Exploring Machine Learning Techniques for Diabetes Prediction in Healthcare

Ukani Vedant[a*], Sukhadiya Devarsh[b], Nishith Parmar[c]

[a] *Information Technology, Gandhinagar Institute of Technology - 382721, India*

[b] *Information Technology, Gandhinagar Institute of Technology - 382721, India*

[c]*Assistant Professor, Gandhinagar Institute of Technology - 382721, India*

Abstract

Diabetes is a chronic illness affecting millions of people worldwide and has the potential to cause a global healthcare crisis. It is a long-term health condition and is characterised by high blood glucose levels, resulting in symptoms such as frequent urination, blurry vision, increased thirst, and increased hunger. Diabetes harms major organs such as the eyes, kidneys, heart, and brain. Detecting diabetes early on can significantly reduce the risk of developing this serious health condition. Machine learning is a scientific field that teaches machines to learn from their experiences. This work intends to create a system for predicting early diabetes by comparing the results of various techniques for machine learning like K-Nearest Neighbor (KNN), Logistic Regression, Decision Tree, and Support Vector Machine (SVM). We utilised two datasets for our machine learning classification tasks: the PIMA Indian diabetes dataset, which is publicly available as an open-source dataset, and a 130-US Hospitals dataset. Further, the combined dataset is employed for each algorithm, and the performance of these algorithms is evaluated on various factors like sensitivity, accuracy, recall, f-measure, and specificity. The number of accurately and inaccurately identified samples is used to estimate accuracy. Based on the evaluation results, the algorithm with the highest performance measure is selected for predicting diabetes.

Keywords: Diabetes, Decision tree, KNN, Logistic Regression, SVM, Accuracy

1. Introduction

1.1. Diabetes Mellitus

Diabetes mellitus is often known as diabetes. Diabetes is a chronic condition that causes problems in the body when the pancreas does not generate enough insulin, which leads to poor carbohydrate metabolism and elevated blood glucose levels. Even among youngsters, diabetes is a major health concern. It is one of the world's worst illnesses. Diabetes can also lead to other illnesses, including heart attacks, renal problems, blindness, etc. Diabetic mellitus is classified as a non-communicable disease (NCD). According to the 2017 statistics, around 426 million people worldwide suffered from diabetes. Approximately 2.5–3.5 million people died due to diabetes. Doctors and researchers claimed that by 2045, this number would rise to 630 million. Four out of every ten people have diabetes, and two in eight have prediabetes. Early diagnosis and treatment can help manage diabetes and stop complications, which can ultimately save lives. Diabetes can be classified into the following types:

- **Type-1 diabetes:** Type-1 diabetes is caused by a deficiency of insulin, as the pancreas is unable to produce sufficient insulin due to the destruction of beta cells by the immune system. This results in elevated blood sugar levels, as the glucose cannot be utilised by the cells without adequate insulin. While the actual reason is unknown, genetic factors are considered to play a key impact. Type-1 diabetes typically affects children and young adults, with 5% to 10% of people being affected. If left untreated then it can lead to severe issues. Management of type-1 diabetes requires regular insulin injections, which vary in type and dosage depending on individual requirements, to control blood sugar levels.

- **Type-2 diabetes:** Type-2 diabetes is a prevalent condition where the body produces enough insulin, but cells do not use it efficiently, resulting in increased blood glucose levels. Our pancreas may produce less insulin over time. This form of diabetes accounts for around 90% of all cases and is commonly observed in both young and elderly individuals. Type-2 diabetes poses a high risk of developing several medical complications, such as diabetic retinopathy (which damages blood vessels in the eyes), diabetic neuropathy, and kidney and cardiovascular diseases. This is an ongoing disease with no known cure yet. Patients are advised to make certain lifestyle modifications to maintain normal blood glucose levels.

* Ukani Vedant. Tel.: +91 9328400365

E-mail address: patelvedant321@gmail.com

- **Gestational diabetes:** Other than in women who already have diabetes, gestational diabetes is a type of diabetes that arises in pregnancy. This type of diabetes is similar to type 2 diabetes and is more likely to develop in obese women. While gestational diabetes is reversible and typically resolves after childbirth, it can lead to perinatal complications and pose health risks for both the mother and newborn.

1.2. Machine Learning in Diabetes Prediction

The use of machine learning applications in healthcare is beneficial for clinicians and patients as it enables better decision making in less time and at lower costs. These applications aim to provide high-quality services that are affordable for patients. While healthcare systems are complex, emerging techniques have made evaluation easier than before. Various algorithms are being combined to improve the accuracy of experiments. Early predictive analysis is essential for determining the severity of a disease. Machine learning algorithms are used as a hybrid model approach to compare and analyse patterns and accuracy using various tools for decision making in the medical field. Three main types of machine learning algorithms are:

- **Supervised learning**: In supervised learning, the system is trained on a set of input-output pairs, known as training data, to learn a function that can accurately predict the output value for new, unseen inputs. The goal of supervised learning is to find a target function that minimises the error between the predicted output and the actual output. The input variables, or independent variables, are also known as features or attributes, and the output variable, or dependent variable, is the target variable. The domain of the target function is the set of all possible input values or instances. Classification and regression are the two main kinds of algorithms used in supervised learning. The target variable in classification is categorical data, and the purpose is to predict the class or category of a new input instance. In regression, the target variable is a continuous variable, and the goal is to predict a numerical value for a new input instance.

- **Unsupervised learning**: Unsupervised learning is a technique that is used to identify hidden patterns and structures within a dataset whose input is known but whose output is unknown. It is commonly used on transactional data, where clustering algorithms like k-means and hierarchical clustering are used to identify similar groups of data. However, noisy and missing data can make the extraction of meaningful information from a dataset difficult. Data mining plays an important role in unsupervised learning because it helps identify patterns and relationships among variables in large sets of data. There are two different types of methods in unsupervised learning: the method that is used to group the objects into clusters with the most similarities is known as clustering, and the method that is used to find relationships between variables is known as an association.

- **Reinforcement Learning**: Reinforcement learning is a broad term that refers to a class of machine learning techniques in which a model learns by interacting with its environment to optimise a reward function. Unlike other types of machine learning, this system has no advanced knowledge of the environment's behaviour and must learn through trial and error. Throughout this trial-and-error phase, the system takes action and receives feedback in the form of a reward signal. Reinforcement learning is especially beneficial for creating autonomous systems that can learn from and adapt to new surroundings without explicit programming or oversight. Because of its independence from the environment, reinforcement learning is appealing for a wide range of applications, including decision-making, gameplay, and robotics.

2. Literature Review

In this related work section, we are briefly going to discuss the earlier work done by researchers in this field of diabetes mellitus prediction using machine learning. One such study by N. Fazakis et al. [1] proposed an IoT-enabled framework for workers which monitors users' health, well-being, and functional ability in an unobtrusive manner. The framework was empowered with AI tools and applied various machine learning algorithms, including ensemble learning, to predict the occurrence of Type 2 Diabetes Mellitus (T2DM). The author in this work used the English Longitudinal Study of Ageing (ELSA) database, and different features of models trained on this dataset are compared; mainly the AUC curve was used to compare the performance of different algorithms.In the realm of machine learning, KM Jyoti Rani [2] conducted an analysis of a dataset obtained from the UCI machine learning repositories. The author aimed to improve the accuracy of the dataset by utilising five different techniques, including Support Vector Machine (SVM), Random Forest, Decision Tree, K Nearest Neighbor (KNN), and Logistic Regression. After conducting the analysis, the proposed method found that the Decision Tree classifier outperformed the other techniques, achieving a 98% accuracy rate on the training dataset. Deepti Sisodia et al. [3] evaluated three different classifiers on the PIDD (Pima Indian Diabetes Dataset): SVM, Naive Bayes, and Decision Tree. The author analysed the techniques in this work using multiple measures, and the results were validated using the Receiver Characteristic (ROC) curve. The obtained results show that Naive Bayes beats all other algorithms in terms of accuracy, with 76.30% accuracy. In [5], Mujumdar, Aishwarya, and Vaidehi V. use a pipeline approach to try to enhance the classifier's performance utilising a new dataset instead of an old dataset in [5]. The model was separated into five parts, including Dataset Collecting, Data Pre-processing, Clustering, Build Model, and Evaluation. AdaBoost classifier was found to be the best model using the pipeline, with an accuracy of 98.8%. Mitushi Soni et al. [4] presented their work on early diabetes mellitus (DM) prediction utilising six different classifiers. The author applied feature subset selection approaches, which aid in the elimination of irrelevant features. The feature importance plot indicates that skin thickness is less significant in the case of a random forest classifier, and the results demonstrate that random forest outperforms others with 77% accuracy in comparison.

3. Proposed Methodology

This section discusses the techniques and methods used to build the proposed diabetes prediction system. In this section, we will discover the several classifiers that machine learning uses to predict diabetes. We will also discuss our proposed techniques for improving accuracy. The methodology begins by collecting the dataset required for the study. The dataset used in this research was obtained from the UCI machine learning repository. Once the dataset is obtained, it is merged to form a comprehensive and unified dataset that incorporates all the necessary variables and information. This merged dataset serves as the foundation for further analysis and model development. Next, data preprocessing techniques are applied to the merged dataset to handle missing values and outliers and ensure data quality. This includes checking for missing values and transforming instances of diabetes into numerical values (e.g., 1 or 0). Feature selection is performed to identify the most relevant features for diabetes prediction, reducing the dimensionality of the dataset. After preprocessing, the dataset is split into two subsets: the training data and the test data. The training data, which constitutes 80% of the dataset, is used to train various machine learning algorithms, including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Logistic Regression (LR). Each algorithm is built and trained based on the training set. Subsequently, the trained classifier models are tested using the remaining 20% of the data, referred to as the test set, to evaluate their predictive performance. The performance evaluation includes measuring accuracy, precision, recall, and F1-score, among other evaluation measures. Through comparison and evaluation of the experimental results obtained for each classifier, the best-performing algorithm for diabetes prediction is determined. The overall process, including data collection, dataset merging, data preprocessing, data splitting, algorithm selection, model building, and model evaluation, is summarised in Figure 1 below. This methodology provides a systematic approach to developing and evaluating predictive models for diabetes using machine learning algorithms.

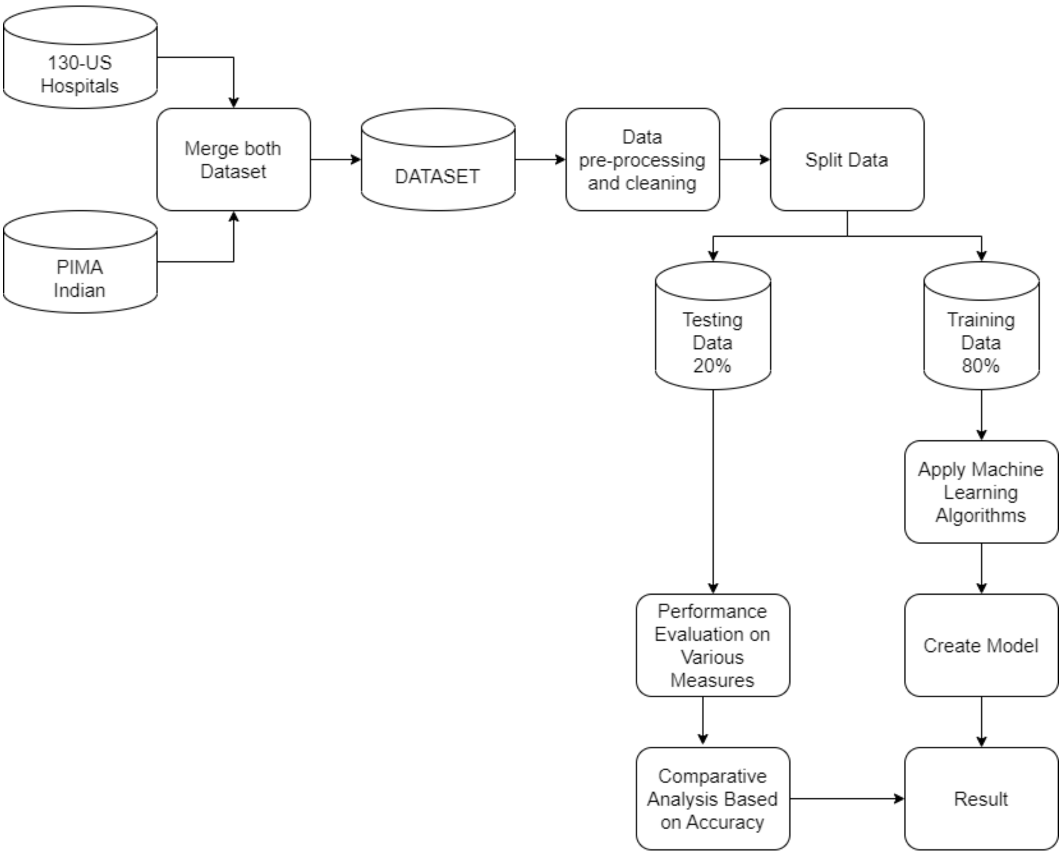

Fig. 1. Working Sequences of a Machine Learning-based Diabetes Prediction Model

3.1 Dataset Description

This study has made use of the publicly available Pima Indian Diabetes Dataset, which includes records on 768 female patients aged 21 and older as well as a parameter that shows the presence or absence of diabetes. The dataset, however, is slightly biassed, with around 500 cases labelled negative (0) and 268 instances labelled positive (1) for diabetes. The attributes of the dataset are illustrated in table 1. In addition to this dataset, the articles provided a dataset of 130 US hospitals that has been prepared to analyse outcomes pertaining to patients with diabetes. Both datasets were pooled to train a diabetes prediction model, yielding a total of 6768 characteristics.

Table 1. Dataset Attribute Description

No.	Attribute	Description
1	Pregnancies	No. of times pregnant
2	Glucose (mg/dL)	Plasma glucose concentration
3	BloodPressure ($mm\ Hg$)	Diastolic blood pressure
4	SkinThickness (mm)	Skin fold thickness
5	Insulin ($\mu U/ml$)	2 hours serum insulin
6	BMI (kg/m^2)	Body mass index
7	DiabetesPedigreeFunction	Diabetes pedigree function
8	Age ($years$)	Age in years
9	Outcome	class ‘0’ - non-diabetic, class ‘1’- Diabetic

Table 2. Dataset Description

Dataset	No. of Instances	No. of Attributes
PIMA	768	9
130-US Hospital	6000	9
Total	6768	9

3.2 Pre-processing

In the situation of a medical dataset, where missing values are common, data preparation is crucial. The accuracy of predictions made from the dataset may be compromised due to missing values or inconsistent data. Data preprocessing techniques are used to overcome this issue. Because some variables, such as BMI and skin thickness, cannot have zero values, the mean value of their corresponding attributes is derived from the dataset and used to replace missing values. In the case of merged diabetes datasets, two preprocessing steps are carried out. The first step involves removing all instances with a zero value, as such values are not possible and are likely due to data entry errors. The second step is data splitting, where the cleaned data is normalised and separated into training and testing sets. This is done using the holdout validation technique, where 80% of the data is used for training and 20% for testing. The training process generates a model based on logic and algorithms using the feature values in the training data, while the test dataset is used to validate the accuracy of the model. By following these preprocessing steps, we can ensure that the machine learning techniques applied to the dataset produce accurate and reliable results.

3.3 Classification Algorithms and Techniques

3.3.1 K Nearest Neighbors

K-nearest neighbor (KNN) is a versatile and widely used machine learning algorithm renowned for its simplicity and effectiveness in classification and regression tasks. With its intuitive approach, KNN finds its strength in identifying the closest neighbours to a given data point and utilising their characteristics to make predictions. By employing distance metrics such as Euclidean or Manhattan distance, KNN measures the proximity between the input data point and every other data point in the training set. The k nearest neighbors, determined by their distances, are then chosen to contribute to the prediction process. For classification tasks, KNN assigns the majority class among the nearest neighbours as the predicted label for the input data point. In regression tasks, the predicted value is often determined by calculating the mean or median of the target values of the nearest neighbours. The choice of k, the distance metric, and appropriate data normalisation techniques greatly influence the algorithm's results.

3.3.2 Decision Tree

Decision trees are widely employed supervised machine learning algorithms capable of handling classification and regression tasks. They operate by constructing a hierarchical structure composed of nodes that make decisions based on features or attributes. At each node, the algorithm carefully selects the feature that yields the highest information gain, allowing for the division of data into distinct classes. This process continues recursively, as the decision tree grows by choosing the most suitable attribute to separate the data at each subsequent node. The ultimate goal is to accurately classify all the data points by correctly assigning them to their respective classes.

3.3.3 Support Vector Machine

A popular supervised machine learning approach for classification and regression analysis is the Support Vector Machine (SVM). Its primary goal is to group the points in the provided dataset into clearly defined groups. Support vectors are well selected hyperplanes that maximise the margin or distance between the closest data points of various classifications. SVM seeks to improve generalisation and increase its capacity to correctly categorise unobserved cases in both the training and test sets by optimising this margin. The chosen hyperplane, often referred to as the Optimal Separating Hyperplane, reduces the chance of misclassifying training instances while also ensuring robustness when processing brand-new, untested data points.

3.3.4 Logistic Regression

The supervised learning classification process known as logistic regression is frequently used to calculate the likelihood of a binary response based on one or more predictors. These predictors may be continuous or discrete in type. Its main goal is to categorise or separate data objects into separate categories. When dealing with binary classification jobs, like evaluating whether a patient is positive or negative for a certain medical condition like diabetes, this technique is quite helpful. Finding the best-fitting model that explains the relationship between the target variable and the predictor variables is the goal of logistic regression. It uses the sigmoid function, which is based on the theory of linear regression. It always gives the probability for a particular data set between 0 and 1, and, according to the probability, we put the instance in a particular class.

3.4 Evaluation

It is important to provide various measures to discuss the pros and cons of classification models and evaluate their performance. A variety of measures are used to assess the performance of machine learning models. Accuracy, confusion matrix, precision, recall, F1 score, and specificity are among these measurements. The confusion matrix, which is clearly depicted below in Table 3, makes it simple to determine how the built predictive model performs. It outputs a matrix that summarises the overall performance of the model. The confusion matrix is a helpful tool for evaluating classification model performance, providing a summary of predicted and observed classes. Based on whether the expected and actual values match, this matrix divides predictions into four individual instances: True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP). True positive, true negative, false positive, and false negative are used to determine other metrics such as sensitivity, specificity, accuracy, precision, and F1 score.

Table 3. Confusion matrix

Actual class	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Accuracy: Accuracy represents the percentage of right predictions out of the total number of predictions. Mathematically, it is given as

$$Accuracy = (TP + TN)/(TP + FN + FP + TN) \tag{1}$$

Specificity: The proportion of participants who are correctly classified as negative is referred to as specificity. This can be expressed in terms of a mathematical equation as

$$Specificity = TN/(TN + FP) \tag{2}$$

Recall/Sensitivity: The proportion of participants who are correctly classified as positive is referred to as sensitivity. It is expressed as

$$Recall = TP/(TP + FN) \tag{3}$$

Precision: Precision is calculated by dividing the number of correct positive outcomes by the number of positive results predicted by the classifier. This can be defined mathematically as

$$Precision = TP/(TP + FP) \tag{4}$$

F1-Score: The F1 score is the harmonic mean of precision and recall, and it measures the test's correctness. It is presented in mathematical terms like

$$F1\ Score = 2 * (precision * recall)/(precision + recall) \tag{5}$$

4. Experimental Results

As can be seen from the below plot, the testing data showed that KNN had an accuracy of 82.67%, the decision tree had an accuracy of 76.17%, logistic regression had an accuracy of 78.88%, and SVM had an accuracy of 76.35%.

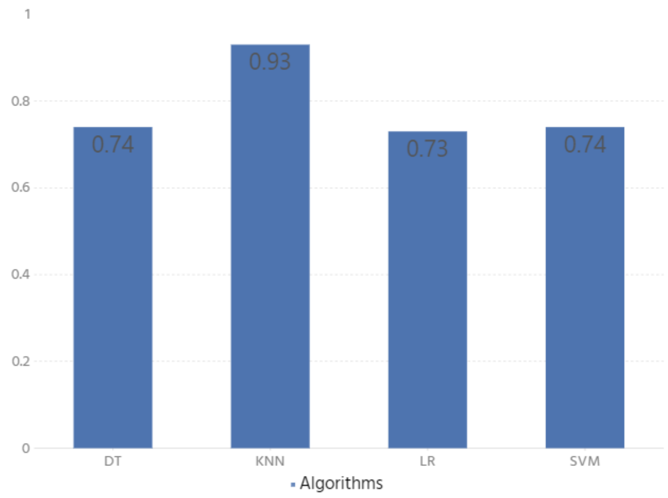

Fig. 2. Machine Learning Algorithm Accuracy Comparison Bar Graph

Based on these findings, it is possible to infer that KNN outperformed all other algorithms in terms of accuracy, while the decision tree had the lowest accuracy. For these algorithms, quality metrics such as accuracy, recall, precision, F1 Score, and specificity were computed and shown in Table 4. Figures 3–6 show the confusion matrix of prediction using support vector machines, k-nearest neighbors, decision tree, and logistic regression, with LR displaying the largest number of true positives and KNN displaying the highest number of true negatives. Additionally, KNN had the fewest false negatives, whereas LR had the fewest false positives. The introduction of a new dataset has improved the accuracy of KNN but has not impacted decision tree, SVM, or logistic regression. KNN achieved an accuracy rate of 93.04% on the testing dataset when K = 11 and input data were used. These findings provide valuable insights into the performance of different classification algorithms and can assist in selecting the most suitable algorithm for a specific application. It can be concluded that KNN is superior to decision tree, SVM, and logistic regression in predicting whether a person is diabetic or not, with "1" representing positive results and "0" representing negative results.

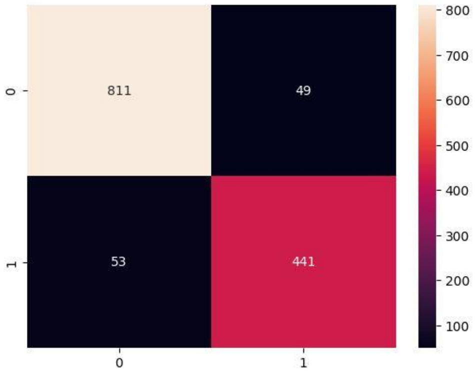

Fig. 3. KNN Confusion matrix

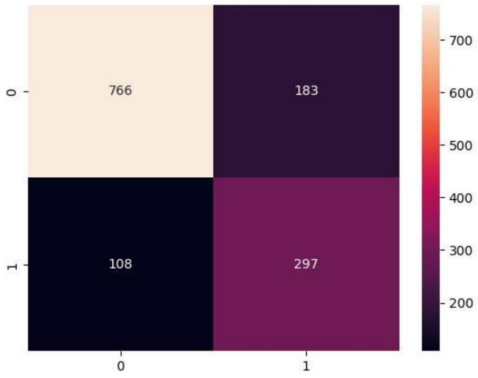

Fig. 4. DT Confusion matrix

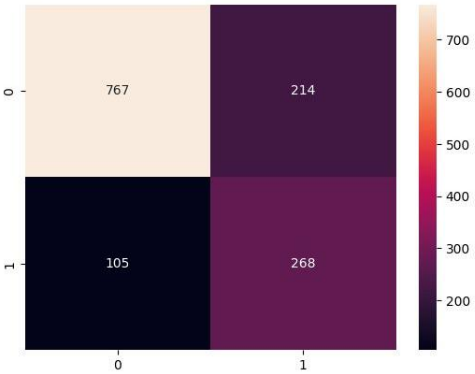

Fig. 5. LR Confusion matrix

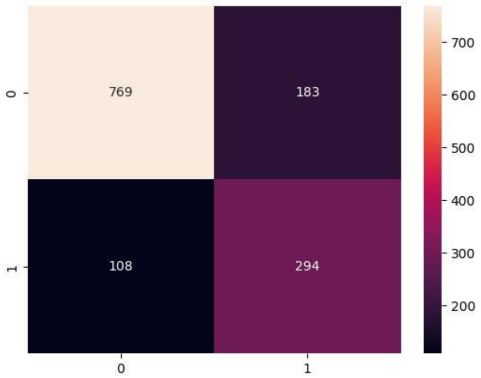

Fig. 6. SVM Confusion matrix

Table 4. Performance measure of classifier in the merged dataset (RTML and PIMA Indian)

Classifier	Recall	Precision	Specificity	Accuracy %	F1 Score
Decision Tree	0.87	0.80	0.61	74.86	0.84
K-Nearest Neighbor	0.93	0.94	0.90	93.04	0.94
Logistic Regression	0.87	0.78	0.55	73.67	0.82
Support Vector Machine	0.87	0.80	0.61	74.79	0.84

5. Conclusion

One of the most important real-world medical problems is the detection of diabetes at an early stage. Experimental results can be used to improve health care intake through early predictions and making early decisions to cure diabetes and save human lives. In this paper, our main aim was to design and implement diabetes prediction using machine learning techniques and do a comparative performance analysis of the models trained using those techniques, which has been achieved successfully. The proposed approach uses four machine learning classification techniques: SVM, KNN, Decision Tree, and Logistic Regression classifiers. The effectiveness of these classifiers is examined by conducting experiments on an open-source Pima Indian dataset and a 130-US hospitals dataset. The performance metrics, namely specificity, sensitivity, accuracy, recall, and f-measure, are computed for the given machine learning and ensemble techniques. We found that the KNN classifier has the highest accuracy through the confusion matrix evaluation test, with 93.04% classification accuracy and an F1 score and precision of 0.94 and 0.94, respectively. It is observed that the model's accuracy improves with this dataset compared to the existing one. In the future, it is planned to improve the accuracy of the prediction by testing our classification techniques on the huge dataset, and its performance can be improved. Some future scopes of this work include deploying the best performing KNN classifier into a website and smartphone application to predict diabetes instantly. We can achieve better results by collecting more data from a larger group of patients.

References

1. N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis and K. Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction," in IEEE Access, vol. 9, pp. 103737-103757, 2021, doi: 10.1109/ACCESS.2021.3098691.
2. KM Jyoti Rani, "Diabetes Prediction Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456- 3307, Volume 6 Issue 4, pp. 294-305, July-August 2020.
3. Deepti Sisodia, Dilip Singh Sisodia, Prediction of Diabetes using Classification Algorithms, Procedia Computer Science, Volume 132, 2018, Pages 1578-1585, ISSN 1877-0509.
4. Mitushi Soni, Dr. Sunita Varma, 2020, Diabetes Prediction using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 09 (September 2020).
5. Mujumdar, Aishwarya & Vaidehi, V. (2019). Diabetes Prediction using Machine Learning Algorithms. Procedia Computer Science. 165. 292-299. 10.1016/j.procs.2020.01.047.
6. Hassan, Abdulhakim & Malaserene, I & Leema, Anny. (2020). Diabetes Mellitus Prediction using Classification Techniques. International Journal of Innovative Technology and Exploring Engineering. 9. 2278-3075. 10.35940/ijitee. E2692.039520.
7. Sarker, Iqbal & Faruque, Md & Alqahtani, Hamed & Kalim, Asra. (2019). K-Nearest Neighbor Learning based Diabetes Mellitus Prediction and Analysis for eHealth Services. 162737. 10.4108/eai.13-7-2018.162737.
8. Zou, Quan & Qu, Kaiyang & Luo, Yamei & Yin, Dehui & Ju, Ying & Tang, Hua. (2018). Predicting Diabetes Mellitus with Machine Learning Techniques. Frontiers in Genetics. 9. 10.3389/fgene.2018.00515.
9. Tasin, Isfafuzzaman & Nabil, Tansin & Islam, Sanjida & Khan, Riasat. (2022). Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters. 10.1049/htl2.12039.
10. A. Prabha, J. Yadav, A. Rani and V. Singh, "Non-invasive Diabetes Mellitus Detection System using Machine Learning Techniques," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 948-953, doi: 10.1109/Confluence51648.2021.9377138.
11. Birjais, R., Mourya, A.K., Chauhan, R. *et al.* Prediction and diagnosis of future diabetes risk: a machine learning approach. *SN Appl. Sci.* 1, 1112 (2019).
12. Dritsas E, Trigka M. Data-Driven Machine-Learning Methods for Diabetes Risk Prediction. Sensors (Basel). 2022 Jul 15;22(14):5304. doi: 10.3390/s22145304. PMID: 35890983; PMCID: PMC9318204.
13. Xue, Jingyu & Min, Fanchao & Ma, Fengying. (2020). Research on Diabetes Prediction Method Based on Machine Learning. Journal of Physics: Conference Series. 1684. 012062. 10.1088/1742-6596/1684/1/012062.
14. M. Rout and A. Kaur, "Prediction of Diabetes Risk based on Machine Learning Techniques," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 246-251, doi: 10.1109/ICIEM48762.2020.9160276.
15. Sarwar, Muhammad Azeem & Kamal, Nasir & Hamid, Wajeeha & Shah, Munam. (2018). Prediction of Diabetes Using Machine Learning Algorithms in Healthcare. 1-6. 10.23919/IConAC.2018.8748992.
16. Shafi, Salliah & Selvam, Venkatesan & Ansari, Gufran & Ansari, Mohd Dilshad & Rahman, Md Habibur. (2022). Prevalence and Early Prediction of Diabetes Using Machine Learning in North Kashmir: A Case Study of District Bandipora. Computational Intelligence and Neuroscience. 2022. 1-12. 10.1155/2022/2789760.
17. Beghriche, Tawfiq & Mohamed, Djerioui & Youcef, Brik & Bilal, Attallah & Brahim Belhaouari, Samir. (2021). An Efficient Prediction System for Diabetes Disease Based on Deep Neural Network. Complexity. 2021. 1-14. 10.1155/2021/6053824.
18. Choudhury, Ambika & Gupta, Deepak. (2019). A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques: IC3 2018. 10.1007/978-981-13-1280-9_6.
19. Joshi, Tejas & Pramila, M & Chawan, Pramila. (2018). Diabetes Prediction Using Machine Learning Techniques. 2248-9622. 10.9790/9622-0801020913.