



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 8, August 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Diabetic Prediction Using Different Machine Learning Approaches

Dr. T.S. BASKARAN, Ms. B. MADHUVANTHI

Associate Professor & Research Supervisor, PG & Research Department of Computer Science, A Veeriyar Vandayar Memorial Sri Pushpam College (Autonomous), Poondi, Thanjavur, "Affiliated to Bharathidasan University, Tiruchirappalli", Tamil Nadu, India

Research Scholar, PG & Research Department of Computer Science, A. Veeriyar Vandayar Memorial Sri Pushpam College (Autonomous), Poondi, Thanjavur, "Affiliated to Bharathidasan University, Tiruchirappalli", Tamil Nadu, India

ABSTRACT: Diabetes is caused by an extended condition that regulates blood sugar. More than 30 million people in India are affected by diabetes and many are at risk. Therefore, early diagnosis and treatment are needed to prevent diabetes and avoid related health problems. An effective forecast is needed to evaluate the danger of diabetes among individuals based on their lifestyle and family history. The proposed work predicts and diagnoses different risk factors related to this disease using machine learning (ML) techniques. ML techniques provide efficient results for extracting knowledge from the diagnostic model. Suggestions and prescriptions depend on the lifestyle and machine learning technique applications.

KEYWORDS: Diabetes, classification algorithm and Machine learning technique

I. INTRODUCTION

Data mining using advanced pattern recognition and classification has played a main part in the development of expert systems in the medical field. Medical Expert Systems is an active area of research where data analysts and experts in ML are continuously determined to make them more accurate. The improved diagnostic systems with improved performance save the time of the medical experts. Moreover, these systems assist doctors and physicians in their clinical routine. According to the most recent data from the International Diabetes Federation, approximately 537 million adults (20-79 years) are living with diabetes. The total number of people living with diabetes is projected to rise to 643 million by 2030 and 783 million by 2045, 3 in 4 adults with diabetes live in low- and middle-income countries. Almost 1 in 2 (240 million) adults living with diabetes are undiagnosed. Diabetes caused 6.7 million deaths. Diabetes caused at least USD 966 billion in health expenditure 9% of total spending on adults. More than 1.2 million children and adolescents (0-19 years) are living with type 1 diabetes, 1 in 6 live births (21 million) are affected by diabetes during pregnancy and 541 million adults are at increased risk of developing type 2 diabetes [1]. Diabetes mellitus, more commonly referred to as diabetes, is one of the deadliest and most chronic diseases in the world. That occurs when the pancreas is no longer able to make insulin, or when the body cannot make good use of the insulin it produces. Insulin is a hormone made by the pancreas that acts like a key to let glucose from the food we eat pass from the bloodstream into the cells in the body to produce energy. All carbohydrate foods are broken down into glucose in the blood. Insulin helps glucose get into the cells [4]. Diabetes mellitus is classified as (i) Pre-diabetes - Blood glucose (also known as blood sugar) levels before diabetes are higher than normal. When the blood glucose level reaches a certain level, there is diabetes. It is a disease that occurs when the body does not produce or use the hormone insulin properly. Doctors sometimes denote pre-diabetes as impaired glucose tolerance (IGT) or impaired fasting glucose (IFG), depending on what test was used when it was detected. This condition increases the risk of developing type 2 diabetes and heart disease [5]. Type 1 - Also known as juvenile diabetes or insulin-dependent diabetes mellitus (IDDM), Type 1 diabetes is an autoimmune disease. The body's immune system does not recognize the cells in the pancreas that make insulin (beta cells) and attacks and destroys these cells. As a result, the body is unable to produce enough insulin, a hormone that regulates the amount of sugar in the blood. People with this disease must inject themselves with insulin to stay alive [11]. Type 2 - Also known as adult-onset diabetes or non-insulin-dependent

diabetes mellitus (NIDDM) is a chronic condition in which the body doesn't produce enough insulin and/or is unable to use insulin properly. Type 2 diabetes usually occurs in people who are over 40 years of age, overweight, and have a family history of diabetes, although today it is increasingly found in younger people. Type 2 diabetes is the most common form of diabetes. It's caused by a combination of factors but usually begins with insulin resistance [14]. Gestational diabetes - Gestational diabetes is a type of diabetes that can develop during pregnancy in women who don't already have diabetes. Diabetes means blood glucose is too high. Too much glucose in the blood is not good for the baby. Gestational diabetes is usually diagnosed in the 24th to 28th week of pregnancy. Managing gestational diabetes can help the baby stay healthy [3].

Early prediction of such diseases can be controlled over the diseases and save human life. To achieve this goal, this research work mainly explores the early prediction of diabetes by taking into account various risk factors related to this disease. Machine learning techniques provide an efficient result to extract knowledge by constructing predicting models from diagnostic medical datasets collected from diabetic patients. Extracting knowledge from such data can be useful to predict diabetic patients. Various machine learning techniques can predict diabetes mellitus. However, it is very difficult to choose the best technique to predict based on such attributes [13].

Hence, the main objective of this research is to predict diabetes in the future based on his or her present daily lifestyle activities and other indicators that might be used to predict the disease, suggestions, and prescriptions for the diabetic.

II. RELATED WORK

AyushAnand and Divya Shakti has discussed the diabetes risks likely to be developed from person daily lifestyle activities. The significant identification factor for blood pressure to diagnoses diabetes in CART prediction model [3].

VeenaVijayan V and Anjali C has predict diabetic which uses decision stump as base classifier in AdaBoost algorithm. The accuracy of decision support system can be improved with implementation of other classifier [2].

Maham Jahangir, Hammad Afzal, Mehreen Ahmed, KhawarKhurshid and Raheel Nawazhas presents an application of automatic multilayer perceptron (AutoMLP) which is combined with an outlier detection method Enhanced Class Outlier Detection using distance based algorithm to create a novel prediction framework for predict diabetes [7]. Md. Faisal Faruque, Asaduzzaman, and Iqbal H. Sarker has employ four popular machine learning algorithms, namely Support Vector Machine, Naive Bayes, K-Nearest Neighbor and C4.5 Decision Tree, on adult population data to predict diabetic mellitus [13].

NehaPrernaTigga and ShrutiGarg has discussed about to assess the risk of diabetes among individuals based on their lifestyle and family background. The risk of Type 2 diabetes was predicted using different machine learning algorithms as Random Forest algorithms are highly accurate, the algorithms were also applied to the Pima Indian Diabetes database and questionnaire like online and offline database [18].

Amine Rghioui 1, Jaime Lloret, Sandra Sendra and AbdelmajidOumnad has develop a monitoring system for diabetic patients using 5G technology and machine learning algorithms [19].

Saiteja Prasad Chatrati, Gahangir Hossain, AyushGoyal, AnupamaBhan, Sayantan Bhattacharya, Devottam Gaurav, and Sanju Mishra Tiwari has develop the application automatically detects the patient's disease status based on the algorithmic predictions from their blood pressure and glucose measurements as predictors or inputs [15].

Amine Rghioui, AssiaNaja, Jaime LloretMauri, and AbedlmajidOumnad has designed a new system for the diabetic patient monitoring system measures glucose level for the diabetic patients using glucose sensors and the data captured transmitted using NodeMCU with the help of the IoT platform to the database for storage, treatment, and discusses predictive analytics using four different machine-learning algorithms [20].

Munish Bhatia, Simranpreet Kaur, Sandeep K. Sood, and VeerawaliBehal has develop a comprehensive system of home- based early detection and prediction of urine-based diabetes (UbD). Diabetic Prediction and Decision Making allow an individual not exclusively to track his/her diabetes measure on regular basis but the prediction procedure is also accomplished so that prudent steps can be taken at early stages [16].

SkWasimHaidar, Dr. Surendra Pal Singh, and Dr. Prashant Johri has presents a new cloudenabled disease diagnosis and prediction model TLBO-KELM model. The TLBO-KELM model gets executed to diagnose and predict the existence of diseases using the patient data [17].

III. PROPOSED SYSTEM

In this section, develop a health care surveillance system to help diabetic patients respond to their situation and manage their illness themselves. The proposed system automatically records several data related to the health of diabetic patients. Figure 1 shows the proposed architecture for the diabetic system.

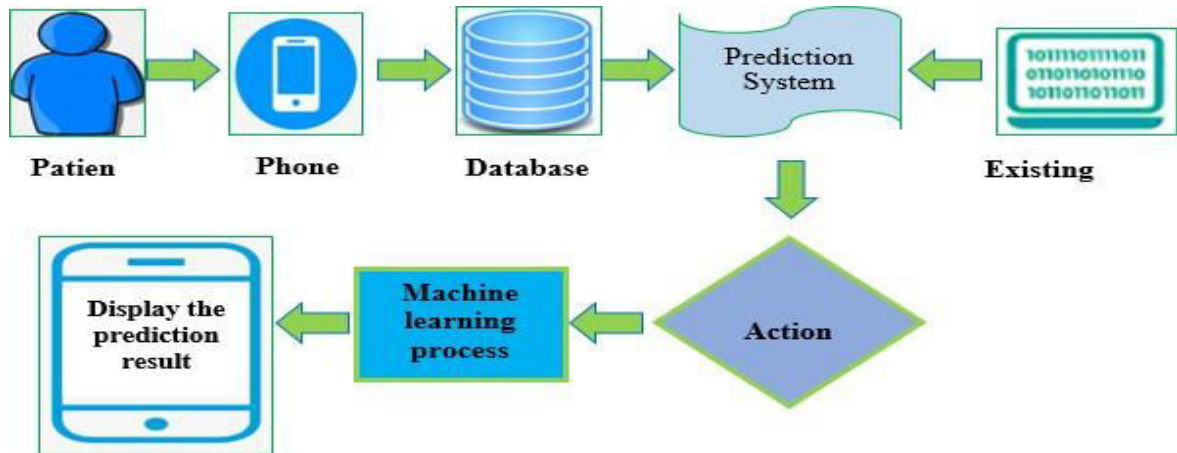


Figure 1 Diabetes system

1. The patient end that contains the mobile terminals powered with short range wireless connectivity with the CGM and activity data are transferred from these sensor devices to the smart phone automatically, and nutritional intakes are entered manually using a interface that is also used to access all components of the system.
2. The transformed data and history of the data stored by the diabetes database.
3. The prediction system contains the main menus to perform the action selection and machine learning algorithm to predict. The prediction result display the screen.

1) Action selection:

The diabetes system menu contains the CGM, physical activity, medicine, food, consolidate report and tips for manage diabetes. Figure 2 shows the action selection of the main menu.

Input data from the patient:

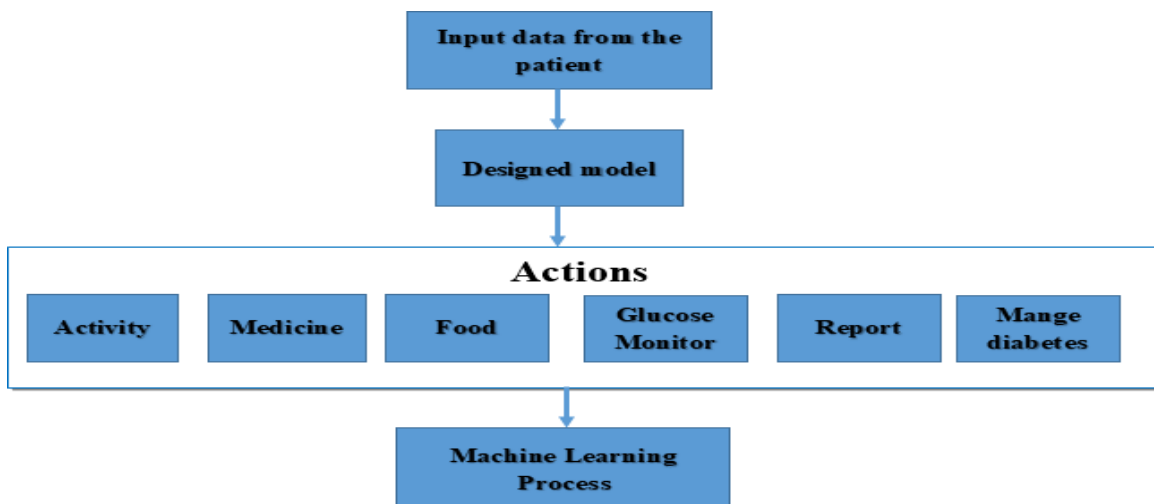


Figure 2 Diabetes system menu

Owing to the near-ubiquitous use of smart phone, we can use them to collect patient information such as CGM measurements taken at regular intervals, and the corresponding daily events that impact the CGM levels such as insulin, meals, exercise, and sleep. The ambient sensing features of a smart phone can help us collect some user information (e.g., exercise and sleep) automatically. Some other data (e.g., insulin dose) need to be entered manually. Because of its popularity, we mainly use touch-based gestures as the default input method. The collected data are pre-processed and automatically uploaded to the cloud.

Diabetes system actions:

- **Activity:** Activity trackers are devices that translate movement into different forms of data. Most trackers will provide estimates of steps, distance, and active minutes. The data can be viewed on a system.
- **Medicine:** Recommended the medicine and calculators intended to recommend adose of insulin.
- **Food:** It calculates a recommended daily calorie intake. It also has a well-designed food diary. The clear picture of how many calories have consumed during the day.
- **Glucose Monitor:** A sensor associated with the CGM is inserted under the skin, usually on the belly or arm. It tests glucose frequently in regular intervals (typically 5 to 15 minutes), and a wireless transmitter sends the glucose reading to a display system.
- **Report:** The report can be generated by past and current status of health.
- **Manage diabetes:** The diabetes system determining the required nutrition of patients and recommend meals to meet these needs, take their medication or insulin dose level, identify the factors that influence blood sugar, and educate patients on them.

2) Machine learning process

Figure 3, shows architecture diagram of machine learning process for diabetes prediction system.

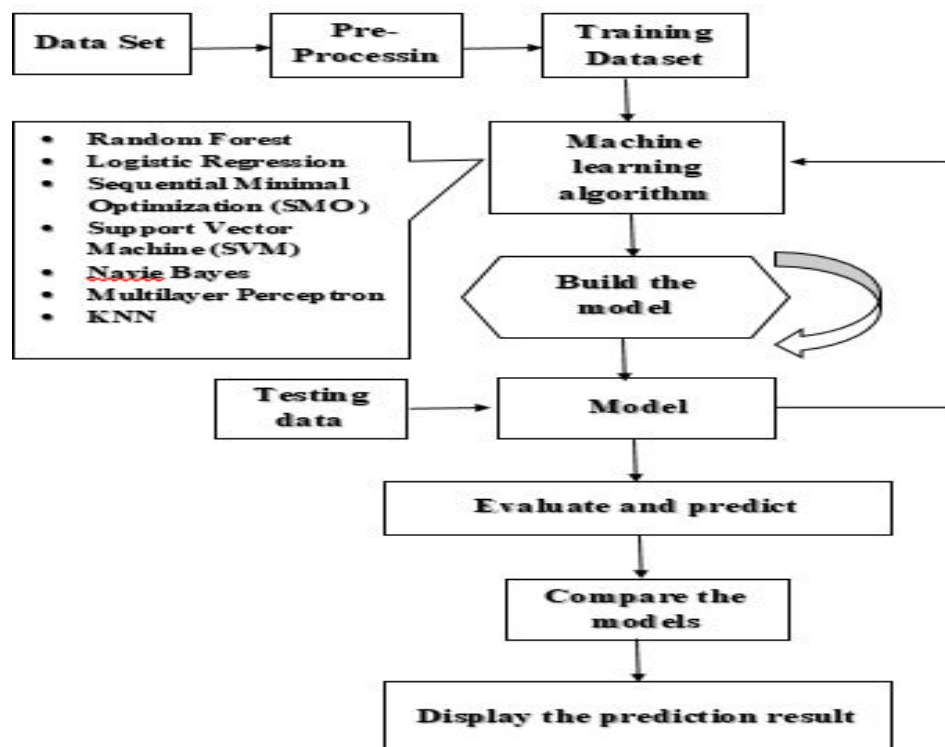


Figure 3 Working of diabetes system

Dataset collection

After defining the problem, collect the data from the Activity, Diabetes, Food, and MedicineData Storage.

Training and testing data

The training data set in Machine Learning is used to train the model for carrying out abundantactions. Detailed features are fetched from the training set to train the model. These structures are therefore combined into the prototype. In sentiment analysis, single words or sequences of consecutive words are taken from the tweets. Therefore, if the training set is labelled correctly, then the model will be able to acquire something from the features. So for testing the model such type of data is used to check whether it is responding correctly or not [11].

Diabetes Machine learning Techniques

For diabetic classification, we fine-tuned three widely used state-of-the-art techniques. Mainly, a comparative analysis is performed among the proposed techniques for classifying an individual in either of the diabetes categories. The details of the proposed diabetes techniques are as follows.

- **Random Forest**

As its name implies, it is a collection of models that operate as an ensemble. The critical idea behind RF is the wisdom of the crowd, each model predicts a result, and in the end, the majority wins. It has been used in the literature for diabetic prediction and was found to be effective. Given a set of training examples $X=x_1,2,\dots,x_m$ and their respective targets $Y=y_1,y_2,\dots,y_m$, Rf classifier iterates B times by choosing samples with replacement by fitting a tree to the training examples [12]. The training algorithmconsists of the following steps depicted in equation (1).

- (i) For $b=1\dots B$, samples with replacement n training examples form X and Y.
- (ii) Train a classification tree f_b on X_b and Y_b .

$$f = \frac{1}{B} \sum_{b=1}^B f_b(x') \tag{1}$$

- **Logistic Regression**

It is appropriate to use logistic regression when the dependent variable is binary, as we have to classify an individual in either type 1 or type 2 diabetes. Besides, it is used for predictive analysis and explains the relationship between a dependent variable and one or many independent variables, as shown in equation (2). Therefore, we used the sigmoid cost function as a hypothesis function ($h(X)$). The aim is to minimize cost function $J(\theta)$. It always results in classifying an example either in class 1 or class 2 [8].

$$h(x) = \frac{1}{1 + e^{-\theta^T x}}$$

$$J(\theta) = - \frac{1}{m} [\sum_{i=1}^m y^i \log h_{\theta}(x^i) + (1 - y^i) \log (1 - h_{\theta}(x^i))] \tag{2}$$

- **Sequential Minimal Optimization (SMO)**

SMO is a popular method to train support vector machines (SVM). Sequential Minimal Optimisation (SMO) addresses the quadratic optimization problem (QP) that resultfromusing SVM to classify datasets. SMO is able to decompose

the QP problem into smaller QP problems to be resolved using two Lagrange multipliers for each. The SMO algorithm is able to train a support vector machine using different kernels such as Polynomial Kernels, and Gaussian Kernels such as RBF. Nominal attributes that can be used to label the data are transformed into binary representations and all independent attributes are normalised. SMO algorithm is able to analyse large datasets using the methods describe and performs faster in comparison to linear SVMs and non-linear SVMs. SMO has been applied successfully applied in image classification and signal processing in a variety of domain areas e.g. biomedical information's with promising results [6].

- **Support Vector Machine (SVM)**

The occurrences of points in area is denoted by the SVM algorithm that are then plotted so that the classes are separated by strong gap. The goal is to determine the maximum-margin hyperplane which provides the greatest parting between the classes. The occurrences which is closest to the maximum-margin hyperplane are called support vectors. The vectors are chosen which are based on the part of the dataset that signifies the training set. Support vectors of two classes enable the creation of two parallel hyperplanes.

Therefore, larger the periphery between the two hyperplanes, better will be the generalization error of the classifier. SVMs are implemented in a unique way as compared with other machine learning algorithms [9].

- **Naive Bayes**

The Naïve Bayes Algorithm is a probabilistic algorithm that is sequential in nature, following steps of execution, classification, estimation and prediction. For finding relations between the diseases, symptoms and medications, there are various data mining existing solution, but these algorithms have their own limitations; numerous iterations, binning of the continuous arguments, high computational time, etc. Naïve Bayes overcomes various limitations including omission of complex iterative estimations of the parameter and can be applied on a large dataset in real time [11]. The algorithm works on the simple Naïve Bayes equation (3).

$$\text{Posterior Probability } (C|X) = \frac{\text{Likelihood } (X|C) * \text{Class Prior Probability } P(C)}{\text{Predictor Prior Probability } P(X)} \quad (3)$$

- **Multilayer Perceptron**

MLP is one of the most common ANN which is widely used for different tasks like pattern classification, pattern recognition etc. One of the most important features of MLP is that we can specify any number of output classes. The network architecture chosen for this problem is MLP having eleven input nodes and three output nodes. Each node present in the input layer is connected to every other node in the hidden layer through some weights. The value of the weighted input sum to a particular node maybe large, therefore it is important to scale down the weighted sum by reducing it before producing the resulted output of that particular node.

- **K-Nearest Neighbor (KNN)**

K-Nearest Neighbour is considered on the lazy learning methods. KNN depends on neighbour to predict the unknown tuple. KNN identify k neighbours for unknown tuple; it uses distance measure like Euclidean distance to calculate the distance between data points [25]. The most commonly used is the Euclidean distance, expressed in equation (4).

The Euclidean distance between (x_1, y_1) and (x_2, y_2) is,

$$(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Evaluate and predict

The three main metrics used to evaluate a classification model are accuracy, precision, and recall measure given in equation (5), (6) and (7) can be calculated. These matrices gave True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [10].

- **Accuracy:** The percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

- **Precision:** The fraction of relevant examples (true positives) among all of the examples which were

predicted to belong in a certain class.

$$\text{Precision} = \frac{TP}{(TP) + (FP)} \tag{6}$$

- **Recall:** The fraction of examples which were predicted to belong to a class with respect to all of the examples that truly belong in the class.

$$\text{Recall} = \frac{TP}{(TP) + (FN)} \tag{7}$$

Compare the models

After taking the input dataset the model will predict the data by applying the ML algorithms and provide the best result in the form of comparison between to predict the best accuracy, precision and recall to treat diabetes. The best model output can stored in database.

Display the prediction result

The system result depend on the menu option. The system provides knowledge about diabetes and some suggestions on the disease.

IV. RESULT AND DISCUSSION

In this session to evaluate one action. To select the Manage diabetes, this action can performed by the patient input data (testing data) and compare machine learning technique generate the models, evaluate the performance of different techniques.

Table 1 shows the performance evaluation of individual algorithms, while observing the accuracy, the highest value is shown by Logistic regression (81.77%) and the lowest value is shown by Support Vector Machine (75.76%).

Table 1 Performance evaluation of individual algorithm

Algorithm	Precision	Recall	Accuracy (%)
Random forest	0.767	0.766	76.56
Logistic regression	0.814	0.818	81.77
Sequential Minimal Optimization	0.791	0.797	79.67
Support Vector Machine	0.757	0.759	75.76
Navie Bayes	0.781	0.781	78.13
Multilayer Perceptron	0.786	0.792	79.17
K-Nearest Neighbour	0.764	0.768	77.43

In order to evaluate the performance of different machine learning techniques, have shown the results in Figure 4 on the basis of precision and recall. The Figure 4 shows the results of various machine learning techniques such as Random forest, Logistic regression, Sequential

Minimal Optimization, Support Vector Machine, Navie Bayes, Multilayer Perceptron, and K-Nearest Neighbour. If observe Figure 4, see that classifier Logistic regression achieves better results then other classifiers to predict diabetes. According to Figure 4 Logistic Regression achieves the 81% precision and 82% recall on this dataset.

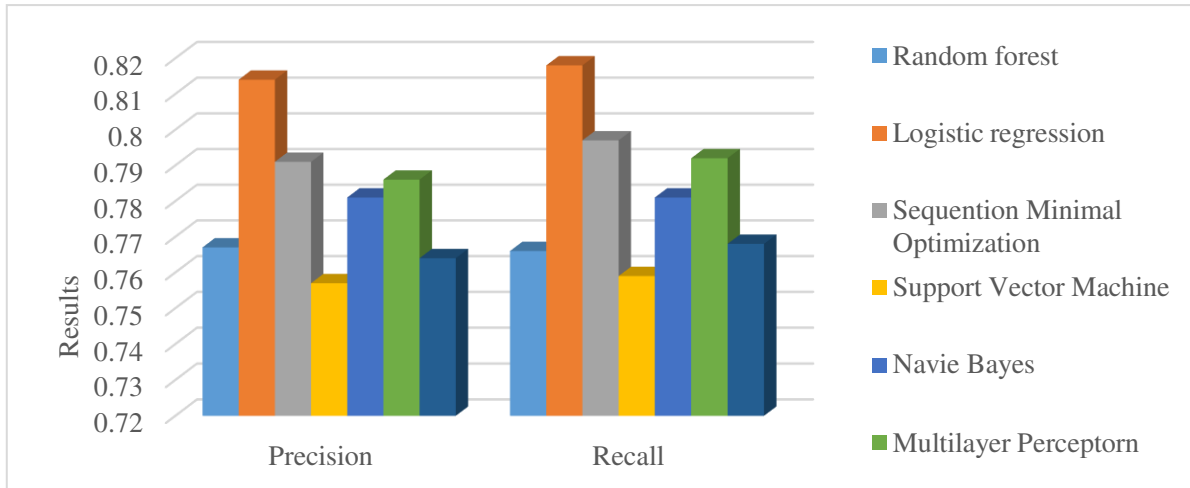


Figure 4 Prediction results of various machine learning techniques

In addition to precision and recall, also calculate the direct accuracy rate in percentage of all these classifiers shown in Figure 5. If we observe Figure 5, also see that Logistic Regression technique outperformance then other techniques.

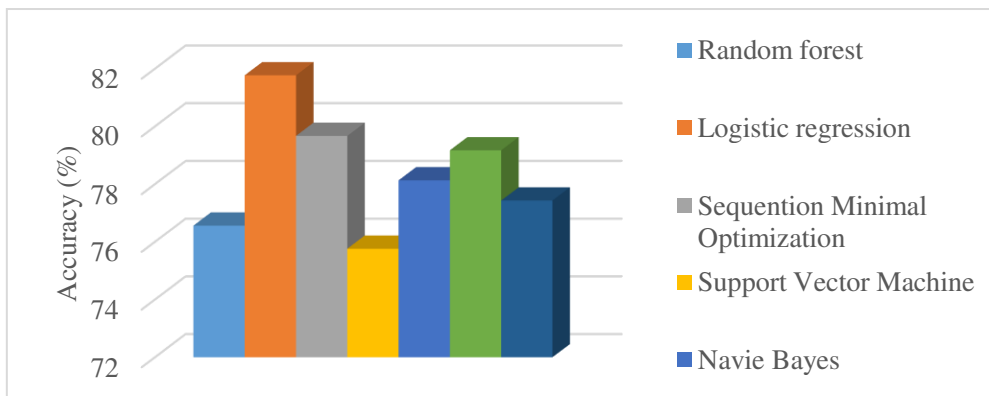


Figure 5 Accuracy result of various machine learning technique

Overall, have chosen the best machine learning techniques to predict diabetes to achieve highperformance, based on the evaluation criteria discuss above.

V. CONCLUSION

The major contribution of this paper is the development of a user-friendly graphical user interface-based system for the patients where the application monitors and analyses their glucose measurements, physical activities, medicines, and food intake. The application automatically detects the patient’s diabetes status based on the algorithmic predictions from their measurements as predictors or inputs.

FUTURE SCOPE

In future this system can be designed for any prediction of any other disease such as cancer, thyroid, lung diseases etc., if these are an android application of such disease prediction would be of great use in the near future.

Data set:

- <https://data.world/uci/human-activity-recognition/workspace/file/filename>
- /activity_labels.csv
- <https://www.indianmedicinedatabase.com/>
- <https://www.kaggle.com/vaishnavivenkatesan/food-and-their-calories>
- <https://www.kaggle.com/uciml/pima-indians-diabetes-database>

REFERENCES

1. <https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html>.
2. Anand, "PREDICTION OF DIABETES BASED ON PERSONAL LIFESTYLE INDICATORS," no. September, pp. 4–5, 2015.
3. V. V. V, "Prediction and Diagnosis of Diabetes Mellitus -A Machine Learning Approach," no. December, pp. 122–127, 2015.
4. Swain and A. Sample, "COMPARATIVE ANALYSIS ON PREDICTION OF DIABETES MELLITUS USING MACHINE," pp. 3312–3317, 2016.
5. D. Sejdinovic, L. G. Pokvic, A. Badnjevic, and M. Malenica, "CLASSIFICATION OF PREDIABETES AND TYPE 2 DIABETES USING ARTIFICIAL CLASSIFICATION OF PREDIABETES AND TYPE 2 DIABETES USING ARTIFICIAL NEURAL," no. September, 2017.
6. H. Afzal and M. Ahmed, "An Expert System for Diabetes Prediction using Auto Tuned Multi-Layer Perceptron," no. September, pp. 722–728, 2017.
7. M. Chokr and S. Elbassuoni, "Calories Prediction from Food Images," pp. 4664–4669, 2017.
8. D. Paul, "Prediction using Machine Learning," 2018 IEEE 9th Annu. Inf. Technol. Electron. Mob. Commun. Conf., pp. 924–928, 2018.
9. S. Antonio and H. Study, "Predicting Diabetes in Healthy Population through Machine Learning," 2019.
10. P. Sonar and K. Jayamalini, "DIABETES PREDICTION USING DIFFERENT," 2019 3rd Int. Conf. Comput. Methodol. Commun., no. Iccmc, pp. 367–371, 2019.
11. Mujumdar and V. Vaidehi, "ScienceDirect ScienceDirect ScienceDirect
12. ScienceDirect Diabetes Prediction using Machine Learning Aishwarya Mujumdar Diabetes Prediction using Machine Learning Aishwarya Mujumdar Aishwarya," *Procedia Comput. Sci.*, vol. 165, pp. 292–299, 2019.
13. K. Vijiyakumar, B. Lavanya, I. Nirmala, and S. S. Caroline, "Random Forest Algorithm for the Prediction of Diabetes," 2019 IEEE Int. Conf. Syst. Comput. Autom. Netw., pp. 1–5, 2019.
14. F. Faruque, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus," 2019 Int. Conf. Electr. Comput. Commun. Eng., pp. 1–4, 2019.
15. R. M. Khalil and A. Al-jumaily, "Machine Learning Based Prediction of Depression among Type 2 Diabetic Patients," pp. 0–4, 2020.
16. S. Prasad, G. Hossain, A. Goyal, A. Bhan, and S. Bhattacharya, "Smart home health monitoring system for predicting type 2 diabetes and hypertension," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2020.
17. M. Bhatia, S. Kaur, S. K. Sood, and V. Behal, "Journal Pre-proof," *Artif. Intell. Med.*, p. 101913, 2020.
18. N. P. Tigga and S. Garg, "ScienceDirect ScienceDirect Prediction of Type 2 Diabetes using Machine Learning Prediction of Type 2 Diabetes using Machine Learning Classification Methods Classification Methods," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 706–716, 2020.
19. S. S. and A. O. Rghioui, Amine, Jaime Lloret, "A Smart Architecture for Diabetic Patient Monitoring Using Machine Learning Algorithms," 2020.
20. Publication, "IOT BASED DIABETES DISEASE DIAGNOSIS AND CLASSIFICATION MODEL USING MACHINE IOT BASED DIABETES DISEASE DIAGNOSIS," 2020.
21. Rghioui, A. Naja, J. L. Mauri, and A. Oumnad, "An IoT Based diabetic patient Monitoring System Using Machine Learning and Node MCU An IoT Based diabetic patient Monitoring System Using Machine Learning and Node MCU," 2021.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details