Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

Prediction of Diabetic Disease through Big Data Techniques: A Comparative Study

Dr. P. Suresh Babu[1], Ms. M. Premavathi[2]

[1]Associate Professor, Bharathidasan College of Arts and Science, Erode
[2]Assistant Professor, Bharathidasan College of Arts and Science, Erode

Abstract: Medical data mining is the method of collecting, examining and interpreting the health information for improving the patient care and support decision-making. Medical data processing comprised the data collection, storage, analysis and prediction. Diabetes is a chronic condition characterized by impaired glucose metabolism. Diabetes poses the significant global health challenge. Diabetes resulted in chronic damage and dysfunction of different tissues like eyes, kidneys, heart, blood vessels and nerves. Diabetes prediction has received large research interest to emphasize the importance of predicting diabetes for early intervention and management strategies. Data pre-processing is the process of handling the missing the data points from input dataset. Feature selection is the process of selecting the relevant features from input database. With the selected features, the patient data classification is carried out. Different researchers carried out their research on different diabetes disease detection. But, the accuracy level was not enhanced and time complexity was not reduced. In order to address these issues, machine learning and deep learning based diabetic disease diagnosis is discussed in the survey article.

Keywords: Medical data mining, diabetes, intervention and management, disease diagnosis, feature selection

1. INTRODUCTION

Big data analytics played key issue in healthcare industries. Diabetes is considered as the common disease that caused elder people across worldwide. In International Diabetes Federation, 451 million people across world are diabetic in 2017. Diabetes is the chronic disease linked with abnormal state of human body. The blood glucose level is an impulsive one due to pancreas dysfunction lead to the type 1 diabetes and resistant to insulin. Diabetes is handled through treatment and medication. Individuals with diabetes addressed the secondary health issues like heart diseases and nerve damage. Age, obesity, lack of exercise, hereditary diabetes, living style, bad diet, high blood pressure, etc. resulted in Diabetes Mellitus. People with the diabetes have high risk of diseases.

The main contribution of the review article is formulated as:

- To perform efficient diabetic disease diagnosis methods with minimum time and higher accuracy
- To discuss different diabetic disease diagnosis methods with feature selection and classification process
- To compare the conventional diabetic disease diagnosis performance for attaining enhanced results

The structure of this paper is organized as: Section 2 presents the review of various diabetic disease diagnosis. Section 3 describes the diabetic disease diagnosis dataset used in this study. Section 4 explains the description of different diabetic disease diagnosis. Section 5 explains different performance metrics employed for diabetic disease diagnosis. The experimental results are explained in Section 6. Section 7 concludes the diabetic disease diagnosis work with future direction.

2. LITERATURE SURVEY

Early detection and treatment of diabetes prevent the complications and helped in minimizing the risk of severe health problems. Adamic–Adar Similarity Indexed Wald Boost Data Classification (AASIWBDC) Technique was introduced in [1] for efficient feature selection and data classification. Though the accuracy level was improved, the space complexity was not minimized by AASIWBDC Technique. The data mining and meta-heuristic

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

technique was introduced in [2] to forecast the diabetic patients. However, the time complexity was not reduced by data mining and meta-heuristic technique.

An optimized Light Gradient-Boosting Machine (Light GBM) and K-Nearest Neighbour (KNN) based ensemble algorithm was introduced in [3] for Type 2 Diabetic prediction. But, the classification accuracy was not enhanced by Light GBM and KNN based ensemble algorithm. An interpretable data-driven approach was introduced in [4] to forecast the cardio vascular risk for diabetes disease detection. Though the time complexity was reduced, the error rate was not minimized.

A novel ensemble learning model was introduced in [5] for addressing the diabetes prediction issues. The designed model increased augmenting accuracy to improve the predictive performance for diabetic patients. But, the data classification error was not minimized by novel ensemble learning model. An innovative Ensemble Deep Learning (EDL) clinical decision support system was introduced in [6] for diabetes prediction with high accuracy. But, the prediction error was not minimized by EDL.

Associative Kruskal Wallis and MapReduce Poly Kernel (AKW-MRPK) was introduced in [7] for early disease prediction. However, the computational cost was not reduced by AKW-MRPK. A data-driven approach was introduced in [8] to forecast the blood glucose level of Type 2 diabetes patients. But, the prediction time was not reduced by data-driven approach. An attention-enhanced Deep Belief Network (DBN) was introduced in [9] for early diabetes risk prediction with imbalanced datasets. But, the prediction accuracy was not enhanced by attention-enhanced DBN. A hybrid deep learning model was introduced in [10] for early stage diabetes risk prediction. But, the classification time was not reduced by hybrid deep learning model.

An automatic diabetes prediction system was designed in [11] with private dataset of female patients. But, the feature selection time was not minimized by designed system. The hyper-AdaBoost ML model was introduced in [12] with PIMA Indian Diabetes dataset. However, the computational cost was not reduced by hyper-AdaBoost ML model.

Stacking ensemble model was introduced in [13] for diabetes risk identification at earlier stage. But, the ensemble model failed to reduce the time consumption. A Gradient Boosting Machine and Data Reduction Unit (GBM-DRU) framework was introduced in [14] for diabetes detection. However, accuracy level was not enhanced by designed model. A Deep Neural Network (DNN)-based multi-layer perceptron (MLP) algorithm was presented in [15] for diabetes prediction. But, the complexity level was not reduced. A Stacking Recursive Feature Elimination and Isolation Forest was designed in [16] for diabetes prediction. However, it failed to perform efficient disease prediction in accurate manner.

Ensemble learning method was designed in [17] for diabetes detection. But, the early-stage disease diagnosis remained the key issue. A stacked ensemble model was designed in [18] to achieve improved diabetic diagnosis results. However, the designed model failed to limit ability from diverse data sources. A Bi-directional Long Short-Term Memory (Bi-LSTM) and Convolutional Neural Network was designed in [19] for diabetes diagnosis. However, the developed model failed to perform the multi-class analysis. A machine learning approach was introduced in [20] for efficient diabetes diagnosis. But, the designed approach failed to improve the diabetes prediction results.

Optimized Ensemble Boosting Fuzzified Neural Network (OEBFNN) was introduced in [21] with the support vector feature selection for Diabetic disease detection in early stage. But, the accuracy was not at required level. A monosulphide coated BTO-boosted Ag-based SPR biosensor was employed in [22] for efficient pre- and early-diabetes detection. However, the complexity level was not minimized. The predictive model was introduced in [23] for enhancing the diabetic disease detection.

DiaFog was introduced in [24] with remote users for real-time diabetic mellitus disease (DMD) diagnosis. DMD diagnosis was performed through ensemble deep learning (EDL) technique with IoT, cloud, and fog computing. The machine learning algorithm was introduced in [25] with statistical analysis for normal and

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

diabetic cases to distinguish influential features. Convolutional Neural Networks (CNNs) was designed in [26] with innovative 4D CNN model for early diabetes prediction. A region-specific dataset was used to improve diabetic prediction performance.

A machine learning framework was designed in [27] for diabetes prediction and diagnosis with the PIMA Indian dataset and Medical City Hospital (LMCH) diabetes dataset. A feature extraction approach was introduced in [28] for improving the accuracy through convolutional neural network and long short-term memory model. A deep reinforcement learning framework was designed in [29] for diabetes prediction. But, the designed framework failed to reduce the computational costs. A Convolutional Neural Network (CNN) was developed in [30] for Diabetes Risk Prediction to minimize error rate. However, CNN method failed to perform accurate diabetes prediction.

3. DATASET DESCRIPTION

The first dataset used is the diabetes disease prediction. The URL of the dataset is given as https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset. The prediction dataset comprised the set of medical and demographic data of every patient. Every patient data achieved the diagnosis status like positive or negative depending on diverse features. The patient data information is useful for healthcare professionals to identify patients at risk of developing diabetes. The patient data is provided with the personalized treatment plans. The diabetes disease prediction dataset contains the 9 features and 100000 data samples. The feature descriptions are listed in table 1.

Table 1 Attribute Description

S. No	Attributes	Description
1.	Gender	Gender denotes the biological sex of individual patient with categories like male, female and other
2.	Age	Patient age is diagnosed in older adults. Age value varies from 0-80 in the input dataset
3.	Hypertension	Hypertension is a medical condition where blood pressure in arteries is elevated. Hypertension has values a 0 or 1 1 represents presence of the hypertension 0 symbolizes absence of hypertension
4.	Heart_disease	Heart disease is a medical condition linked with increased diabetes risk. Heart disease has 0 or 1 1 represents presence of Heart disease 0 denotes absence of Heart disease
5.	Smoking_history	Smoking history is the risk factor for diabetes and increased the complications
6.	BMI	Body Mass Index (BMI) is employed to determine the body fat depending on the weight and height. High BMI values are connected to diabetes risk
7.	HbA1c_level	Hemoglobin A1c level is used to compute the person average blood sugar level over past 2-3 months High levels represents greater diabetes risk Greater than 6.5% of HbA1c level symbolizes the diabetes
8.	Blood_glucose_level	Blood glucose level denotes the amount of glucose in bloodstream at particular time period High blood glucose levels are essential diabetes indicator

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

9	Diabetes	Diabetes symbolizes the target variable 1 represents diabetes presence 0 denotes diabetes absence

The second dataset used is Diabetes 130-US Hospitals for Years 1999-2008. The URL of the dataset is given as https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008. Diabetes 130-US hospitals for years 1999-2008 dataset is employed for performing accurate diabetic disease diagnosis. The dataset consists of 47 attributes and 101766 data instances. The key objective of the input dataset is to identify the patient outcomes relate to the patients with diabetic disease. The dataset comprised the features like patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab tests conducted, HbA1c test result, diagnosis, number of medication, diabetic medication, number of outpatients, inpatient, and emergency visits in year before hospitalization, and so on. Among multiple features, relevant features are chosen for minimizing the diabetes disease diagnosis complexity. The dataset characteristics are multivariate in health and medicine application. The associated tasks of the dataset are used for classification and clustering. The feature type is categorical and integer. The patient information is extracted from input database for addressing the following criteria.

- It is an inpatient encounter
- It is a diabetic encounter where the diabetes is entered into system for efficient diagnosis.
- The length of the stay is atleast 1 day and atmost 14 days.
- Laboratory tests are carried out in encounter process
- Medications are administered in encounter process

4. METHODOLOGY

Diabetes Mellitus is the key problem disease where many people suffered from diabetic disease. The diabetic diseases are identified with different attributed like age, obesity, lack of exercise, diabetes, living style, bad diet, high blood pressure, etc. People with diabetes comprised the high disease risk such as heart disease, kidney disease, stroke, eye problem, nerve damage. The suitable treatment is provided for performing efficient diabetic disease diagnosis. With help of big data analytics, the hidden information is identified to extract the knowledge from input data and forecast the diabetic prediction results correspondingly. The architecture diagram of diabetic disease diagnosis is illustrated in the figure 1.

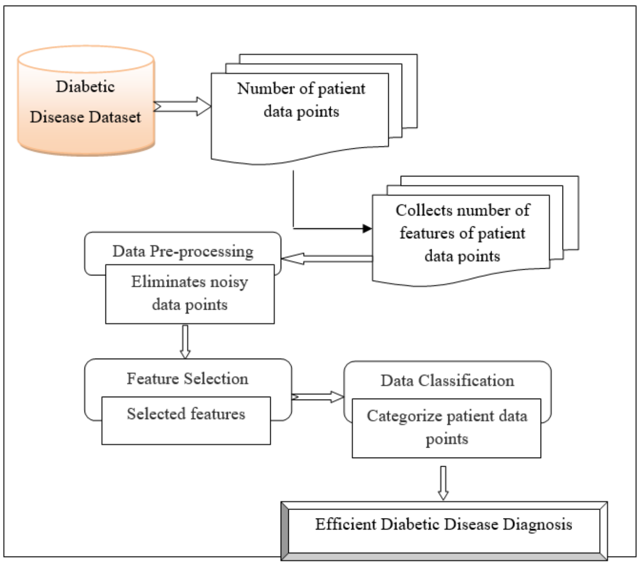

Figure 1 Architecture Diagram of Diabetic Disease Diagnosis

Figure 1 illustrates the architecture diagram of diabetes disease diagnosis with data pre-processing, feature selection and data classification. Initially, the input diabetes dataset is considered as input with number of features for diabetes disease diagnosis. In the first step, the data pre-processing is performed to remove the noisy data points from the input database. In the second step, the feature selection process is performed to enhance the disease diagnosis performance. In the third step, the patient data classification is carried out to categorize into disease presence and disease absence with help of selected features. The different methodologies are discussed for efficient diabetic disease diagnosis.

4.1 Adamic–Adar Similarity Indexed Wald Boost Data Classification

Big data comprised large quantity of information processed in future for attaining better prediction outcomes. Healthcare data analysts were important one for improving healthcare results through collecting, combining and examining data from different sources. Disease diagnosis was essential one for finding the disease through person symptoms and signs. Adamic–Adar Similarity Indexed Wald Boost Data Classification (AASIWBDC) Technique was introduced to carry out feature selection and data classification for increasing the diabetic disease diagnosis efficiency. AASIWBDC Technique employed Adamic–Adar Similarity Index to determine the similarity between features and objectives. With selected features, AASIWBDC Technique used Wald Boost Classification Process to classify the patient data into normal and abnormal groups. Wald Boost Classification Process improved the diabetic disease diagnosis accuracy. AASIWBDC Technique increased the classification performance with high diabetic disease diagnosis accuracy, diagnosis time and error rate using Diabetes 130-US hospitals dataset from 1999–2008. AASIWBDC Technique improved the diabetic disease diagnosis accuracy with lesser diagnosis time and error rate.

4.2 Cloud-Based Optimized Ensemble Model for Risk Prediction

An optimized Light Gradient-Boosting Machine (Light GBM) and K-Nearest Neighbour (KNN) based ensemble algorithm was introduced for forecasting the diabetic progression. Light GBM and KNN based ensemble algorithm was used for Type 2 Diabetes diagnosis. The designed algorithm was employed to classify into high or low risk with help of patient health parameters and serum measurements. LightGBM was used with KNN to categorize the data points. Light GBM and KNN based ensemble algorithm employed optimization methods like cross validation and grid search method to attain better results. LightGBM and KNN were used with voting classifier to identify the final class in effective manner. The analysis was carried out in Microsoft Azure cloud through Azure Machine Learning service. LightGBM and KNN improved scalability, security and integration possibilities into IoT-based smart healthcare systems. LightGBM and KNN increased the versatility and enhanced remote patient monitoring performance. Light GBM and KNN based ensemble algorithm increased Area under Curve (AUC), Receiver Operating Characteristics (ROC) score and classification efficiency.

4.3 Interpretable Data-Driven Approach

An interpretable data-driven approach was introduced to forecast the CVD risk for diabetic prediction. The designed approach present reasonable predictive performance and risk factor identification with CVDs. Diverse machine learning methods like KNN, decision tree, random forest and multilayer perceptron (MLP) are employed for diabetic disease prediction. The conditional tabular generative adversarial network (CTGAN) was employed to generate the synthetic data and enhance training data size. Different filter and wrapper feature selection (FS) method was employed for finding the most relevant features in CVD risk. SHAP and accumulated local effects were used to gain interpretability on predictive performance. ML based feature selection process was carried out for identifying the CVD risk. MLP reduced mean absolute error and mean relative absolute error. CVD risk prediction was carried out through different features, namely age, Hba1c and albuminuria. CVD risk factor identification plays important role in data-driven manner for early interventions and satisfactory treatments.

4.4 Multi-Model Ensemble Learning Approach

An advanced multi-modal ensemble learning approach was introduced to join different classification model for improving the predictive efficiency in precise and reliable manner. The prediction was carried out to find the individual diabetic status. The ensemble learning model was introduced for addressing the diabetes prediction performance. The designed technique joined Random Forest, Extra Trees, and Multilayer Perceptron (MLP) for efficient classification. The designed model was used for classifying the individuals into diabetic, non-diabetic or pre-diabetic. The designed approach was employed with the prediction score to compute the individual likelihood belonging to each respective category. The designed approach improved diabetes prediction performance with augmenting accuracy for diabetic patients.

4.5 Innovative Ensemble Deep Learning Clinical Decision Support System

Diabetes is global health concern with large number of diabetic people at risk. Diabetes is chronic disease and leads to large number of fatalities yearly. Early diabetic prediction is a significant one for preventing the progression and minimizing the complication risks like kidney and heart diseases. Ensemble Deep Learning (EDL) clinical decision support system was introduced for performing efficient diabetes prediction with high accuracy. EDL model employed Deep Learning (DL) architecture like neural network and ensemble learning stacking model. EDL was employed depending on stack ensemble model with stack-ANN, stack-CNN and stack-LSTM to enhance diabetic prediction performance. Pima Indian Diabetes Dataset (PIMA-IDD-I), Diabetes Dataset Frankfurt Hospital Germany (DDFH-G), and Iraqi Diabetes Patient Dataset (IDPD-I) were employed to train the patient data through EDL method. Extra Tree Classifier (ETC) approach was used to mine the relevant features from input database. EDL model improved the performance metrics like accuracy, precision, sensitivity, specificity, F-score, matthews correlation coefficient (MCC), and Area under curve (AUC).

4.6 Mapreduce based Big Data Framework

Machine Learning (ML) is used for performing the pattern recognition, object classification and disease prediction. ML technique is used to achieve the reasonable solution through computation and modeling method. The big data field has received large attention from industrial experts and academicians. The computation process was carried out to perform early disease prediction through ML for big data applications. Associative Kruskal Wallis and MapReduce Poly Kernel (AKW-MRPK) was introduced for performing early diabetic disease prediction. Initially, significant attributes are selected by applying Associative Kruskal Wallis Feature Selection model was introduced to parallelize the polynomial kernel vector through MapReduce depending on qualities gained for early disease diagnosis. AKW-MRPK framework increased the accuracy and speedup efficiency with lesser computational time.

4.7 Attention-Enhanced Deep Belief Networks

Diabetes mellitus is a chronic illness with difficult complications for performing efficient diagnosis. An attention-enhanced Deep Belief Network (DBN) was introduced for early diabetes prediction for highly imbalanced dataset. Sylhet Diabetes Hospital comprised the symptom and demographic information from patients. An ensemble feature selection approach was introduced to find the critical predictors. Generative Adversarial Networks (GANs) were employed to generate the synthetic data for increasing the model robustness. GAN was employed for identifying the underrepresented cases through addressing the class imbalance issues. A hybrid loss function used cross-entropy and focal loss to enhance the classification performance for identifying the hard-to-detect instances. An attention-based DBN model employed the synthetic data with hybrid loss function to enhance accuracy, f1-score, precision and recall. An effective approach was employed for early diabetes detection with clinical tool in healthcare settings.

4.8 Machine learning and explainable AI technique

Diabetes caused 537 million people for deadliest and non-communicable disease. Diabetes factors caused the person to get affected by diabetes such as excessive body weight, abnormal cholesterol level, family history, physical inactivity, bad food habit. People with diabetes for long time obtain different complications like heart

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

disorder, kidney disease, nerve damage, diabetic retinopathy etc. An automatic diabetes prediction system was introduced with female patients in Bangladesh. The designed system employed Pima Indian diabetes dataset from 203 individuals from local textile factory in Bangladesh. A semi-supervised model with extreme gradient boosting was used to forecast the insulin features. SMOTE and ADASYN method were employed to manage the class imbalance issue. The machine learning classification method like decision tree, SVM, Random Forest, Logistic Regression, KNN and ensemble techniques were used to attain the improved prediction results. The website framework and Android smartphone application was designed to predict the diabetes instantly.

5. EXPERIMENTAL ANALYSIS

The experimental result analysis of eight different methods, namely AASIWBDC Technique, Light GBM and KNN based ensemble algorithm, Interpretable data-driven approach, advanced multi-modal ensemble learning approach, EDL clinical decision support system, AKW-MRPK, Attention-Enhanced DBN and Automatic Diabetes Prediction System is performed with four different performance metrics. The performance metrics are listed as, diabetic diagnosis accuracy, precision, recall and diabetic diagnosis time.

5.1 Impact on Diabetic Diagnosis Accuracy:

Diabetic Diagnosis Accuracy is defined as ratio of number of data samples categorized as diabetes presence or absence out to the total number of data samples taken from input dataset. Diabetic diagnosis accuracy is computed as follows,

$$DDA = \left[\frac{TP+TN}{TP+TN+FP+FN}\right] * 100 \qquad (1)$$

From (1), 'DDA' symbolizes the diabetic diagnosis accuracy. 'TP' denote the true positive rate and false positive rate 'FP'. TN' symbolize the true negative rate, false negative rate 'FN'. The accuracy is computed in terms of percentage (%).When the accuracy level is higher, the technique is said to be more efficient.

5.2 Impact on Precision:

Precision is used to measure algorithm ability to accurately identify the true positives. Precision is defined as the ratio of correctly predicted true positive data samples. Precision is calculated as follows,

$$PE = \frac{TP}{TP+FP} \qquad (2)$$

From (2), 'PE denotes a precision. When the precision is higher, the method is said to be more efficient.

5.3 Impact on Recall:

Recall is otherwise termed as sensitivity. Recall is defined as the ability to correctly identify the proportion of actual true positive samples in the diabetes prediction. The recall is defined as,

$$RE = \frac{TP}{TP+FN} \qquad (3)$$

From (3), 'RE' represent the recall. When the recall is higher, the method is said to be more efficient.

5.4 Impact on Diabetic Diagnosis Time:

Diabetic diagnosis time is defined as the amount of time consumed for predicting the diabetes presence or absence from patient data samples. It is mathematically formulated as,

$$DDT = \sum_{i=1}^{n} R_i \; * Time(DP) \qquad (4)$$

From (4), 'DDT' symbolizes the diabetes prediction time. 'R' indicates a number of data samples. '$Time(DP)$' indicates the time to perform the diabetes diagnosis. The diabetic disease diagnosis time is measured in terms of milliseconds (ms).

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

6. RESULT AND DISCUSSION

The result and discussion of existing diabetic disease diagnosis methodologies is based on efficient heart disease detection. The performance measurement is carried out with dissimilar heart disease detection evaluation metrics to determine the efficiency. The performance of existing heart disease detection methods is discussed with table and graphs.

Table 2 Tabulation for Diabetes Disease Prediction Dataset

Methods/Parameters	Diabetic Diagnosis Accuracy (%)	Precision (%)	Recall (%)	Diabetic Diagnosis Time (ms)
AASIWBDC Technique	98	98.5	98.2	29
Light GBM and KNN based ensemble algorithm	74.66	74.44	75.11	45
Interpretable data-driven approach	84.5	86.4	87.5	54
Advanced multi-modal ensemble learning approach	86.4	85.4	86.4	51
EDL clinical decision support system	94.77	93.65	92.85	34
AKW-MRPK	86	86.5	84.6	44
Attention-Enhanced DBN	97.5	98	95	31
Automatic Diabetes Prediction System	92	92.4	93.1	38

Table 3 Tabulation for Diabetes 130-US Hospitals for Years 1999-2008 Dataset

Methods/Parameters	Diabetic Diagnosis Accuracy (%)	Precision (%)	Recall (%)	Diabetic Diagnosis Time (ms)
AASIWBDC Technique	97	96.1	96.8	32
Light GBM and KNN based ensemble algorithm	72.71	72.1	72.3	48
Interpretable data-driven approach	82.7	85.6	86.1	57
Advanced multi-modal ensemble learning approach	84.6	82.1	85.3	55
EDL clinical decision support system	93.14	92.2	91.12	36
AKW-MRPK	84	83.3	83.3	48
Attention-Enhanced DBN	96.2	97	94.6	32
Automatic Diabetes Prediction System	91	91.1	92.8	41

Table 2 and Table 3 describe the performance results for eight different diabetes diagnosis methods with four performance metrics for two different datasets respectively. From above table, it is clear that AASIWBDC Technique achieved better diabetes disease diagnosis performance results than any other conventional methods. For Diabetes Disease Prediction Dataset, AASIWBDC Technique attained 98% of diabetes diagnosis accuracy, 98.5% of precision, 98.2% of recall and 29ms of diabetes diagnosis time. For Diabetes 130-US Hospitals for Years 1999-2008 Dataset, AASIWBDC Technique attained 97% of diabetes diagnosis accuracy, 96.1% of precision,

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

96.8% of recall and 32ms of diabetes diagnosis time. Figure 2 and Figure 3 shows the performance metric analysis of Diabetes Disease Prediction Dataset and Diabetes 130-US Hospitals for Years 1999-2008 Dataset respectively.

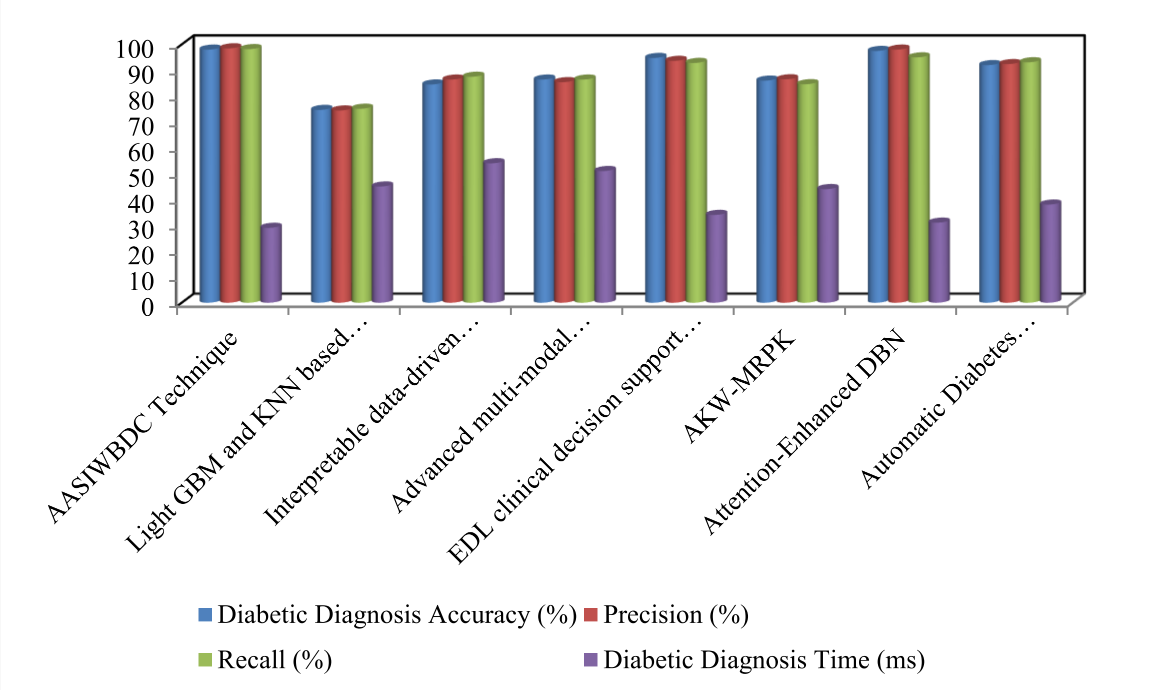

Figure 2 Measurement Analysis of Diabetic Diagnosis Accuracy, Precision, Recall and Diabetic Diagnosis Time of Diabetes Disease Prediction Dataset

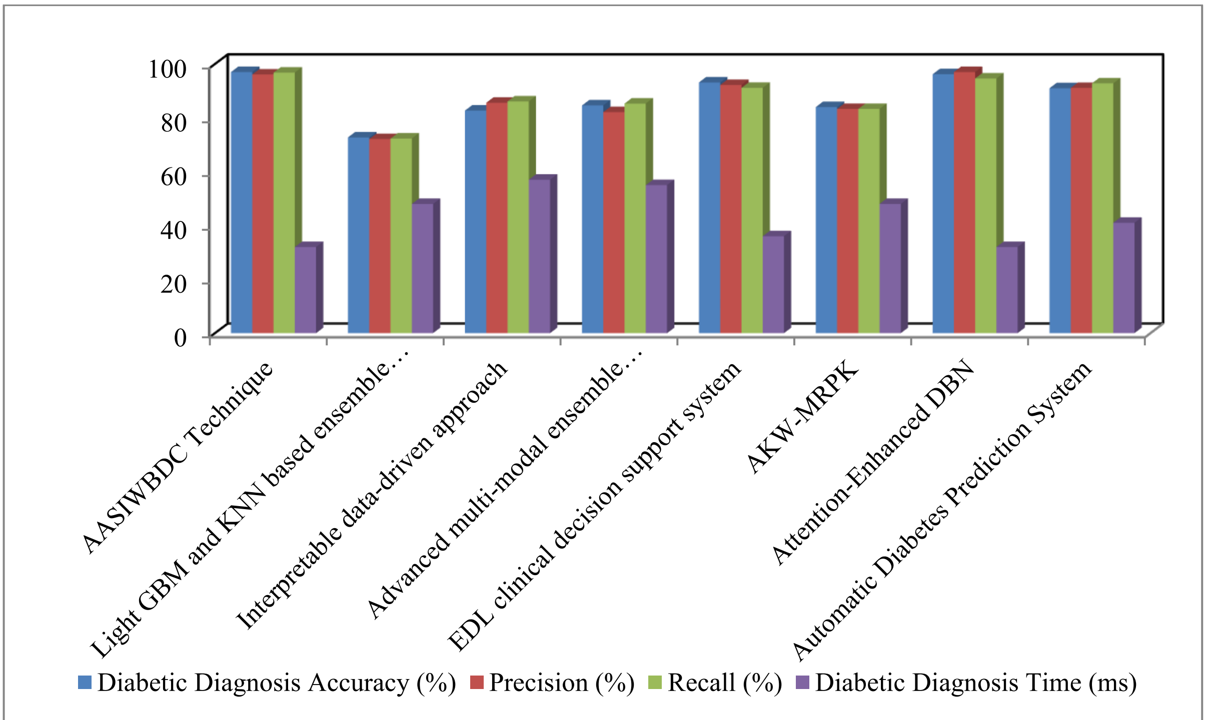

Figure 3 Measurement Analysis of Diabetic Diagnosis Accuracy, Precision, Recall and Diabetic Diagnosis Time of Diabetes 130-US Hospitals for Years 1999-2008 Dataset

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 1
Article Received: 25 January 2026 **Revised**: 18 March 2026 **Accepted**: 12 April 2026

Figure 2 and Figure 3 illustrates the performance metric comparison of eight different existing diabetic diagnosis techniques. From above graphical result analysis, it is clear that AASIWBDC Technique achieved higher diabetic diagnosis performance results than any other existing techniques. This is due to the application of Adamic–Adar Similarity Index and Wald Boost Classification Process to perform relevant feature selection and to classify the patient data into normal and abnormal groups for diabetic disease diagnosis. AASIWBDC Technique increased disease diagnosis accuracy and reduced the time complexity in effective manner. For Diabetes Disease Prediction Dataset, AASIWBDC Technique increased diabetic disease diagnosis accuracy by 12%, precision by 13%, recall by 12% and reduced the diabetic disease diagnosis time by 29% when compared to existing diabetic disease diagnosis techniques. For Diabetes 130-US Hospitals for Years 1999-2008 Dataset, AASIWBDC Technique increased diabetic disease diagnosis by 5%, precision by 4%, recall by 4% and reduced the diabetic disease diagnosis time by 1% when compared to existing diabetic disease diagnosis techniques.

7. CONCLUSION AND FUTURE WORK

In the survey, different diabetic disease diagnosis methods are discussed. In addition, different existing diabetic disease diagnosis methods are explained with methodologies. Different performance parameters are used to examine the diabetic disease diagnosis performance results. When the diabetic disease diagnosis result is compared with another study, ensemble learning concepts has attained better results. In this work, AASIWBDC Technique has attained higher accuracy during diabetic disease diagnosis. AASIWBDC Technique achieves highest precision and recall with lesser diabetic disease diagnosis time performance result than all the other existing techniques. These results strongly suggest that artificial intelligence with ensemble learning can be implemented in future for efficient diabetic disease diagnosis rather than the other conventional methods.

REFERENCES

[1] Ravichandran M, Saravanan Srinivasan, Sandeep Kumar Mathivanan, Hariharan Rajadurai, Benjula Anbu Malar M.B, Saurav Mallik and Hong Qin, “Adamic–Adar similarity indexed Wald boost data classification for diabetic disease diagnosis with big data”, Systems and Soft Computing, Elsevier, Volume 6, December 2024, Pages 1-15

[2] Masoomeh Zeinalnezhad and Saman Shishehchi, “An integrated data mining algorithms and meta-heuristic technique to predict the readmission risk of diabetic patients”, Healthcare Analytics, Elsevier, Volume 5, June 2024, Pages 1-18

[3] V. K. Daliya and T. K. Ramesh, “A Cloud-Based Optimized Ensemble Model for Risk Prediction of Diabetic Progression—An Azure Machine Learning Perspective”, IEEE Access, Volume 13, January 2025, Pages 11560 – 11575

[4] David Chushig-Muzo, Hugo Calero-Díaz, Francisco J. Lara-Abelenda, Vanesa Gómez-Martínez, Conceicao Granja and Cristina Soguero-Ruiz, “Interpretable Data-Driven Approach Based on Feature Selection Methods and GAN-Based Models for Cardiovascular Risk Prediction in Diabetic Patients”, IEEE Access, Volume 12, June 2024, Pages 84292 - 84305

[5] Rishi Jain, Nitin Kumar Tripathi, Millie Pant, Chutiporn Anutariya and Chaklam Silpasuwanchai, “Investigating Gender and Age Variability in Diabetes Prediction: A Multi-Model Ensemble Learning Approach”, IEEE Access, Volume 12, May 2024, Pages 71535 – 71554

[6] Mana Saleh Al Reshan, Samina Amin, Muhammad Ali Zeb, Adel Sulaiman, Hani Alshahrani and Asadullah Shaikh “An Innovative Ensemble Deep Learning Clinical Decision Support System for Diabetes Prediction”, IEEE Access, Volume 12, August 2024, Pages 106193 – 106210

[7] R. Ramani, S. Edwin Raja, D. Dhinakaran, S. Jagan and G. Prabaharan, “MapReduce based big data framework using associative Kruskal poly Kernel classifier for diabetic disease prediction”, MethodsX, Elsevier, Volume 14, June 2025, Pages 1-15

[8] Benzir Md Ahmed, Mohammed Eunus Ali, Mohammad Mehedy Masud, Mohammad Raihan Azad and Mahmuda Naznin, “After-meal blood glucose level prediction for type-2 diabetic patients”, Heliyon, Elsevier, Volume 10, Issue 7, 15 April 2024, Pages 1-18

[9] Olusola Olabanjo, Ashiribo Wusu, Olufemi Olabanjo, Mauton Asokere, Oseni Afisi and Boluwaji Akinnuwesi, “A novel deep learning model for early diabetes risk prediction using attention-enhanced deep belief networks with

highly imbalanced data", International Journal of Information Technology, Springer, Volume 17, 2025, Pages 1933–1955

[10] Mehmet Akif Bülbü, "A novel hybrid deep learning model for early stage diabetes risk prediction", The Journal of Supercomputing, Springer, Volume 80, 2024, Pages 19462–19484

[11] Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam and Riasat Khan, "Diabetes prediction using machine learning and explainable AI techniques", Institution of Engineering and Technology, Volume10, Issue 1-2, February-April 2023, Pages 1-10

[12] Abdulrahman Ahmed Jasim, Layth Rafea Hazim, Hayder Mohammedqasim, Roa'a Mohammedqasem, Oguz Ata and Omar Hussein Salman, "e-Diagnostic system for diabetes disease prediction on an IoMT environment-based hyper AdaBoost machine learning model", The Journal of Supercomputing, Springer, Volume 80, 2024, Pages 15664–15689

[13] Alfredo Daza, Carlos Fidel Ponce Sanchez, Gonzalo Apaza-Perez, Juan Pinto, Karoline Zavaleta Ramos, "Stacking ensemble approach to diagnosing the disease of diabetes", Informatics in Medicine Unlocked, Elsevier, Volume 44, 2024, Pages 1-22

[14] Turke Althobaiti, Saad Althobaiti, Mahmoud M. Selim, "An optimized diabetes mellitus detection model for improved prediction of accuracy and clinical decision-making", Alexandria Engineering Journal, Elsevier, Volume 94, May 2024, Pages 311-324.

[15] Abeer El-Sayyid El-Bashbishy & Hazem M. El-Bakry, "Pediatric diabetes prediction using deep learning", Scientific Reports, volume 14, 2024, Pages 1-20

[16] Nur Farahaina Idris,Mohd Arfian Ismail ,Mohd Izham Mohd Jaya,Ashraf Osman Ibrahim,Anas W. Abulfaraj,Faisal Binzagr, "Stacking with Recursive Feature Elimination-Isolation Forest for classification of diabetes mellitus", PLoS ONE, Volume 19, Issue 5, 2024, Pages 1-18

[17] Ifra Shaheen, Nadeem Javaid, Nabil Alrajeh, Yousra Asim, Sheraz Aslam, "Hi-Le and HiTCLe: Ensemble Learning Approaches for Early Diabetes Detection using Deep Learning and Explainable Artificial Intelligence", IEEE Access, Volume 12, 2024, Pages 66516 – 66538

[18] Khondokar Oliullah, Mahedi Hasan Rasel, Md. Manzurul Islam, Md. Reazul Islam, Md. Anwar Hussen Wadud, Md. Whaiduzzaman, "A stacked ensemble machine learning approach for the prediction of diabetes", Journal of Diabetes & Metabolic Disorders, Springer, Volume 23, 2024, Pages 603–617

[19] Parvathaneni Naga Srinivasu, Shakeel Ahmed , Mahmoud Hassaballah, Naif Almusallam, "An explainable Artificial Intelligence software system for predicting diabetes", Heliyon, Elsevier, Volume 10, Issue 16, 2024, Pages 1-19.

[20] Gangani Dharmarathne, Thilini N. Jayasinghe, Madhusha Bogahawaththa, D.P.P. Meddage, Upaka Rathnayake, "A novel machine learning approach for diagnosing diabetes with a self-explainable interface", Healthcare Analytics, Elsevier, Volume 5, 2024, Pages 1-13

[21] Thilagavathi G and Karthikeyan N.K, "Edge computing for diabetic early disease prediction based on optimized ensemble boosting Fuzzified neural network using recursive support vector feature selection", Biomedical Signal Processing and Control, Elsevier, Volume 108, October 2025, Pages 1-15

[22] Madhusudan Mishra, Sandipta Senapati, Archana Yadav and S.K. Tripathy, "BaTiO3 boosted silver-based SPR sensor for efficient urine-glucose detection in pre-diabetic and early-diabetic stages", Sensors and Actuators A: Physical, Elsevier, Volume 379, 1 December 2024, Pages 1-15

[23] Rafael Brito, Joao Lopes, Lucia Cerqueira, Vitor Barbosa, Carina Matos, Belen Blanco, Tiago Guimarães, Manuel Filipe Santos, "Precision Medicine for Diabetes: Improving the detection of diabetic patients using Predictive Analytics", Procedia Computer Science, Elsevier, Volume 238, 2024, Pages 950-955

[24] Abhilash Pati, Manoranjan Parhi, Binod Kumar Pattanayak, Debabrata Singh, Debabrata Samanta, Amit Banerjee, Sajal Biring, Goutam Kumar Dalapati, "Diagnose Diabetic Mellitus Illness Based on IoT Smart Architecture", Hindawi Publishing Coorporation, Wireless Communications and Mobile Computing, Volume 2022, 2022, Pages 1-20

[25] Mohammad Ehsan Farnoodian, Mohammad Karimi Moridani, and Hanieh Mokhber, "Detection and prediction of diabetes using effective biomarkers", Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, Volume 12, Issue 1, 2024, Pages 1-15

[26] Khoula Al Sadi and Wamadeva Balachandran, "Revolutionizing Early Disease Detection: A High-Accuracy 4D CNN Model for Type 2 Diabetes Screening in Oman", Bioengineering (Basel), Volume 10, Issue 12, December 2023, Pages 14-20

[27] Chollette C. Olisah, Lyndon Smith and Melvyn Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective", Computer Methods and Programs in Biomedicine, Elsevier, Volume 220, June 2022, Pages 1-15

[28] Furqan Rustam, Ahmad Sami Al-Shamayleh, Rahman Shafique, Silvia Aparicio Obregon, Ruben Calderon Iglesias, J. Pablo Miramontes Gonzalez and Imran Ashraf, "Enhanced detection of diabetes mellitus using novel ensemble feature engineering approach and machine learning model", Scientific Reports, Volume 14, Issue 23274, 2024, Pages 1-15

[29] Peter Domanski, Aritra Ray, Kyle Lafata, Farshad Firouzi, Krishnendu Chakrabarty, Dirk Pflüger, "Advancing blood glucose prediction with neural architecture search and deep reinforcement learning for type 1 diabetics", Biocybernetics and Biomedical Engineering, Elsevier, Volume 44, Issue 3, 2024, Pages 481-500

[30] Alawi Alqushaibi, Mohd Hilmi Hasan, Said Jadid Abdulkadir, Amgad Muneer, Mohammed Gamal, Qasem Al-Tashi , Shakirah Mohd Taib and Hitham Alhussian, "Type 2 Diabetes Risk Prediction Using Deep Convolutional Neural Network Based-Bayesian Optimization", Computers, Materials and Continua, Elsevier, Volume 75, Issue 2, 2023, Pages 3223-3238.