baghdad Sci. J.*, vol. 21, no. 5, pp. 1712–1728, 2024, doi: 10.21123/BSJ.2023.9010.
32. D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, no. December 2022, p. 106131, 2023, doi: 10.1016/j.cor.2022.106131.
33. V. O. Nyangaresi, N. K. T. El-Omari, and J. N. Nyakina, "Efficient Feature Selection and ML Algorithm for Accurate Diagnostics," *J. Comput. Sci. Res.*, vol. 4, no. 1, pp. 10–19, 2022, doi: 10.30564/jcsr.v4i1.3852.
34. I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," *Technologies*, vol. 9, no. 4, 2021, doi: 10.3390/technologies9040081.
35. E. Helmud, E. Helmud, F. Fitriyani, and P. Romadiana, "Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 1, pp. 92–97, 2024, doi: 10.32736/sisfokom.v13i1.1985.
36. A. Theissler, M. Thomas, M. Burch, and F. Gerschner, "ConfusionVis: Comparative evaluation and selection of multi-class classifiers based on confusion matrices," *Knowledge-Based Syst.*, vol. 247, p. 108651, 2022, doi: 10.1016/j.knosys.2022.108651.


***Cite this Article: Al-Mhadawi, M.M., Hammoodi, A., Hameed, Y.K., Yas, Q.M. (2026). Multiclass Diabetes Classification using Multimodal Artificial Intelligence. International Journal of Current Science Research and Review,9(2), pp. 824-834. DOI: https://doi.org/10.47191/ijcsrr/V9-i2-28***