Journal of Food and Agricultural Technology Research
https://globalpresspk.com/
ISSN: 3007-0953 (Online)
ISSN: 3007-0945 (Print)

Research Paper

Diabetes Detection and Forecasting using Machine Learning

Aiman Musaed[1], Atta Rahman[1*], Muhammad Athar Javed[2], Mohammed Gollapalli[3], Saeed Matar Alshahrani[4], Aghiad Bakry[1], Mustafa Jamal Gul[5], Mahmoud O. Fares[1], Jamil Radman[1]

[1] Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O Box 1982, Dammam 31441, Saudi Arabia.
[2] Deanship of Academic Development, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.
[3] Faculty of Information Technology, Amtech Institute, Provider No. 14406, 88 Leichhardt St, Spring Hill, QLD 4000, Australia.
[4] College of Computing and Informatics, Saudi Electronic University, P.O. Box 93499, Riyadh 11673, Saudi Arabia.
[5] Department of Business Administration, University of York, Heslington, York YO10 5DD, United Kingdom.

***Corresponding author:** Atta Rahman (aaurrahman@iau.edu.sa)

Abstract: Diabetes Mellitus is a major health concern globally, requiring early diagnosis for its proper treatment. This research seeks to analyze the effectiveness of various machine learning algorithms for detecting diabetes using datasets from the Behavioral Risk Factor Surveillance System (BRFSS). In this context, two datasets have been considered for analysis: the binarized unbalanced dataset (BRFSS2015) and its balanced counterpart. The machine learning algorithms considered in this study are Logistic Regression, Random Forest, Gradient Boosting, K-Nearest Neighbors, Decision Trees, XGBoost, Support Vector Machines (SVM), and Deep Neural Networks. Therefore, the Gradient Boosting classifier has achieved an accuracy of 96.75% on unbalanced data and 85.36% on balanced data, the best among the machine learning algorithms. However, the Deep Neural Network algorithm has achieved 96.86% accuracy on the unbalanced dataset and 85.16% on the balanced dataset.

Keyword: Diabetes detection; machine learning; BRFSS 2015; Gradient Boosting; Deep Neural Networks

Article History
Received: 15 February 2026; Revised: 22 March 2026; Accepted: 23 April 2026 Online: 28 June 2026; Published: 30 June 2026; DOI https://doi.org/10.69501/3gnbk266

1. Introduction

Diabetes is serious health issue and leads further health disorders (Mallhi et al., 2025) and food materials are helpful to control it (Nasir 2024) and an incurable condition that causes considerable harm to health all over the world. Early diagnosis of the disease and accurate prediction of its presence are important for preventing complications such as heart disease and neuropathy. According to Alqarni & Rahman (2023), the issue has been magnified in the post Covid-19 pandemic era due to the significant change in lifestyle such as work from home. Moreover, it is observed that in Saudi Arabia, the diabetes is more prevalent in contrast to other countries due to its demographic reasons such as hot climate that prevent most of the population for outdoor activities (Al-Dossary et al., 2025).

Machine learning methods can detect patterns in data and predict the presence of certain conditions. Therefore, the goal of the research is to analyze the effectiveness of some machine learning models in detecting diabetes using data from two BRFSS 2015 datasets binarized, imbalanced diabetes binary health indicators BRFSS2015 and balanced diabetes binary 5050split health indicators BRFSS2015.

In the original dataset, there are health indicators such as high blood pressure, cholesterol level, BMI, smoking, physical activity, and others. The binary target variable in the dataset indicates whether a person has diabetes. In the second dataset, class imbalance is resolved because the number of diabetic and non-diabetic cases is equal. Thus, this research aims to compare the most widely used algorithms in the literature for similar problems, including Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors (KNN), Decision Trees (DT), XGBoost, SVM, and Deep Neural Networks (DNN).

The rest of the paper is organized as follows: section 2 is dedicated to reviewing the related literature. Section 3 emphasizes materials and methods. The results are presented in section 4 while section 5 concludes the paper.

2. Literature Review

The study conducted by Butt et al. (2021) used machine learning models to predict and classify diabetes using the PIMA Indian Diabetes dataset. In classifying diabetes, they used three machine learning models: Random Forest (RF), Multilayer Perceptron (MLP), and Logistic Regression (LR), with the MLP achieving the highest accuracy of 86.08%. For predicting diabetes, they used three machine learning models: Long Short-Term Memory (LSTM), Moving Averages (MA), and Linear Regression (LR), with the LSTM model achieving the highest prediction accuracy of 87.26%.

Modeling, Simulation and Optimization for Prediction of Type II Diabetes Based on Deep Extreme Learning Machine (DELM) was proposed by Rehman et al. (2022). They have said in their research, "In the human body, glucose is the major component that energizes cells. But insulin is the key component that enters cells to regulate blood glucose levels. Individuals who have type 1 diabetes cannot produce insulin. In contrast, individuals who have type II diabetes cannot react to insulin and produce a sufficient amount of insulin. To properly analyze the disease, a method should be used that has the lowest possible error rate. Using DELM, we achieve high accuracy with the minimum possible error rate. We find that the proposed approach significantly improves performance compared to past studies. In our study, the proposed approach achieved the highest accuracy of 92.8% with 70% training (9500 samples) and 30% for test and validation (4500 samples).

The work by Tigga and Garg (2020) is centered on applying machine learning algorithms to the early detection and treatment of Type 2 diabetes in India. Data were collected from 952 people via online and offline questionnaires containing 18 questions about health, lifestyle, and participants' family histories. Different machine learning algorithms were applied, including the Random Forest Classifier. According to the paper, all of these algorithms are very accurate for healthcare use. From the literature review, machine learning, data mining, neural networks, and genetic algorithms are widely used for diabetes prediction. Many studies are cited by authors who used various datasets and techniques to predict diabetes, achieving varying levels of accuracy. It turned out that the Random Forest Classifier is the most accurate algorithm for both datasets considered in the work.

The primary objective of Alic et al. (2020) was the application of machine learning techniques for predicting the future risk of developing type-2 diabetes. In the study, they analyze previously collected data in the San Antonio Heart Study using Support Vector Machines (SVM), together with ten factors which have proven to be good predictors of future development of diabetes. This particular dataset is imbalanced, having more cases of healthy patients than diabetic ones. In order to solve this problem, the authors make use of 10-fold cross-validation as training and a hold-out approach as validation. In the experiment, the validation accuracy is 84.1% while recall rate is 81.1%, over 100 iterations. The best approach used in this paper is the Support Vector Machines (SVM) approach. The accuracy of the SVM approach is 84.1%. previously, Ali et al., (2025) has used machine learning (ML) for students performance. ML is great tool for detection of food nutritional components with high accuracies such as water (Younas et al., 2025; Younas et al., 2026; Li et al., 2025), phenolic compounds (Younas et al., 2024).

Xie et al. (2019) intended to build machine learning models for prediction of Type 2 diabetes. The research relied on the 2014 Behavioral Risk Factor Surveillance System data that involved 138,146 respondents. Several machine learning models were constructed; such as Support Vector Machine (SVM), Decision Tree, Logistic Regression, Random Forest, Neural Network and Gaussian Naive Bayes. The performance of models was measured through their accuracy, sensitivity, specificity and Area Under the Curve (AUC). The best accuracy (82.4%), specificity (90.2%) and AUC (0.7949) were demonstrated by the Neural Network model. The best sensitivity (51.6%) belonged to the Decision Tree model. In addition, the study found out two novel risk factors: sleeping time and checkup frequency.

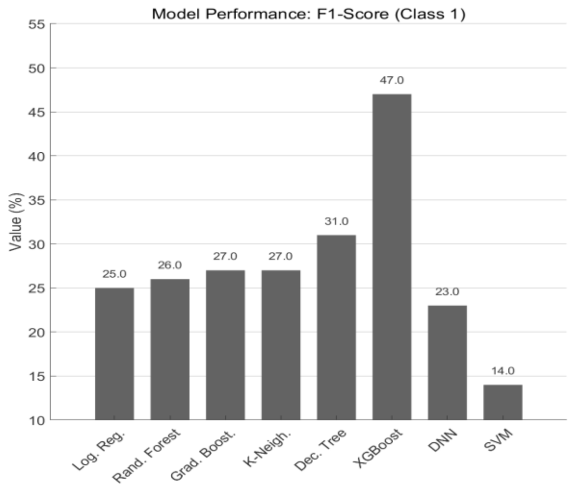

Figure 1. F1-score analysis (class 1)

By Battineni et al. (2019) is about the prediction of Type 2 diabetes among the Pima Indian women by applying machine learning (ML) techniques. The experiment used the Pima Indian diabetes dataset (PIDD) which consists of 768 Pima Indian women. Four types of ML classifier algorithms were used in the process, and those include Naïve Bayes (NB), J48, Logistic Regression (LR), and Random Forest (RF). The classification models were evaluated using various cross validation methods where K=5, 10, 15, and 20. Accuracy, precision, F-score, Recall, and AUC were used as metrics for evaluating the models. Among all the models, the Logistic Regression (LR) showed maximum accuracy of 0.77. The objective of the research was to find out the optimized ML model for predicting diabetes and it was found that LR, RF, and NB were the three most suitable models with AUC 0.83, 0.82, and 0.81, respectively.

Mounika & Sunil (2021) investigated the use of machine learning techniques for the diagnosis of Type-2 diabetes. For its experimentation, it uses the Pima Indian Diabetes Dataset and utilizes two machine learning techniques: Logistic Regression and Random Forest Classifier. The purpose of this research paper is to analyze the effectiveness of these techniques in diagnosing Type-2 diabetes. The conclusion drawn in this paper is that Random Forest gives the maximum accuracy for the PIMA dataset. In particular, Random Forest had an accuracy of 95%, while Logistic Regression had 97%.

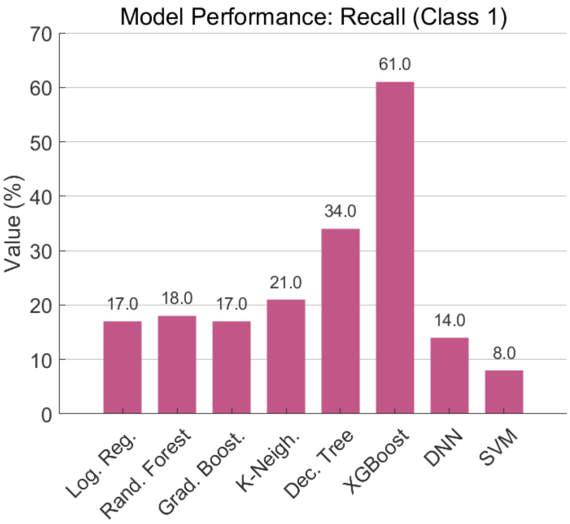

Figure 2. Recall analysis (class 1)

The problem of the accurate prediction of Type 2 Diabetes Mellitus (T2DM) was addressed by Wu et al. (2018). A disease with serious public health consequences. Despite the use of some prediction models and algorithms, they are unable to provide proper levels of accuracy and adaptation to various datasets. The traditional approach involves the use of a single algorithm such as K-means or logistic regression. In their study, the researchers propose a new two-level algorithmic model which utilizes the combination of improved K-means and logistic regression to cluster the data at first and classify it later using WEKA and Pima Indians Diabetes Dataset. In this way, they reach a rather high accuracy rate of 95.42%. It is great about their accomplishment because of the high level of accuracy achieved. Besides, it is a good thing about their record that their model can be adapted to various datasets. Nevertheless, the complex structure of their model could become an obstacle for its practical implementation. Furthermore, the main testing of their model on the Pima Indians Diabetes Dataset leads to the question about the applicability of the model to other people.

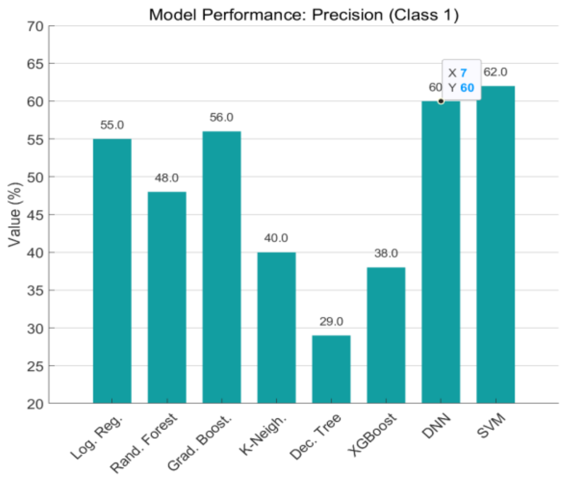

Figure 3. Precision analysis (class 1)

Sisodia and Sisodia (2018) explored the issue of predicting diabetes in the high-risk category of pregnant females using Naive Bayes, Support Vector Machine (SVM), and Decision Tree algorithms in WEKA and Pima Indians Diabetes Database (PIDD). Although previous research has predominantly been centered around individual algorithms, in this study, the researchers managed to make a comparison and provide a broader view of their efficiency. The highest accuracy rate was observed in the case of Naive Bayes algorithm (76.30%). This work has significant value since it compares different machine learning models and is centered around high-risk cases, which might be helpful for healthcare professionals. At the same time, the work has some weaknesses as well. Firstly, the relatively low accuracy levels suggest that there is some space for improvement. Secondly, the researchers' reliance on the Pima Indians Diabetes Database might affect the generalizability of the results to other populations.

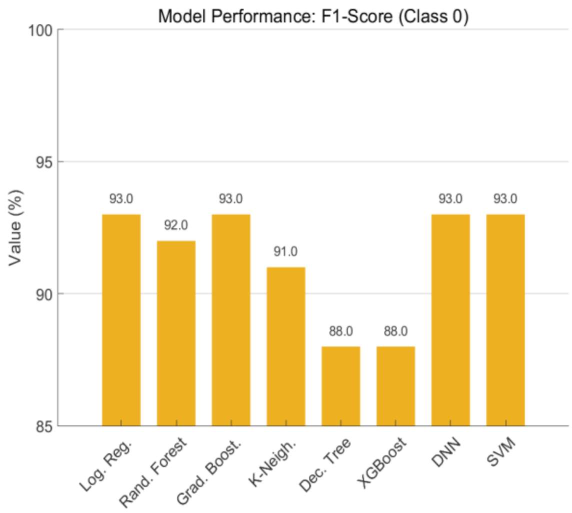

Figure 4. F1-score analysis (class 0)

Swapna et al. (2018) discussed the problem of early diabetes detection through a methodology which does not require any invasiveness by utilizing Heart Rate Variability (HRV) signals obtained from Electrocardiograms (ECG). Instead of employing single algorithm techniques and smaller datasets, this paper uses a methodology which makes use of multiple algorithms, where deep learning models including Convolutional Neural Network (CNN), LSTM and SVM are used to classify and obtain results. Such a methodology was able to achieve an accuracy of 95.7% using CNN 5-LSTM with SVM methodology. Strength of this paper include high accuracy and non-invasive technique which can prove to be revolutionary for healthcare professionals. Weaknesses of this research include the complexity associated with the use of deep learning models and smaller datasets used in the study.

Type 2 diabetes prediction is a crucial issue with serious health consequences, as addressed by Joshi & Dhakal (2021). The authors utilize an approach that combines logistic regression with machine learning, in particular, the classification tree to examine the Pima Indians dataset. Five major predictors of type 2 diabetes, as found by the model, include glucose, pregnancy, BMI, diabetes pedigree function, and age. The 6th model of Logistic Regression reached a prediction accuracy of 78.26%. The benefits of this model are in reasonable prediction accuracy and the ability to supplement existing preventive measures. However, limited testing on one dataset may be a downside.

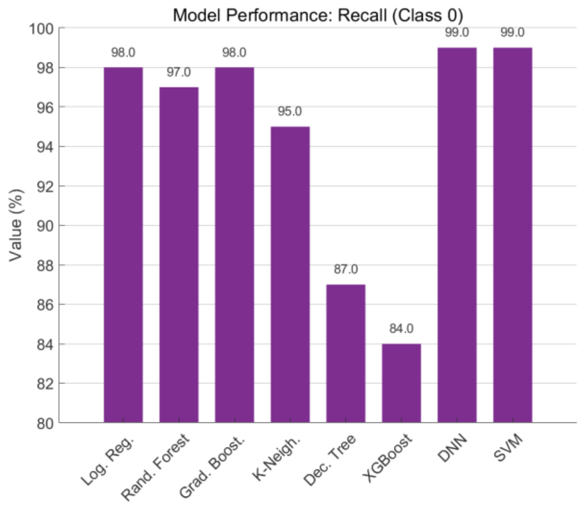

Figure 5. Recall analysis (class 0)

Predicting T2DM with the use of the Glmnet model has been developed by Kopitar et al., (2020). This model showed quite high results in terms of the area under the curve (AUC), achieving 0.859 in AUC. The major advantage of the Glmnet model is in better performance in datasets where predictors have high correlation and sparsity.

In Fazakis et al. (2021) work, weighted VotingLRRFs model was used for T2DM prediction. It showed an AUC of 0.884, meaning a very good accuracy of predictions. One of the greatest benefits of the Weighted VotingLRRFs model is optimization of multiple measures of classification quality at once, including AUC and Sensitivity.

Awad et al. (2020) used a mathematical model to predict the T2DM burden for the country up to 2050 considering the key risk factors. The prediction was made using age-structure models and survey data. The prevalence of T2DM for the country for the period from 2020 to 2050 triples, namely from 15.2% to 23.8%. According to the report, the fastest growing key risk factor is obesity. It is reported that in 2020, 56.7% of T2DM cases were caused by obesity according to the forecasts, and it accounts for 71.4% in 2050. It is also stated that the proportion of the national health care expenditures allocated on T2DM treatment increases from 21.2% to 28.8% in 2050. Robust sensitivity and uncertainty analyses proved these results.

The work of Nicolucci et al. (2022) applied machine learning for building predictive models for complications of T2DM based on EMRs from 23 diabetes centers of Italy for 15 years involving 147,664 patients. Predictions were made for six groups of complications that include diabetes eye complication, cardiovascular disease, cerebrovascular disease, peripheral vascular disease, nephropathy, and diabetic neuropathy with the use of supervised tree-based learning method (XGBoost). Models proved to have very high accuracy, which is higher than 70%, and AUC values greater than 0.80, nephropathy prediction model obtained even 0.97 AUC value. Besides, the models had high predictability in case of predicting early occurrences of the complications (two years), where sensitivity varied from 83.2% to 88.5%. External validation carried out in five diabetes centers revealed good performance of the models and robustness but also some predictive performance variability. This research proves the effectiveness of using machine learning algorithms for predicting complications of diabetes mellitus.

In their study, Wee et al. (2024) made use of various machine learning and deep learning techniques for the diagnosis of diabetes and explored the effects of the latest techniques based on both invasive and non-invasive data sets. In case of machine learning techniques, they used Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR), attaining accuracy up to 82% with respect to RF technique. As far as deep learning techniques are concerned, they made use of Convolutional Neural Networks (CNN), Deep Neural Networks (DNN) and Deep Belief Networks (DBN). They got 92.31% and 98.16% accuracy with respect to CNN and DNN models respectively.

Study conducted by Sharma et al. (2020) involves usage of the Pima Indians Diabetes database that comprises 768 samples. In this experiment, the capability of diabetes prediction of three methods was compared: Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN).

For each approach, the same number of training and testing samples from the data set with eight attributes regarding diabetes were considered. As can be seen from the results obtained, the Random Forest approach, with accuracy of 83.67%, demonstrated the best results in comparison with the Support Vector Machine and Deep Learning approaches, for which the accuracy was 65.38% and 76.81% correspondingly.

It can be seen from the analysis of scientific literature that deep learning and machine learning approaches are promising in the prediction of different diseases including type II diabetes such as; Musleh et al., (2024&2026); Ahmed et al. (2025a&b); Youldhash et al. (2024&2025); Gollapalli et al. (2024a&b); Alnuaimi et al. (2024); Rahman et al. (2024); Singh et al. (2024) and Alhashem et al. (2023).

3. Materials and Methods

This study utilized two datasets: diabetes binary health indicators BRFSS2015 and diabetes binary 5050 split health indicators BRFSS2015. The first dataset is a binarized version of the original imbalanced dataset containing health indicators for diabetes from the Behavioral Risk Factor Surveillance System (BRFSS) 2015, with a binary target variable indicating whether an individual has diabetes. The second dataset is a balanced version derived from the first, containing an equal number of diabetic and non-diabetic individuals. Both datasets share the same set of columns.

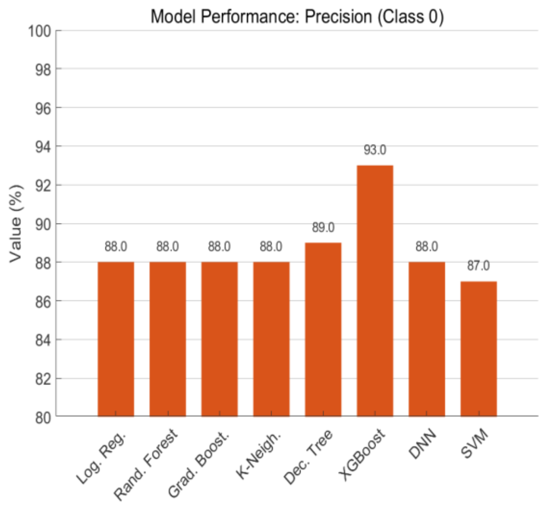

Figure 6. Precision analysis (class 0)

3.1. Data Description

The datasets utilized in the present research include diabetes binary health indicators BRFSS2015 (imbalanced) and diabetes binary 5050 split health indicators BRFSS2015 (balanced), each of which includes 22 features described in Table 1. The target feature is Diabetes_012, where 0 denotes absence of diabetes, 1 denotes pre-diabetes and 2 denotes diabetes; in case of binary classification, it was transformed to binary target feature (0-no diabetes, 1-diabetes/pre-diabetes). Imbalanced dataset consists of 253,680 data points, where about 13.9% of them are diabetes patients, while balanced dataset includes 43,784 data points, of which half are diabetic, and half are non-diabetic patients. Class distribution in the two datasets is presented in Fig. 1(a,b) and bar charts with comparison of number of samples in both datasets are presented in Fig. 2.

Summary statistics on the unbalanced data set are presented in Table 2. The mean value of Diabetes_binary is 0.139, which proves the unbalance of this data set. From the clinical predictors, HighBP (mean = 0.429), HighChol (0.424) and BMI (mean = 28.38) are present with relatively high frequency, whereas CholCheck (0.963) and AnyHealthcare (0.951) are close to 1. Standard deviations of PhysHlth (8.72) and MentHlth (7.41) are indicative of considerable variability in self-reported number of health days. Summary statistics on the balanced data set (see Table 3) show that the mean value of Diabetes_binary is 0.5 by definition. In contrast to the unbalanced data set, the balanced one demonstrates higher means for HighBP (0.563 vs. 0.429), BMI (29.86 vs. 28.38), GenHlth (2.84 vs 2.51), and PhysHlth (5.81 vs. 4.24), reflecting the enrichment of diabetic individuals who tend to have worse health indicators.

Correlation analysis of the relationship between each variable and the target variable (Diabetes_binary) is provided in Tables 4 and 5. In terms of the imbalanced data set (Table 4), the positive correlations of the greatest strength with diabetes have been found in GenHlth (0.294), HighBP (0.263), DiffWalk (0.218), BMI (0.217), and HighChol (0.200). Among negative correlations (protective factors): Income (−0.164), Education (−0.124), and PhysActivity (−0.118). As far as the balanced data set (Table 5) goes, the correlations have become much more pronounced: GenHlth (0.408), HighBP (0.382), BMI (0.293), Age (0.273). The negative correlations have become more pronounced as well (Income −0.224, Education −0.170 PhysActivity −0.159). This increase is expected because balancing removes the class imbalance bias, revealing the true associations between predictors and diabetes.

The types of data and ranges for all continuous and categorical variables are listed in Table 1. In brief, Age is continuous and measured in years; BMI is continuous; PhysHlth and MentHlth are the number of unhealthy days in the last 30 days (0-30). The list of categorical variables includes Education (1-6); Income (1-8); GenHlth (1-5, where 1 is excellent and 5 is poor). Binary variables such as Sex, Smoker, HighBP, and DiffWalk are either 0 or 1. Together, these variables provide a comprehensive profile of health behaviors, comorbidities, and sociodemographic factors associated with diabetes.

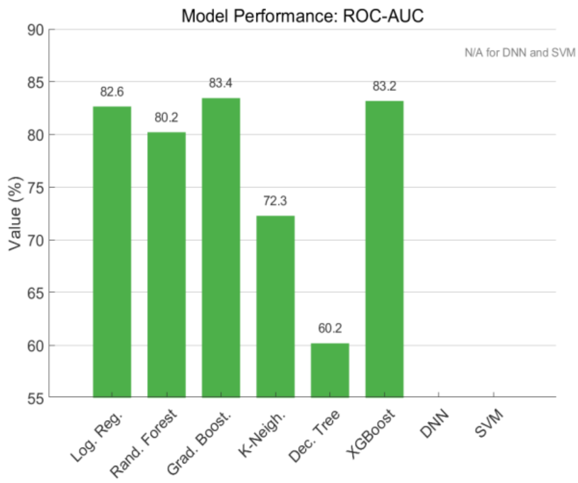

Figure 7. ROC-AUC analysis

3.2. Preprocessing Steps for All Models

Prior to modeling, preprocessing of both datasets was done in the following way. The splitting of the entire set was done into training and testing sets, in which 20% of the entire data was used for testing purposes to evaluate the performance of the models. Further, the features were standardized using StandardScaler, ensuring that there is no bias towards the features of large scales.

3.3. Models

Two datasets were used in this research; that is, diabetes binary health indicators BRFSS2015 and diabetes binary 5050split health indicators BRFSS2015. The former is an imbalanced dataset where the data is binarized with the target column denoting if someone has diabetes or not. The latter is the data that is made balanced, based on the former and consists of an equal number of diabetics and non-diabetics. Both datasets contain the same set of columns.

Before implementing the models, both the complete dataset and the balanced subset were pre-processed as below. First, both datasets were divided into training and testing datasets, where 20% of the data was kept for testing purposes and feature standardization was done using the StandardScaler to avoid bias towards large-scale values.

Different classifiers were trained. In order to train the logistic regression, the solver "liblinear" was chosen since it is appropriate for binary classification and L2 regularization was performed to avoid overfitting. Different variations of the random forest classifier were examined. The parameters used include the following: the number of trees (n_estimators) equals 10, 20, and 30; the maximum depth of the tree (max_depth) is 10, 20, and 30; the minimum number of samples required to perform the split into an internal node (min_samples_split) was set as 2, 5, 10, 15, and 20; and the minimum number of samples at each leaf node (min_samples_leaf) is 1, 2, 4, 6, and 8. In turn, for training the gradient boosting classifier, the number of boosting stages (n_estimators) equals 100 and 200, the learning rate is 0.01 and 0.1, and the maximum depth of the tree (max_depth) equals 3, 5, and 7.

An SVM with RBF kernel was utilized; the cost hyperparameter was chosen to be 1.0 in order to keep the compromise between misclassification error and decision surface complexity, and the hyperparameter for kernel coefficient, gamma, was set to the ‘scale’ parameter, which automatically adjusts depending on the number of features. The K nearest neighbors (KNN) method was applied with different numbers of neighbors to find the optimal parameter for classification. The decision tree classifier was applied as well, with max_depth set to 30 to avoid overfitting and yet capture enough patterns of the data, min_samples_split = 2, and min_samples_leaf = 1 to enable detailed splitting at the leaves level. Moreover, the XGBoost model was implemented; the synthetic minority oversampling technique (SMOTE) was used for balancing the training data, and the following parameters were set: n_estimators = 100 and 200, max_depth = 3, 5, and 7, learning_rate = 0.01 and 0.1, subsample = 0.8 and 1.0, and colsample_bytree = 0.8 and 1.0. At last, the deep learning model was built utilizing Keras library.

The architecture of the neural network included three fully connected dense layers having 12, 8, and 1 neuron each. The network used ReLU as an activation function for the hidden layers and a sigmoid function for the output layer. For optimizing the model, Adam optimizer was used and it was trained for 50 epochs.

3.4 Models Evaluation

The effectiveness and robustness of the constructed predictive models will be evaluated using a comprehensive framework for evaluating the models. The framework makes use of several performance metrics that help in determining certain aspects of model performance. The main performance metrics to be used are given as Rahman (2025); Hantom & Rahman (2024); Rahman et al. (2025); Arooj et al. (2024), Rahman et al. (2026).

The effectiveness and robustness of the constructed predictive models were evaluated using a comprehensive framework using several performance metrics. The accuracy is the measurement of correctness of prediction in all classes as the ratio of correct predictions to total observations:

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

Precision measures the accuracy of positive predictions, which is particularly important when false positives carry high cost:

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

Recall (sensitivity), critical in medical diagnostics, captures the model's ability to identify all relevant instances:

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

The F1-score, the harmonic mean of precision and recall, provides a balanced measure for uneven class distributions:

$$\mathbf{F1\ score} = \frac{\mathbf{2 * precision \times recall}}{\mathbf{precision + recall}} \quad (2)$$

The receiver operating characteristic area under the curve (ROC-AUC) measure assesses the performance of the model in terms of discrimination ability at all classification decision levels. This area is equal to the probability that the model ranks positive observation higher than the negative one, where the ROC graph shows the true positive rate (recall) versus the false positive rate (1-specificity).

Finally, the confusion matrix was used for visualizing the results as an essential graphical tool, which allows observing the predicted classes versus the actual classes to estimate recall, precision, and different kinds of errors.

4. Results

This section will present the results of applying the mode to different criteria. Figure 1 shows the F1 score analysis on criterion 1 with different suggested approaches like LR, RF, GB, KNN, DT, XGBoost, DNN, and SVM. In this way, XGBoost was more efficient with an F1 score of 47% compared to the others approaches which had F1 scores between 14% to 31%.

Figure 2 provides the recall analysis for class 1 using various proposed algorithms, including LR, RF, GB, KNN, DT, XGBoost, DNN and SVM, respectively. In this regard, XGBoost outperformed the others with a 61% recall, followed by DT at 34%. SVM underperformed at 8% followed by DNN at 14%, while the rest were in the range of 17% to 21%.

Figure 3 provides precision analysis for class 1 using various proposed algorithms, including LR, RF, GB, KNN, DT, XGBoost, DNN and SVM, respectively. In this regard, SVM outperformed the others with a 62% precision value, followed by DNN at 60%. DT underperformed at 29% followed by XGBoost at 38% and KNN at 40%, while the rest were in the range of 48% to 56%.

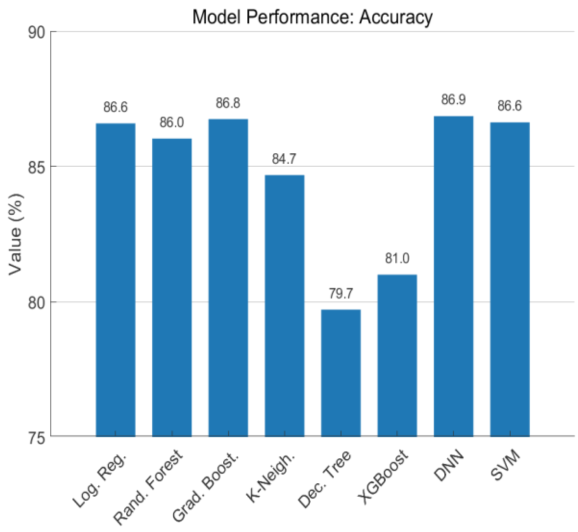

Figure 8. Accuracy analysis

Figure 4 provides the F1-score analysis for class 0 using various proposed algorithms, including LR, RF, GB, KNN, DT, XGBoost, DNN and SVM, respectively. In this regard, DNN, SVM, LR and GB outperformed the others with a 93% F1-score followed by RF and KNN at 92% and 91%, respectively. While DT and XGBoost were on relatively poorer sides at 88%. However, in contrast to class 1, all the algorithms performed exceptionally well.

Figure 5 provides the recall analysis for class 0 using various proposed algorithms, including LR, RF, GB, KNN, DT, XGBoost, DNN and SVM, respectively. In this regard, DNN and SVM outperformed the others with a 99% recall value followed by LR and GB at 98% each. RF and KNN obtained recall values as 97% and 95%, respectively. While DT and XGBoost were on relatively poorer sides at 87% and 84%, respectively. However, in contrast to class 1, all the algorithms performed exceptionally well.

Figure 6 provides precision analysis for class 0 using various proposed algorithms, including LR, RF, GB, KNN, DT, XGBoost, DNN and SVM, respectively. In this regard, XGBoost outperformed the others with a 93% recall, followed by DT at 89%. LR, RF, GB, KNN, and DNN were at 88% while SVM at 87%.

Figure 7 provides ROC-AUC analysis using various proposed algorithms, including LR, RF, GB, KNN, DT, XGBoost, DNN and SVM, respectively. GB outperformed at 83.4%, followed by XGBoost, LR and RF, with a slight difference: 83.2%, 82.6% and 80.2% ROC-AUC, respectively. DT and KNN were on the relatively poor side, 60.2% and 72.3%, while DNN and SVM were undefined.

Figure 8 provides accuracy analysis using various proposed algorithms, including LR, RF, GB, KNN, DT, XGBoost, DNN and SVM, respectively. DNN and GB outperformed the others with 86.9% and 86.8% accuracy, respectively. After that, LR and SVM outperformed at 86.6%, followed by RF and KNN at 84.7%, respectively. However, DT and XGBoost were on relatively poorer sides at 79.7% and 81%, respectively.

5. Discussion

The logistic regression model, when applied to the original imbalanced dataset, returned an accuracy of 86.59% and ROC AUC of 82.64%. It showed excellent precision and recall for the non-diabetic class (class 0) while having poor performance regarding class 1 (diabetic). The random forest classifier model provided an accuracy of 86.03% and ROC AUC of 80.20%, experiencing problems with the same class 1 (diabetic).

The gradient boosting technique generated the best result of all methods when working with this dataset - accuracy is 86.75%, ROC AUC is 83.44%, while providing great performance for class 0 and better performance for class 1. The k neighbors classifier returned an accuracy of 84.68% and ROC AUC of 72.27%. This method showed good performance for class 0 but had much lower indicators when processing class 1 (diabetic). The decision tree classifier was the method with the worst indicators: accuracy is 79.71%, ROC AUC is 60.17%.

XGBoost had accuracy of 81.00% and ROC AUC of 83.17%, with a highly accurate precision for class 0 and good recall for class 1, showing that this model is properly tuned for handling imbalanced data. The deep neural network was the second most accurate model, with accuracy of 86.86% (ROC AUC not specified), good results for class 0 and poor results for class 1. SVM got an accuracy of 86.63%, good for class 0, and poor for class 1, which shows difficulties in handling non-linear separations.

The changes in performance on the balanced binary dataset are quite considerable. The accuracy of the logistic regression went down to 74.84% (ROC AUC: 82.49%), while the metrics of precision and recall became balanced. The accuracy of the random forest algorithm went down to 73.77% (ROC AUC: 80.82%) but was also balanced. The gradient boosting algorithm showed high accuracy, which is equal to 75.36% (ROC AUC: 83.16%). In addition, the performance of the algorithm was balanced for both classes.

The kNNclassifier had accuracy equal to 70.99% (ROC AUC: 76.94%). Thus, this algorithm had the metrics balanced but low because of its difficulties with the balanced datasets due to the distance principle. The decision tree algorithm showed the lowest accuracy equal to 64.81% (ROC AUC: 64.83%). In addition, the metrics were balanced and lower, which means that the problem with overfitting is persistent. XGBoost had the accuracy equal to 73.00% (ROC AUC: 83.26%). The metrics of the precision of class 0 and recall of class 1 were high. Finally, accuracy of SVM was lowered down to 74.81% with equal precision and recall, once again indicating limitations in dealing with non-linear class separations without any further optimization.

In future, hybrid models, ensemble learning approaches and transfer learning can be investigated for further refinement of the results such as conducted in the studies including Rahman et al. (2022); Nasir et al. (2022a&b); Dash et al. (2019); Khan et al. (2020); Ahmed et al. (2023); Ahmed et al. (2026a&b); Qayyum et al. (2026); Khan et al. (2025) and Ahmed et al. (2025).

6. Conclusion

Given both datasets, the Gradient Boosting Classifier stands out to be the best algorithm for diagnosing diabetes. It performed exceptionally well in the binary dataset with an accuracy of 86.75%, while on the other hand, it performed at 75.36% on the balanced dataset. The consistency of its precision and recall on the two datasets across both classes makes it very robust and effective.

The capability of this algorithm to perform well in imbalanced and balanced datasets makes it a good fit for any application since the class distribution can be different. The better performance of its ROC-AUC values makes it a better performing algorithm in the discrimination of the diabetic and the non-diabetic people. This is important for the diagnosis of diabetes. The good performance of the deep neural network in the two cases with an accuracy of 86.86% on the binary and 75.16% on the balanced dataset also makes it robust and effective.

Conflicts of Interest: The authors declare no conflicts of interest.

Reference

Ahmed, M. I. B., Olatunji, S. O., Rahman, A., Alhiyafi, J., Kudos, S. A., Alqahtani, M., Alshammari, G., & Alharbi, A. (2026). *Early diagnosis of hypothyroidism in the Eastern Province of Saudi Arabia using computational intelligence techniques*. Journal of Visualized Experiments, 70065.

Ahmed, M. I. B., Olatunji, S. O., Rahman, A., Alqahtani, M., Alshammari, G., Alharbi, A., & Mohammad, N. (2026). *A computational intelligence-based early diagnosis of asthma disease: A Saudi Arabian case study*. Journal of Visualized Experiments, 232, e70040.

Ahmed, M. I. B., Zaghdoud, R., Rahman, A., Ali, F., Alhashim, H., Almubarak, M. Y., Alqahtani, M., Alhabboub, Y., & Alzahrani, K. (2025). *A hybrid machine learning model in diagnosing brain strokes*. Journal of Advances in Information Technology, 16(11), 1664–1674.

Ahmed, M. S., Rahman, A., AlGhamdi, F., AlDakheel, S., Hakami, H., AlJumah, A., AlIbrahim, Z., Youldash, M., Alam Khan, M. A., & Basheer Ahmed, M. I. (2023). *Joint diagnosis of pneumonia, COVID-19, and tuberculosis from chest X-ray images: A deep learning approach*. Diagnostics, 13(15), 2562.

Al-Dossary et al. (2024). *Evaluating Mental Health and Well-Being through Social Media Analysis*. International Journal of Computer Science and Network Security, 24(11), 21-34.

Alhashem, A., Abdulbaset, A., Almudarra, F., Alshareef, H., Alqasoumi, M., Alqahtani, M., Alshammari, G., & Alharbi, A. (2023). *Diabetes detection and forecasting using machine learning approaches: Current state-of-the-art*. International Journal of Computer Science and Network Security, 23(10), 199–208.

Ali, F., Imran, Z., Ali, A., Ahmed, M., Iqbal, A., Nazeer, A., ... & Mariussi, M. (2025). A Novel Concept Study on Key Factors Affecting Student Performance Using Machine Learning Algorithms. Journal of Food and Agricultural Technology Research, 4(01), 38-51.

Alic, A., Imran, M., & Ali, S. H. (2020). *Machine learning approach to predict future risk of type-2 diabetes*. Health Informatics Journal, 26(3), 1925–1937.

Alnuaimi, M. M., Alqahtani, N. S., Gollapalli, M., Rahman, A., Alahmadi, A., Alshammari, G., Alqahtani, M., & Alharbi, A. (2024). *Transfer learning empowered skin diseases detection in children*. Computer Modeling in Engineering & Sciences, 141(3), 2609–2623.

Alqarni, A. & Rahman, A. (2023). Arabic Tweets-Based Sentiment Analysis to Investigate the Impact of COVID-19 in KSA: A Deep Learning Approach. Big Data and Cognitive Computing 7 (1), 16.

Arooj, S., Khan, M. F., Shahzad, T., Khan, M. A., Nasir, M. U., Zubair, M., Rahman, A., Alqahtani, M., Alshammari, G., & Alharbi, A. (2024). *Data fusion architecture empowered with deep learning for breast cancer classification*. Computers, Materials & Continua, 77(3), 2813–2830.

Aslam, S., Khan, T. J., Mehmood, H., Baig, M. I., & Saleem, S. (2020). *Application of deep learning techniques for diabetes analysis*. Journal of Computational Science, 47, 101247.

Awad, M., Al-Mawali, A., Al-Lawati, J., & Al-Mandhari, A. (2020). *Forecasting the type 2 diabetes mellitus epidemic in Oman up to 2050: Mathematical modeling approach*. International Journal of Environmental Research and Public Health, 17(3), 1018.

Battineni, G., Sagaro, G. G., Amenta, F., & Tayebati, S. K. (2019). *Comparative machine-learning approach: A follow-up study on type 2 diabetes predictions by cross-validation methods*. Machines, 7(4), 74.

Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). *Machine learning based diabetes classification and prediction for healthcare applications*. Journal of Healthcare Engineering, 2021, 9930985.

Dash, S., Biswa, S., Banerjee, D., & Rahman, A. (2019). *Edge and fog computing in healthcare – A review*. Scalable Computing: Practice and Experience, 20(2), 191–205.

Fazakis, N., Kalatzis, O., Doulamis, S., Voulodimos, A., Protopapadakis, E., & Doulamis, N. (2021). *Machine learning tools for long-term type 2 diabetes risk prediction*. IEEE Access, 9, 10307–10315.

Gollapalli, M., Almahmoudi, A., Saraireh, L., Rahman, A., Biabani, S. A., Iqbal, T., Alqahtani, M., Alshammari, G., & Alharbi, A. (2024). *A retrospective data analysis on the breast cancer patients epidemiology and awareness in the eastern regions of Saudi Arabia*. International Journal of Design & Nature and Ecodynamics, 19(5), 1741–1751.

Gollapalli, M., Rahman, A., Kudos, S. A., Foula, M. S., Alkhalifa, A. M., Albisher, H. M., Al-Hariri, M. T., & Mohammad, N. (2024). *Appendicitis diagnosis: Ensemble machine learning and explainable artificial intelligence-based comprehensive approach*. Big Data and Cognitive Computing, 8(9), 108.

Hantom, W. H., & Rahman, A. (2024). *Arabic spam tweets classification: A comprehensive machine learning approach*. AI, 5(3), 1049–1065. https://doi.org/10.3390/ai5030052

Mallhi, I. Y., Umair, M. K., Norani, M. A., Shahzaib, S., Ali, M. U., Hameed, S., ... & Iqbal, M. J. (2025). Nutritional Strategies for Type 2 Diabetes: A Comprehensive Review on Current Evidence and Future Perspectives. Journal of Food and Agricultural Technology Research, 4(02), 1-18.

Joshi, R. D., & Dhakal, C. K. (2021). *Predicting type 2 diabetes using logistic regression and machine learning approaches*. International Journal of Environmental Research and Public Health, 18(14), 7346.

Khan, M. A., Abbas, S., Atta, A., Ditta, A., Alqahtani, M., Alshammari, G., Alharbi, A., & Rahman, A. (2020). *Intelligent cloud-based heart disease prediction system empowered with supervised machine learning*. Computers, Materials & Continua, 65(1), 139–151.

Khan, M.A.A. et al. (2025). *Harnessing Deep Learning to Arabic Sign Language Recognition*. International Journal of Intelligent Engineering and Systems 18 (8), 743-759.

Kopitar, L., Kocbek, P. K., Cilar, L., Sheikh, A., Stiglic, G., & Stiglic, M. (2020). *Early detection of type 2 diabetes mellitus using machine learning-based prediction models*. Scientific Reports, 10(1), 6440.

Li, W., Younas, S., Ali, F., Arqam, U., Abbas, M. S., Yousaf, M., ... & Imran, M. (2025). Non-Destructive Estimation of Water Fractions by Machine Learning Models During Freeze-Drying. Journal of Food Process Engineering, 48(7), e70174.

Mounika, V., & Sunil, D. S. (2021). *Prediction of type-2 diabetes using machine learning algorithms*. IEEE Access, 9, 97980–97986.

Musleh, D., Rahman, A., Almossaeed, H., Balhareth, F., Alqahtani, G., Alobaidan, N., Altalag, J., Alhiyafi, J., Alshahrani, S., & Alqahtani, M. (2026). *AI-enabled diagnosis using YOLOv9: Leveraging X-ray image analysis in dentistry*. Big Data and Cognitive Computing, 10(1), 16.

Musleh, D., Rahman, A., Alkherallah, M. A., AlBo-Hassan, M. K., Alawami, M. M., Alqahtani, M., Alshammari, G., Alharbi, A., & Alhiyafi, J. (2024). *Machine learning approach to cyberbullying detection in Arabic tweets*. Computers, Materials & Continua, 80(1), 1033–1054.

Nasir, A., Khan, M. A., Ali, U., Shafqat, A., Abid, A., Shahzad, M., ... & Riasat, K. (2024). Enhancing Beetroot Jam with Stevia: Antioxidant Properties and Quality Assessment During Storage. Journal of Food and Agricultural Technology Research, 3(02), 1-19.

Nasir, M. U., Khan, S., Mehmood, S., Khan, M. A., Rahman, A.-u., Hwang, S. O., Hamadi, H. A., & Mansoor, W. (2022). *IoMT-based osteosarcoma cancer detection in histopathology images using transfer learning empowered with blockchain, fog computing, and edge computing*. Sensors, 22(14), 5444.

Nasir, M. U., Zubair, M., Ghazal, T. M., Khan, M. F., Ahmad, M., Rahman, A.-u., Hamadi, H. A., Khan, M. A., & Mansoor, W. (2022). *Kidney cancer prediction empowered with blockchain security using transfer learning*. Sensors, 22(19), 7483.

Nicolucci, A., Romeo, L., Bernardini, M., Vespasiani, M., Rossi, M. C., Petrelli, M., Ceriello, A., Di Bartolo, P., Frontoni, E., & Vespasiani, G. (2022). *Prediction of complications of type 2 diabetes: A machine learning approach*.

Diabetes Research and Clinical Practice, 190, 110013.

Qayyum, F., Ali, F., Altaf, B., Rahman, A., Alqahtani, M., Alshammari, G., & Alharbi, A. (2026). *A feed-forward neural network–based framework for integrating special education programs into agriculture and food systems toward sustainable development*. Journal of Food and Agricultural Technology Research, 5(1), 12–21.

Rahman, A. (2025). *Solar panel surface defect and dust detection: Deep learning approach*. Journal of Imaging, 11(9), 287.

Rahman, A., Alatallah, F. A., Almubarak, A. J., Alkhazal, H. A., Alzayer, H. A., Alqahtani, M., Alshammari, G., Alharbi, A., & Youldash, M. (2025). *Deep learning-based automated inspection of generic personal protective equipment*. Computers, Materials & Continua, 85(2), 3507–3525.

Rahman, A., Youldhash, M., Alshammari, G., Sebiany, A., Alzayat, J., Alsayed, M., Alqahtani, M., & Alharbi, A. (2024). *Diabetic retinopathy detection: A hybrid intelligent approach*. Computers, Materials & Continua, 80(3), 4561–4576.

Rahman, A.-u., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M. A., & Mosavi, A. (2022). *Rainfall prediction system using machine learning fusion for smart cities*. Sensors, 22(9), 3504.

Rehman, A., Athar, A., Khan, M. A., Abbas, S., Fatima, A., Rahman, A., & Saeed, A. (2022). *Modelling, simulation, and optimization of diabetes type II prediction using deep extreme learning machine*. Journal of Ambient Intelligence and Smart Environments, 12(2), 125–138.

Sharma, A., Shah, M., & Das, R. N. (2020). *A review on machine learning algorithms for diabetes diagnosis*. International Journal of Advanced Research in Computer and Communication Engineering, 9(1), 12–15.

Singh, K. R., Dash, S., & Atta-ur-Rahman. (2024). *Detection of arrhythmia from ECG signal using bat algorithm-based deep neural network*. In B. Pati et al. (Eds.), ICACIE 2022 (Lecture Notes in Networks and Systems). Springer.

Sisodia, D., & Sisodia, D. S. (2018). *Prediction of diabetes using classification algorithms*. Procedia Computer Science, 132, 1578–1585.

Swapna, S. G., Vinayakumar, R., & Soman, K. P. (2018). *Diabetes detection using deep learning algorithms*. ICT Express, 4(4), 243–249.

Tigga, N. P., & Garg, S. (2020). *Prediction of type 2 diabetes using machine learning classification methods*. Procedia Computer Science, 167, 706–716.

Wee, B. F., Sivakumar, S., Lim, K. H., Tan, C. K., Tan, C. S., Tan, S. Y., Tan, S. C., & Wong, K. W. (2024). *Diabetes detection based on machine learning and deep learning approaches*. Multimedia Tools and Applications, 83, 24153–24185.

Wu, H., Yang, S., & Huang, Y. (2018). *Type 2 diabetes mellitus prediction model based on data mining*. Informatics in Medicine Unlocked, 10, 100–107.

Xie, Z., Ansari, O. N., & Sun, Y. (2019). *Building risk prediction models for type 2 diabetes using machine learning techniques*. Preventing Chronic Disease, 16, E130.

Youldash, M., Rahman, A., Alsayed, M., Sebiany, A., Alzayat, J., Aljishi, N., & Alshammari, G. (2024). *Early detection and classification of diabetic retinopathy: A deep learning approach*. AI, 5(4), 2586–2617.

Youldhash, M., Rahman, A., Alsayed, M., Sebiany, A., Alzayat, J., Aljishi, N., Alshammari, G., & Alqahtani, M. (2025). *Deep learning empowered diagnosis of diabetic retinopathy*. Intelligent Automation & Soft Computing, 40(1), 125–143.

Younas, S., Ali, F., Arqam, U., Wang, X., Ahmad, M., Tariq, F., ... & Ali, Z. (2026). Comparative assessment of ensemble machine learning models for detecting water dynamics in freeze-dried mushrooms using Vis-NIR spectroscopy. Journal of Food Measurement and Characterization, 1-16.

Younas, S., Ahmad, M., Arqam, U., Manzoor, M. S., Wang, X., Li, W., ... & Imran, M. (2025). Prediction of water status during freeze drying with genetic and whale algorithms of machine learning. International Journal of Refrigeration, 178, 347-358.

Younas, S., Manzoor, M. S., Arqam, U., Ali, F., Murtaza, A., Wahab, M. A., ... & Wang, X. (2024). Rapid and non-destructive determination of total phenolic contents using UV–NIR spectroscopy of dehydrated mushroom (Lentinus edodes). Journal of Food Process Engineering, 47(7), e14699.