



Ensemble Machine Learning for the Classification and Prediction of Mellitus Diabetes

Koteswara Rao Kodepogu^{1*}, Eswar Patnala², Jagadeeswara Rao Annam³, Shobana Gorintla⁴, Veerla Vijaya Rama Krishna⁵, Vipparla Aruna⁶, Vijaya Bharathi Manjeti⁷, Anil Kumar Pallikonda¹

¹ Department of CSE, PVP Siddhartha Institute of Technology, Vijayawada 520007, India

² Department of CSE, Koneru Lakshmaiah Education Foundation, Vijayawada 522301, India

³ CSE, CVR College of Engineering, Vastunagar, Mangalpalli (V), Ibrahimpatnam (M), Hyderabad 501510, India

⁴ Department-CSE NRI Institute of Technology, Pothavarapadu Agiripalli Mandal, Vijayawada 521212, India

⁵ Department of Electronics & Communication, Lakireddy Bali Reddy College of Engineering, Mylavaram 521230, India

⁶ Department of CSE NRI Institute of Technology Pothavarapadu, Vijayawada 521212, India

⁷ Department of CSE, GITAM School of Technology, Visakhapatnam 530045, India

Corresponding Author Email: kkrao@pypsiddhartha.ac.in

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300701>

ABSTRACT

Received: 27 January 2025

Revised: 15 April 2025

Accepted: 22 April 2025

Available online: 31 July 2025

Keywords:

diabetes mellitus, ensemble machine learning, classification, adaboost, decision tree classifier

Nowadays, among the diseases with the greatest rate of growth and diabetes mellitus is the primary cause of illness and transience globally. Diabetes mellitus is the aggregate term for the metabolic disorders are defined by consistently increased blood glucose levels. The most important thing is to identify diabetic patients early on, since this reduces the individuals' chance of developing serious diseases. Early disease detection is made possible in large part by machine learning. This research presents the use of ensemble machine learning for the classification and prophecy of mellitus diabetes. The Pima Indians Diabetes Dataset, which was acquired from the UCI ML Origin, was used in this investigation. This is a proposal for a decision support system that classifies using the AdaBoost algorithm and Decision Stump as a decision tree classifier. 768 instances and 8 attributes made up the global dataset used for training by the system. It originated from the Irvine (UCI) ML origin at the University of California. For example, the ensemble AdaBoost and decision tree classifiers scored better with 95% F1-Score, 94% accuracy, 94% precision, and 95% recall. Experimental results essentially demonstrated that the comparison of overall performance outperforms well-known classifiers.

1. INTRODUCTION

People with diabetes mellitus face several health issues as a result of this chronic condition. It is predicted to increase by 642 million by 2040, having already reached 415 million worldwide [1]. The second-highest number of diabetes world has been identified in India. A heart attack, blindness, kidney failure, limb amputation, and even early death can result from late detection and inadequate treatment of diabetes mellitus, which typically shows no symptoms in its early stages [2]. Extremely elevated body fluid sugar echelons are an assurance of the clinical condition well-known as diabetes.

The bloodstream contains excess glucose, which can transform into a rich source of energy as it breaks down [3]. However, only if they get inside the body's cells are they broken down. The glucose molecules create deposits and turn into plaque when they stay in the bloodstream for a long duration of time rather than entering the cells [4]. Heart attacks, obesity, and other serious health problems are brought on by this plaque structure. Usually, persons over 40 experience this type of disease, but sometimes it occurs earlier.

Diabetes mellitus occurs in three main types, which are as follows: The immune system of the patient mistakenly targets

and kills the beta cells in the patient's digestive gland that produce insulin in the chronic illness known as type 1 diabetes [5].

When a patient has type-2 diabetes, resulting in decreased insulin production, which raises their blood sugar levels [6]. Recent research indicates that 80% of instances of type-2 diabetes can be avoided with early detection. "Pre-diabetes" is the term used to describe blood sugar levels that are too high to be classified as type 2 diabetes [7].

High blood sugar levels during pregnancy put a woman at risk for gestational diabetes. Insulin is normally secreted by the body to maintain blood glucose levels when an individual consumes food [8]. To support the growth of the fetus, the woman's placenta produces a hormone known as human placental lactogen during pregnancy. This hormone supports in the breakdown of the woman's body fats [9]. An oral glucose tolerance test is used to determine gestational diabetes. Typically, this test is performed between weeks 24 and 28 of gestation. The metabolic changes that occur in the latter stages of pregnancy are linked to this type of insulin resistance. As stated in the reason, the WHO intelligences that a significant proportion of children have type 1 diabetes. Therefore, diabetes can be considered a serious chronic disease [10].

Global health and welfare depend on the prevention and early detection of diabetes; therefore, an effective diabetes diagnosis and a mechanism for predicting the various conditions will affect the severity of diabetes are essential [11]. Predictive models for diabetes development therapies based on those risk factors [12].

In order to give people with a high-quality life, machine learning algorithms are an optimal option for early diabetes detection with a high level of confidence. Particularly in the medical sciences, classification is an essential tool for analysis and decision-making. Without explicit programming, machine learning (ML) algorithms can learn on their own [13]. Model validation is the process by which the tested data set can be used to assess the training model.

They evolve into either supervised or unsupervised algorithms, depending on the situation. Through examining datasets from individuals who have diabetes mellitus. In order to make an accurate diagnosis, they want to employ algorithms for machine learning. The categorization and prediction of mellitus diabetes using ensemble machine learning is presented in this study. This study made use of the Pima Indians Diabetes Dataset, which was obtained from the UCI ML Origin.

It includes 768 data of Pima Indian female patients who are 21 years of age or older. In addition to a binary class label indicative of the presence (1) or absence (0) of diabetes, eight clinical characteristics are included in each record: age, pregnancies, BMI, blood pressure, skin thickness, insulin level, glucose level, and diabetes pedigree function.

This dataset's organized nature and medically relevant features for diabetes prediction make it perfect for classification tasks with ensemble machine learning models. This paper makes use of AdaBoost and Decision Trees (DT). F1-Score, Accuracy, Precision, and Recall were the performance analysis parameters that were used. The structure of the remaining paper is as follows: Section 2 provides an explanation of the literature survey. Section 3 provides an explanation of the diabetes mellitus classification and prediction model. Results analysis is explained in Section 4. Section 5 presents the conclusion.

2. LITERATURE SURVEY

Lekha and Suchetha [14] investigate the application of a modified convolution neural network (CNN) algorithm that is one-dimensional (1-D) and integrates methods for both feature descent and classification. It is discovered that the method suggested in this research significantly reduces the limitations of applying these techniques separately, therefore enhancing the classifier's performance even further. This study employs a modified 1-D CNN to analyse real-time breath data obtained from several gas sensors. Testing and system performance are carried out and assessed.

A sparse representation classifier (SRC) and facemask chunk colour topographies are rummage-sale in Zhang and Zhang's study [15] to diagnose diabetic mellitus (DM) in a novel non-invasive way. A non-invasive apprehension device with image amendment is castoff to take the first image, which is created by precisely positioning four facemask blocks all over the place of the face. Results of the experiment are displayed for a dataset that includes 284 DM samples and 142 healthy samples. The SRC's average accuracy is 97.54% when using a combination of the face blocks to differentiate between

the Healthy and DM classes. Fitriyani et al. [16] state that a DPM that may antedate type 2 diabetes and hypertension in advance can be generated utilizing information on an entity's risk issues.

The data distribution is balanced using the SMOTET omek, diseases are predicted using an ensemble approach, and outlier data is eliminated using an iForest based outlier detection method. When unforeseen health concerns emerge in their early stages, the mobile application then receives the prediction result, allowing for prompt and suitable action to be made to reduce and eliminate a person's risks.

Nuankaew et al. [17] suggest a method to predict type 2 diabetes using variables that represent individual health conditions. In particular, as a requirement for a successful prediction model, based on the idea that every person has a diversity of wellbeing issues fetched on by various personal traits, this study suggests a unique prediction technique termed AWOD. Two datasets, each with 392 records, were examined in order to verify the recommended methodology. According on the comparative findings, the suggested approach obtained accuracy rates of 93.22% and 98.95% for Datasets 1 and 2, respectively., outperforming previous machine learning-based strategies.

Guo et al. [18] suggest a new Multi-Feature Complementary Learning (MFCL) model for the identification of DM. By generating feature-specific projections, many features are independently projected into a shared observation space and effectively combined into a single vector. Finally, an optimization approach is suggested to produce fused feature vectors by alternatively optimizing the projection variables. In terms of DM detection, the proposed method outperforms state-of-the-art techniques with an accuracy of 92.85% [19]. A unique DL architecture is suggested to identify if a being has diabetes Centered on a retinal scan. We use a very modest dataset to create our multi-stage CNN-based prototypical DiaNet, which can achieve over 84% accuracy on this test. Medical professionals in the field have confirmed that it effectively locates the areas of retinal pictures that influence its decision-making. Prognostic indicators for diabetes and other comorbidities, according to our studies, such as ischemic heart disease and hypertension, can be present in retinal images.

Anaya-Isaza and Zequera-Diaz [20] required a classification index such as the TCI. They then created 12 distinct data augmentation techniques using the deep CNN pattern, including 8 recently presented techniques and 4 conventional methods. By regulating the DM probability percentages for both foot pictures, a 100% detection was achieved. proving that the network's performance was enhanced by both the recommended and conventional methods [21].

A worker-centric, IoT-enabled, empowered framework for tracking users' health, well-being, and functional capacities is proposed with the aid of AI approaches. Diabetes is a chronic condition that is prevalent in both developed and developing nations and lowers people's quality of life. It has a high fatality rate. The ensemble Weighted Voting LRRFs ML model, which has an area under the ROC curve (AUC) of 0.884, is proposed to enhance diabetes prediction. A model for diabetes forecast using fused ML is presented by Ahmed et al. [22].

The two model types that made up the conceptual framework are SVM and ANN models. The proposed fused machine learning model performs better than the previously mentioned methods, with a prediction accuracy of 94.87.

To screen for type 2 diabetes, Zanelli et al. [23] offer a Light CNN-based model that uses a one unprocessed pulse taken from the photoplethysmogram.

We evaluate various model architectures that include designed photoplethysmography (PPG) features or biological sex and age as inputs in addition to the baseline design. When trained using age, biological sex, and raw PPG waves without the use of transfer learning, our model attained an AUC of 75.5, which is competitive with the state-of-the-art. Nurmi and Lohan [24] examined if diabetes can be accurately predicted by a single sensor or by a combination of sensors. When employing the XGBoost approach, we discovered that a multisensory combination consisting of glucose, electrocardiogram (ECG), and accelerometer (ACC) sensors produces the highest prediction accuracy of 98.2%. Despite their faster convergence, three-sensor combos that combine glucose, ECG, and performance curves have the same accuracy as four-sensor combinations in terms of ACC. To identify type 2 diabetes, the amount of acetone in exhaled breath is measured. To identify the disease, a new sensing module with a number of sensors is used to track the acetone concentration. CNN and other DL algorithms are frequently used to automatically evaluate medical data and generate predictions. When compared to present non-invasive methods, the suggested detection strategy and deep learning algorithm provide higher accuracy [25].

3. CLASSIFICATION AND PREDICTION OF MELLITUS DIABETES

Figure 1 demonstrates the block diagram for the categorization and prediction of mellitus diabetes using ensemble machine learning. Data from the UCI ML origin is used in the training phase. There are 8 characteristics and 768 illustrations in the dataset.

The attributes are,

- The total number of pregnancies a patient woman
- Level of glucose/sugar
- Diastolic blood pressure (measured in millimeters Hg)
- The patient's BMI
- The thickness of the skin fold in mm
- The value of insulin in two hours
- Pedigree function: a hereditary component
- The patient's age in years

Normalization is used to pre-process the diabetic dataset in the first phase. It is referred to as the pre-processing phase. The diabetes outcome prediction is impacted by these attributes' null values. Therefore, in order to create an effective machine learning model, data preparation is essential. There were some missing bits of information in the gathered dataset. An approach to deal with missing values is by After removing the missing value attribute and replacing zero for the missing values, replace them with the column's mean value. Using min-max normalization, diabetes datasets are pre-processed to transform all variables within the same range (0 and 1).

In machine learning, finding and eliminating relevant features from unprocessed data is known as feature extraction. A more useful dataset is then produced using these features, and it can be used for a number of tasks, including classification and prediction.

To extract the dataset's finest features, we utilized Scikit-learn's SelectKBest feature selection algorithm. All but the k highest score features are eliminated by the SelectKBest. The

"f_classif" function, which provides the ANOVA F-value between the label and feature for classification tasks, is the classification scoring function that is used. The optimal risk variables for Foretelling diabetes mellitus were determined using the F-value and p-value of each predictor.

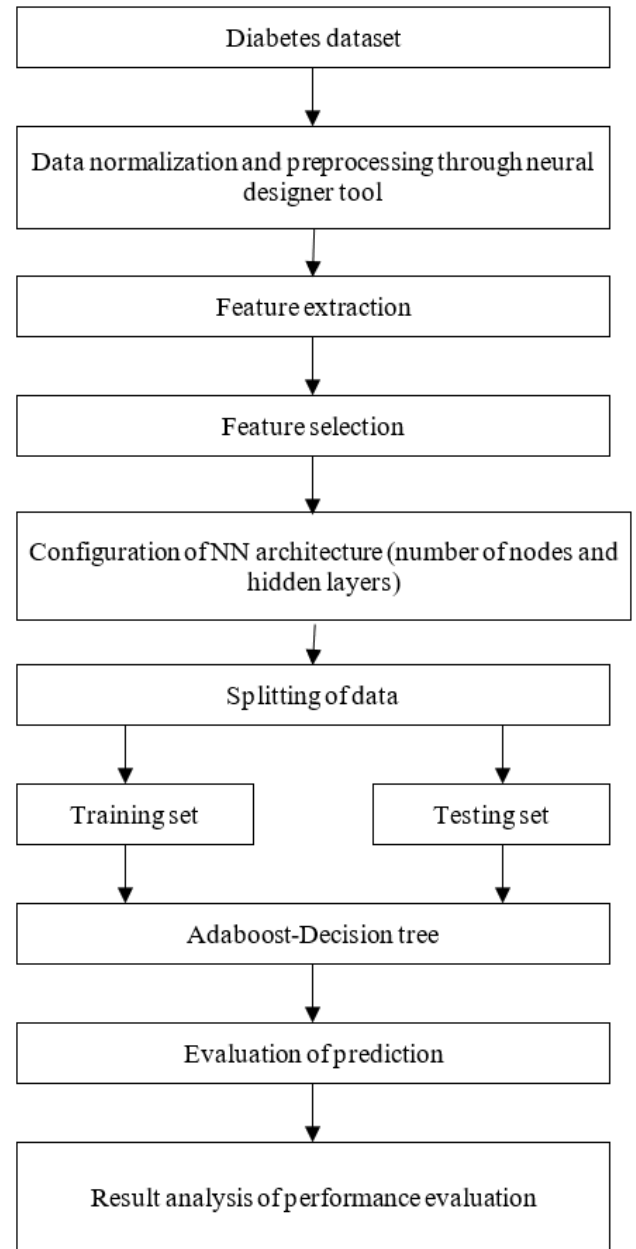


Figure 1. Block diagram of classification and prediction of mellitus diabetes

The three levels of a NN architecture are the neural network layer, the scaling layer, and the probabilistic layer. This neural network architecture has an 8:3:1 structure. Alternatively, there are three processing nodes, eight input nodes, and one output node. The complex structure of the model is represented by the three hidden neurons. For testing and training, a percentage split option is offered. 25% of the 768 instances are utilized for testing, while 75% are used for training. AdaBoost and DT utilized in this work for classification and prediction of mellitus diabetes. Here, a decision support system is suggested that classifies data using the AdaBoost algorithm and Decision Stump as a decision tree classifier.

A classification technique having root, internal, and leaf nodes is called a decision tree. The leaf node will be given the class label. Different test conditions are included in the root node and non-terminal node to differentiate between different attributes. The information obtained determines which root node is chosen. The branch is chosen based on the result of applying the condition to each record while keeping the beginning node. The property with the highest information gain value will be the root node. The purpose of this node is to add the extra attributes as children.

The training data is first given equal weight by the Adaptive Boosting (AdaBoost) algorithm. Developing a weak learner into a strong learner is the primary objective. Consider about the diabetes dataset's overall number of cases. Both positive and negative class labels are used. Decision tree is considered as classifier algorithm. Determine each instance's weight in the diabetic dataset. Give each sample in the diabetes dataset the same weight that was determined in the previous step. The decision tree classifier should be trained. Adjust (raise) the samples' weight if the diabetes sample is incorrectly identified. Reduce the weight of the cases if the diabetic sample is properly arranged. The iteration keeps going until the training error reaches zero or a user-specified value.

The effectiveness of categorizing algorithms was assessed by taking into account a number of recital meters, including F1-score, recall, accuracy, and precision.

4. RESULT ANALYSIS

This section evaluates the effectiveness of the ensemble machine learning approach for diabetes mellitus categorization and prediction that was previously addressed. Performance metrics such as F1-score, recall, accuracy, and precision are used to assess performance. The UCI machine learning repository supplied the global dataset's 768 instances and 8 attributes for training that the system uses. 25% of the 768 instances are utilized for testing, while 75% are used for training. Below are the definitions and computation formulas for these parameters:

Accuracy: It is described as the ratio of accurate forecasts to all forecasts. In Eq. (1), the formula is as follows:

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

Precision: It's described as the ratio of all samples that were accurately predicted to be true out of all samples that were projected to be true, even if they were incorrect. In Eq. (2), the precision parameter is represented,

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

Recall: This refers to the proportion of pertinent outcomes that the machine learning system forecasts. Another name for it is the true positive rate.

The following is the formula;

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

F1-score: It is a dataset that blends recall and precision in ML to assess a model's performance.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Table 1 contrasts the performance of the mentioned classification and prediction of mellitus diabetes using ensemble machine learning (AdaBoost+Decision tree) and prediction of mellitus diabetes using K-Nearest Neighbor (KNN) models in terms of precision, accuracy, F1-score, and recall parameters.

Figure 2 shows the graphical representation of accuracy parameter for Classification and Forecast of mellitus diabetes using described (AdaBoost+Decision tree) model and KNN model. In this graph, described model achieves high accuracy value compared to other models.

Table 1. Performance analysis

Parameters	Classification and Forecast of Mellitus Diabetes Model		
	AdaBoost+Decision Tree	KNN	Confidence Interval
Accuracy (%)	95	89	115 and 125mg/dL.
Precision (%)	94	90	115 and 125mg/dL.
Recall (%)	94	91	115 and 125mg/dL.
F1-Score (%)	95	88	115 and 125mg/dL.

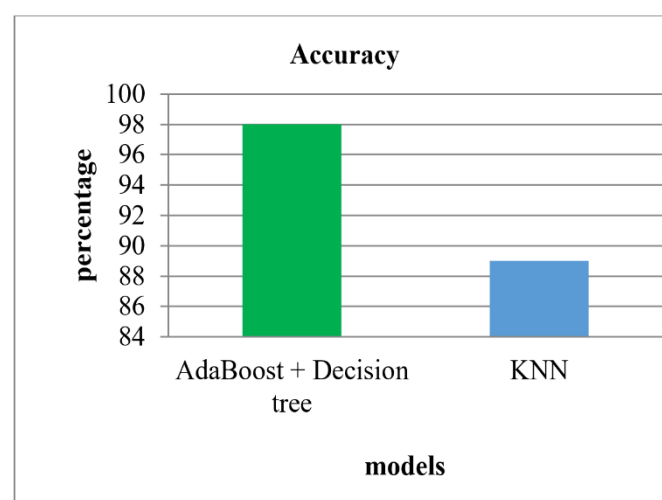


Figure 2. Accuracy parameter comparative analysis

Precision parameter comparative analysis is presented in Figure 3 for two models, based on classification and prediction of mellitus diabetes. Highest precision value (97%) is obtained for described model.

Comparative graphical analysis is represented in Figure 4 for classification and forecast of mellitus diabetes by means of described (AdaBoost+Decision tree) model and KNN model. 96% of recall value is obtained for described (AdaBoost+Decision tree) model.

F1-Score parameter graphical representation is elaborated in Figure 5 for two models. It is observed that described model is more efficient with high F1-Score compared to the other model.

Here's a simple comparison of computational efficiency (like training time) between **AdaBoost+Decision Tree** and **KNN**:

AdaBoost+Decision Tree

- **Training Time: High**
 - Trains multiple decision trees in sequence.
 - Each tree focuses on fixing errors of the previous one.
- **Prediction Time: Fast**
 - Final model is a weighted sum of trees, so prediction is quick.
- **Best For:** When training time is acceptable, but fast predictions are needed.

K-Nearest Neighbors (KNN)

- **Training Time: Very Low**
 - Just stores the data-no real "training."
- **Prediction Time: Slow**
 - Must calculate distance from test point to *all* training points.
- **Best For:** When training is quick and dataset is small.

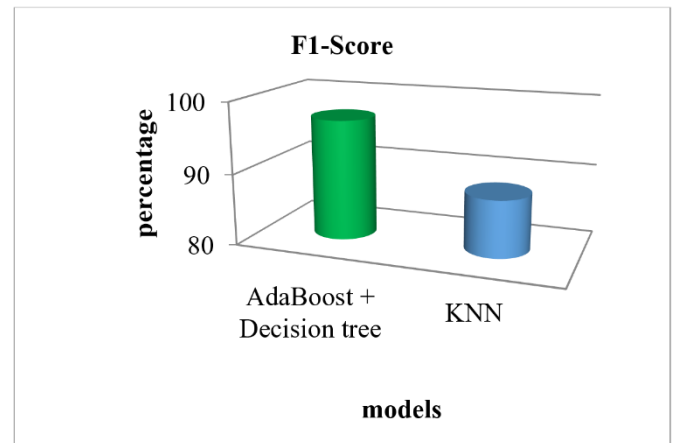


Figure 5. F1-Score parameter comparative analysis

5. CONCLUSION

The classification and prediction of diabetes mellitus using ensemble machine learning is described in this work. Early detection of diabetes lowers the hazards to the patient's health. Physicians, patients, and the patient's family may benefit from the prediction's outcomes. We introduced a technique for diabetes prediction that uses machine learning techniques. This is a proposal for a decision support system that classifies using the AdaBoost algorithm and Decision Stump as a decision tree classifier. The global dataset used for training was gathered from the UCI machine learning repository and consists of 768 instances and 8 characteristics. Out of 768 instances, 75% are used for teaching and 25% are used for testing. Precision, accuracy, F1-score, and recall are performance indicators used for performance evaluation. The equivalent attained values for the Accuracy, Precision, Recall, and F1-Score parameters are 98%, 97%, 96%, and 97%, respectively.

6. FUTURE WORK

Future research can go in a number of areas, despite the fact that ensemble machine learning algorithms have demonstrated encouraging results in the classification and prediction of diabetes mellitus. Future research can first concentrate on combining bigger and more varied datasets from other clinical and demographic sources. As a result, ensemble models will be more robust and generalizable across populations. Second, new health indicators including genetic markers, lifestyle factors, and data from continuous glucose monitoring can be incorporated into feature engineering to further enhance it and perhaps increase predicted accuracy.

The creation of hybrid ensemble models, which blend deep learning techniques with more conventional techniques like Random Forest and Gradient Boosting, is another avenue. Complex temporal patterns and non-linear correlations in patient data may be captured by such integration. Furthermore, explainability in ensemble learning is still a problem; future research should concentrate on interpretable ensemble frameworks that can give physicians clear insights for decision support.

Lastly, real-time predictive technologies that are implemented via cloud-based or mobile platforms can be investigated to facilitate early intervention and individualized

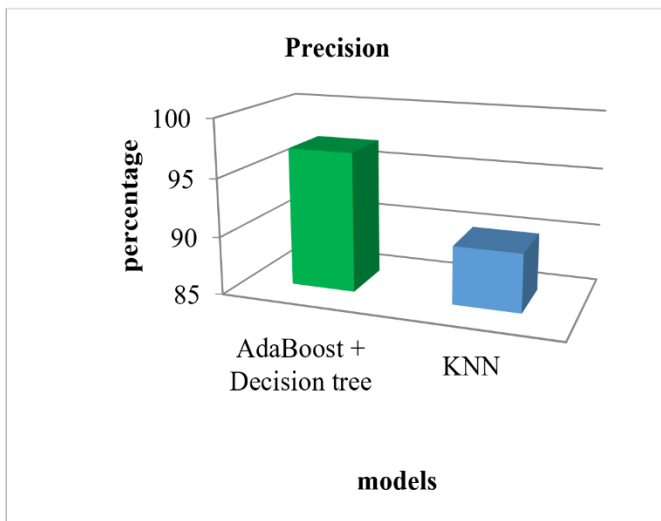


Figure 3. Precision parameter comparative analysis

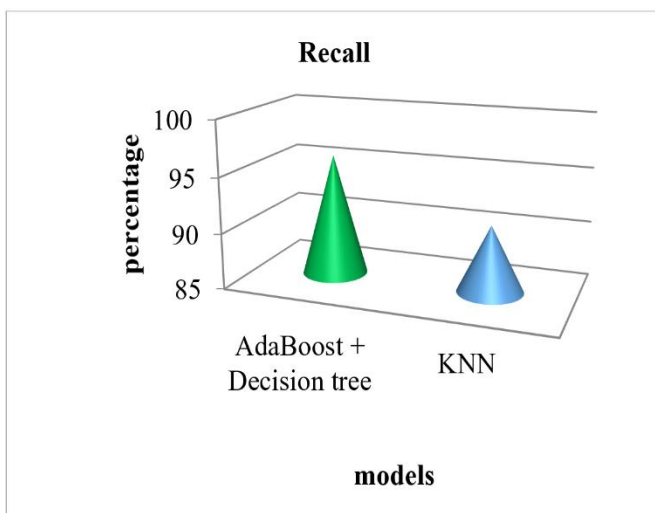


Figure 4. Recall parameter comparative analysis

From overall analysis, it is observed that, described classification and prediction of mellitus diabetes using ensemble machine learning (AdaBoost+Decision tree) is more efficient in terms of all parameters. The obtained values for the Accuracy, Precision, Recall, and F1-Score parameters are 98%, 97%, 96%, and 97%, respectively.

healthcare. To guarantee the dependability and practical usefulness of ensemble models in diabetes management, validation through clinical trials and cooperation with medical experts will be crucial.

REFERENCES

- [1] Marzouk, R., Alluhaidan, A.S., El_Rahman, S.A. (2022). An analytical predictive models and secure web-based personalized diabetes monitoring system. *IEEE Access*, 10: 105657-105673. <https://doi.org/10.1109/ACCESS.2022.3211264>
- [2] Alian, S., Li, J., Pandey, V. (2018). A personalized recommendation system to support diabetes self-management for American Indians. *IEEE Access*, 6: 73041-73051. <https://doi.org/10.1109/ACCESS.2018.2882138>
- [3] Kovács, L., Eigner, G., Siket, M., Barkai, L. (2019). Control of diabetes mellitus by advanced robust control solution. *IEEE Access*, 7: 125609-125622. <https://doi.org/10.1109/ACCESS.2019.2938267>
- [4] Yu, Z., Luo, W., Tse, R., Pau, G. (2023). DMNet: A personalized risk assessment framework for elderly people with type 2 diabetes. *IEEE Journal of Biomedical and Health Informatics*, 27(3): 1558-1568. <https://doi.org/10.1109/JBHI.2022.3233622>
- [5] Güemes, A., Cappon, G., Hernandez, B., Reddy, M., Oliver, N., Georgiou, P., Herrero, P. (2019). Predicting quality of overnight glycaemic control in type 1 diabetes using binary classifiers. *IEEE Journal of Biomedical and Health Informatics*, 24(5): 1439-1446. <https://doi.org/10.1109/JBHI.2019.2938305>
- [6] Lee, B.J., Kim, J.Y. (2015). Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning. *IEEE Journal of Biomedical and Health Informatics*, 20(1): 39-46. <https://doi.org/10.1109/JBHI.2015.2396520>
- [7] Bernardini, M., Romeo, L., Misericordia, P., Frontoni, E. (2019). Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine. *IEEE Journal of Biomedical and Health Informatics*, 24(1): 235-246. <https://doi.org/10.1109/JBHI.2019.2899218>
- [8] Gomes Filho, E., Pinheiro, P.R., Pinheiro, M.C.D., Nunes, L.C., Gomes, L.B.G. (2019). Heterogeneous methodology to support the early diagnosis of gestational diabetes. *IEEE Access*, 7: 67190-67199. <https://doi.org/10.1109/ACCESS.2019.2903691>
- [9] Lu, H.Y., Ding, X., Hirst, J.E., Yang, Y., Yang, J., Mackillop, L., Clifton, D.A. (2023). Digital health and machine learning technologies for blood glucose monitoring and management of gestational diabetes. *IEEE Reviews in Biomedical Engineering*, 17: 98-117. <https://doi.org/10.1109/RBME.2023.3242261>
- [10] Montaser, E., Diez, J.L., Rossetti, P., Rashid, M., Cinar, A., Bondia, J. (2019). Seasonal local models for glucose prediction in type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics*, 24(7): 2064-2072. <https://doi.org/10.1109/JBHI.2019.2956704>
- [11] Samreen, S. (2021). Memory-efficient, accurate and early diagnosis of diabetes through a machine learning pipeline employing crow search-based feature engineering and a stacking ensemble. *IEEE Access*, 9: 134335-134354. <https://doi.org/10.1109/ACCESS.2021.3116383>
- [12] Barakat, N., Bradley, A.P., Barakat, M.N.H. (2010). Intelligent support vector machines for diagnosis of diabetes mellitus. *IEEE Transactions on Information Technology in Biomedicine*, 14(4): 1114-1120. <https://doi.org/10.1109/TITB.2009.2039485>
- [13] Wang, Q., Cao, W., Guo, J., Ren, J., Cheng, Y., Davis, D.N. (2019). DMP_MI: An effective diabetes mellitus classification algorithm on imbalanced data with missing values. *IEEE Access*, 7: 102232-102238. <https://doi.org/10.1109/ACCESS.2019.2929866>
- [14] Lekha, S., Suchetha, M. (2017). Real-time non-invasive detection and classification of diabetes using modified convolution neural network. *IEEE Journal of Biomedical and Health Informatics*, 22(5): 1630-1636. <https://doi.org/10.1109/JBHI.2017.2757510>
- [15] Zhang, B., Zhang, D. (2013). Noninvasive diabetes mellitus detection using facial block color with a sparse representation classifier. *IEEE Transactions on Biomedical Engineering*, 61(4): 1027-1033. <https://doi.org/10.1109/TBME.2013.2292936>
- [16] Fitriyani, N.L., Syafrudin, M., Alfian, G., Rhee, J. (2019). Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *IEEE Access*, 7: 144777-144789. <https://doi.org/10.1109/ACCESS.2019.2945129>
- [17] Nuankaew, P., Chaising, S., Temdee, P. (2021). Average weighted objective distance-based method for type 2 diabetes prediction. *IEEE Access*, 9: 137015-137028. <https://doi.org/10.1109/ACCESS.2021.3117269>
- [18] Guo, C., Jiang, Z., Zhang, D. (2022). Multi-feature complementary learning for diabetes mellitus detection using pulse signals. *IEEE Journal of Biomedical and Health Informatics*, 26(11): 5684-5694. <https://doi.org/10.1109/JBHI.2022.3198792>
- [19] Islam, M.T., Al-Absi, H.R., Ruagh, E.A., Alam, T. (2021). DiaNet: A deep learning based architecture to diagnose diabetes using retinal images only. *IEEE Access*, 9: 15686-15695. <https://doi.org/10.1109/ACCESS.2021.3052477>
- [20] Anaya-Isaza, A., Zequera-Diaz, M. (2022). Detection of diabetes mellitus with deep learning and data augmentation techniques on foot thermography. *IEEE Access*, 10: 59564-59591. <https://doi.org/10.1109/ACCESS.2022.3180036>
- [21] Fazakis, N., Kocsis, O., Dritsas, E., Alexiou, S., Fakotakis, N., Moustakas, K. (2021). Machine learning tools for long-term type 2 diabetes risk prediction. *IEEE Access*, 9: 103737-103757. <https://doi.org/10.1109/ACCESS.2021.3098691>
- [22] Ahmed, U., Issa, G.F., Khan, M.A., Aftab, S., Khan, M.F., Said, R.A., Ghazal, T.M., Ahmad, M. (2022). Prediction of diabetes empowered with fused machine learning. *IEEE Access*, 10: 8529-8538. <https://doi.org/10.1109/ACCESS.2022.3142097>
- [23] Zanelli, S., El Yacoubi, M.A., Hallab, M., Ammi, M. (2023). Type 2 diabetes detection with light CNN from single raw PPG wave. *IEEE Access*, 11: 57652-57665. <https://doi.org/10.1109/ACCESS.2023.3274484>
- [24] Nurmi, J., Lohan, E.S. (2023). Machine-learning-based diabetes prediction using multisensor data. *IEEE Sensors*

Journal, 23(22): 28370-28377.
<https://doi.org/10.1109/JSEN.2023.3319360>
[25] Bhaskar, N., Bairagi, V., Boonchieng, E., Munot, M.V.
(2023). Automated detection of diabetes from exhaled

human breath using deep hybrid architecture. IEEE
Access, 11: 51712-51722.
<https://doi.org/10.1109/ACCESS.2023.3278278>