International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Diabetes Prediction Using Medical Data and Disease Influence Measures using Machine Learning

[1]Dr. Avinash J. Agrawal, [2]Dr. Rashmi R. Welekar, [3]Dr. Namita Parati, [4]Dr. Pravin R. Satav, [5]Dr. Leena H. Patil, [6]Dr. Shailendra S. Aote

Submitted: 15/05/2023 **Revised:** 01/07/2023 **Accepted:** 20/07/2023

Abstract: This project intends to create a diabetes predictive model utilizing medical data and investigate the impact of various factors on the condition using machine learning techniques. Millions of individuals throughout the world suffer from the common chronic illness known as diabetes. The results of patients and public health initiatives can be considerably improved by early detection and an understanding of the underlying causes.To do this, a large dataset of medical records from people with and without diabetes that included a variety of demographic, lifestyle, and clinical factors was gathered. To pre-process the data and identify useful features, feature engineering approaches were used. Accurate prediction models for diabetes risk assessment were created using a variety of machine learning methods, such as Decision Trees, Random Forests, and Support Vector Machines.The main causes of the development of diabetes were also determined by looking into disease influence measures. This study intends to clarify the relative importance of several risk factors, such as age, BMI, family history, and glucose levels, by examining feature importance and correlation coefficients.In this paper various disease prediction methods were assessed and contrasted depending on how well they predicted diseases. The analysis' findings have been given in great detail to aid in the development process.

Keywords: *Support Vector Machine, Decision Trees, Random Forests, and machine learning for diabetes prediction*

I. Introduction

Numerous variables, including as alterations in dietary habits, lifestyle changes, and sedentary activity, have contributed to its gradually rising incidence. Effective diabetes care and the avoidance of the disease's consequences depend on early detection and precise diabetes prediction [1].Development of predictive models based on medical data is now possible thanks to machine learning, which has shown to be a potent tool in the healthcare industry. Machine learning algorithms can find patterns and relationships in massive datasets of patient medical information that can help with disease prediction.

The emergence of Diabetes Mellitus, which can affect people at any age, is one of the effects of the changing lifestyles of humans on their health. Once this syndrome manifests, it remains a lifelong condition marked by inadequate pancreatic insulin secretion, which causes increased blood sugar levels [2]. Diabetes Mellitus, which affects people of all ages, is regarded as a global health issue. In 2000, diabetes contributed to a mortality rate of 6%; this percentage is projected to increase to 45% in the following 20 years. Numerous variables, including altered eating patterns, sedentary lifestyles, smoking, and consuming lots of high-protein and junk food, among others, might contribute to the development of diabetes [23].

---

[1]Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India
[2]Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India
[3]Maturi Venkata Subba Rao (MVSR) Engineering College, Hyderabad, Telangana State, India.
[4]Government Polytechnic, Murtijapur, Maharashtra, India
[5]Priyadarshini College of Engineering, Nagpur, Maharashtra, India
[6]Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India
[1]agrawalaj@rknec.edu, [2]welekarr@rknec.edu, [3]namianand006in@gmail.com, [4]prsatav@gmail.com, [5]lhpatil10@gmail.com, [6]shailendra.aote@gmail.com

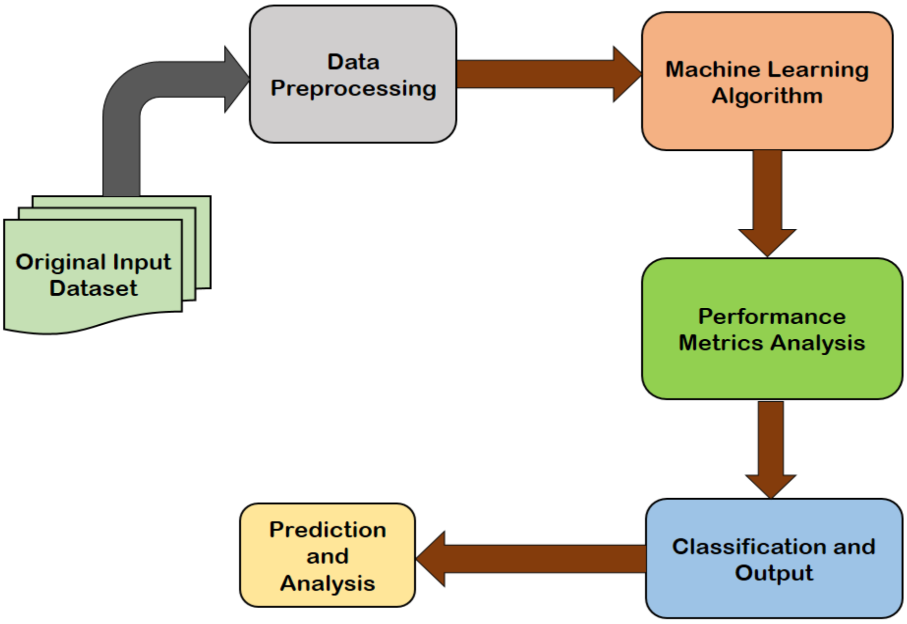

Fig. 1: Proposed model for Prediction and analysis

Monitoring blood sugar levels and taking into account other elements like age, physical activity, and lifestyle choices are both essential for identifying the disease. One-third of diabetics are thought to be affected by microvascular and macrovascular problems that cause harm to their organs and tissues [3]. For the purpose of preventing vascular problems, medical professionals must recognize pre-diabetic individuals and research their glucose tolerance and insulin resistance.

Diabetes patients often experience neuropathy and nephropathy, two serious consequences that can cause peripheral nerve damage and cardiovascular failure. Neuropathy affects more than 50% of diabetics. Renal organs are affected and may fail as a result of diabetic nephropathy, which is characterized by elevated albumin levels in the urine. Since [4] the precision of the prediction significantly depends on the quantity of training data taken into account, effective predictions of neuropathy and nephropathy require considerable data gathering from several diabetic individuals.Historical information on a variety of diabetes patients is necessary for illness prediction and is gathered from many medical organizations into a diabetic dataset. The system is then trained using this dataset such that it can forecast results when given a set of input values [5]. There are several methods for predicting diabetes, one of which uses data mining a method for pulling pertinent information from huge databases. The diabetic prediction problem can be efficiently addressed by adapting data mining approaches. There are also many additional scientific methods available for this prediction task [6].

This work offers a diabetes prediction method based on the Disease Influence Measure (DIM). The suggested approach makes use of DIM to evaluate how various parameters affect the process of disease prediction[7]. By adding DIM, the algorithm improves the precision and comprehension of the prediction outcomes, offering insightful knowledge about the significant elements influencing diabetes. The DIM-based strategy is a viable way to enhance disease prediction and can considerably advance the field of managing diabetic healthcare.With the help of machine learning techniques, we hope to create a diabetes prediction model and investigate the various influences that may contribute to the development of the disease. Various demographic, lifestyle, and clinical factors collected from diabetic patients and non-diabetics will make up the dataset utilized for training and testing the model [8]. Additionally, in order to evaluate the effect of various factors on the prediction process, we will create a disease influence measure (DIM). DIM will offer insightful information about the relative importance of many risk factors influencing the development of diabetes [9].

II. Review Of Literature

Chronic metabolic disease known as diabetes mellitus (DM) is characterized by inadequate insulin production [1]. According to the CDC statistics for 2021, 11.3% of the US population has DM, making it the ailment that affects around 1 in 10 people worldwide. According to the ADA, the cost of diagnosed diabetes in 2017 was anticipated to be $327 billion.Particularly when blood glucose levels are out of control, DM causes a variety of issues in important organs like the heart, brain, kidney, and eyes. Random blood glucose values exceeding 180 (mg/dL) or HbA1c levels higher than 7.0% are used to indicate uncontrolled diabetes mellitus (UDM). Micro- and macrovascular problems, which afflict over 50% of persons with diabetes in the US, are an indication of UDM [2].

Both pharmaceutical and non-pharmacological therapies have been used to lessen the effects of UDM, and new pharmaceutical agents and insulin delivery systems have been developed to enhance glycaemic control. Nevertheless, despite these developments, there is still a substantial variance in glucose control among individuals with varied characteristics, which raises the risk of complications from diabetes.

To lessen the severity of uncontrolled diabetic mellitus (UDM), a variety of pharmaceutical and non-pharmacological therapies have been used [7]. Glycemic control has been greatly enhanced by the development of new pharmacological agents and insulin delivery systems [10]. Nevertheless, despite these developments, there is still a significant variation in glycemic control among patients with varied characteristics, which increases the risk of complications from diabetes [13].In addition to clinical care, early glycemic status prediction of diabetic patients has the potential to lessen UDM's burden [14]. Fasting plasma glucose, haemoglobin A1c, and oral glucose tolerance tests are common glycemic status measurements that have been used most frequently in previous attempts to predict UDM. Due to the overnight fasting requirement, these routine tests might not always be practical. Additionally, more extensive biological marker data have not been properly included into earlier prediction techniques. Although they are known to interact with glycemic status, biological indicators such serum electrolytes and haematological indices play a crucial role in the prediction of UDM utilizing machine learning (ML) algorithms. It may be possible to increase the precision and efficacy of UDM prediction by including these biological markers into ML-based prediction models, providing a fresh and all-encompassing method for managing diabetes.Additionally, physical characteristics like weight and height can reveal obesity, which frequently coexists with diabetes [22]. Erythrocyte counts and other blood indicators can also affect haemoglobin levels, which in turn affects HbA1C levels [23]. If a suitable prediction model is created using information from a representative sample, the correlation between these characteristics and glycemic state offers an alternate means of monitoring UDM. We used the All of Us (AoU) research program, which offers a sizable and racially varied sample of the US population, for this study [24].

In comparison to traditional glucose tests, the prediction of UDM utilizing novel features offers a cost-effective approach for glucose monitoring. By identifying those who are at risk, using a predictive model with a wider range of patient characteristics can help decrease complications caused by diabetes and improve the quality of life for diabetes patients. In this situation, our work used a supervised machine learning approach and several patient features to effectively predict UDM. It emphasized the value of patient features and physiological indicators in predicting UDM in the absence of routine glycemic status monitoring.

III. Publically Available Datasets

The Pima Indian Diabetes Dataset (UCI Machine Learning Repository, 1998) is among the best datasets to evaluate machine learning algorithms for diabetes prediction. Based on diagnostic indicators like pregnancy, blood sugar, blood pressure, skin its thickness, diabetic pedigree function, insulin, body mass index, and age, the Pima Indian dataset, which was made public by the National Institute of Diabetes and Digestive and Kidney Diseases in 2013, can be used to determine whether a patient has diabetes.

Table 1: Description of Dataset

Features	**Details**
Pregnancies	Number of pregnancies that have occurred.
Glucose	After two hours, the plasma glucose concentration (a test of glucose tolerance).
Blood Pressure	Diastolic blood pressure is measured in mm Hg (heartbeats per minute).
Skin Thickness	(mm) The thickness of the triceps' skin folds.
Insulin	Serum insulin concentration (mu U/ml) after two hours.
BMI	Index of body mass.
Diabetes Pedigree Function	Diabetes Family History.
Age	Age expressed in years.
Outcome	As a result, there is a class variable (0: diabetic Negative, 1: Diabetic Positive).

IV. Proposed System

Pre-processing two different datasets is required for the proposed methodology for diabetes prediction. In this phase, we examine attribute correlation to find pertinent features for diabetes identification. Following that, the data is divided into training and testing sets. Using a variety of methods, the training data is utilized to create predictive machine learning models. We assess the model's effectiveness using various measures. In the end, Flask is used to deploy the top-performing ML model in a web application.

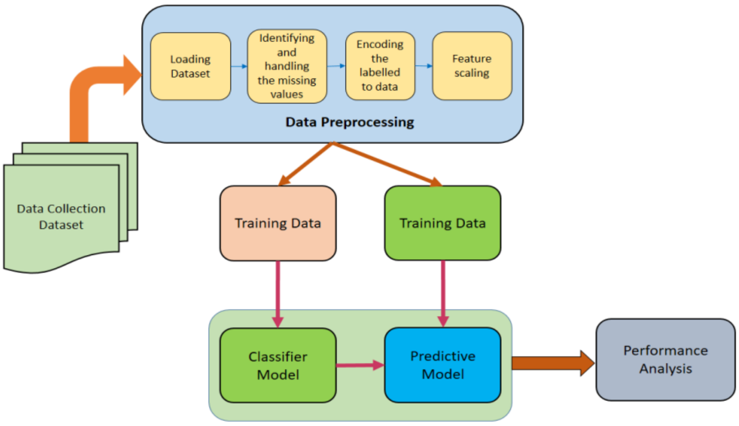

Fig. 2: proposed model and flow of proposed work

1. Data Collection:

To make sure that our model was robust, we gathered two distinct datasets, each with a different amount of features or components. These datasets were assembled from a range of sources, including data on diabetes, global health characteristics, and details from several health institutes.

2. Data Preprocessing:

Preparing raw data for analysis and the creation of prediction models requires a critical step called data preparation. It entails preparing the data for machine learning algorithms by cleaning, converting, and organizing it. Enhancing data quality, handling missing values, eliminating inconsistencies, and creating features that more accurately depict the underlying patterns in the data are the main goals of data preprocessing.

2.1 Data Cleaning: Data cleaning entails dealing with outliers, duplicate records, and missing values. Depending on the circumstances, missing data can be eliminated or imputed. To prevent bias in the analysis, duplicate records are often eliminated.

2.2 Data Transformation: Data transformation is sometimes necessary to make sure that the data adheres to a particular distribution. To scale the data inside a certain range, common transformations include logarithmic, square root, or normalization.

2.3 Feature Selection: Selection of Features: The effectiveness and performance of the model can be enhanced by choosing the features that are the most pertinent. To lessen complexity and overfitting, duplicate or irrelevant features may be deleted.

2.4 Feature Engineering: By developing new features from already existing ones, we can gain new knowledge and improve the model's capacity to recognize patterns.

2.5 Data Splitting: Usually, the dataset is used to construct training and testing sets. The testing set is used to evaluate the model's performance on test data after it has been trained using the training set.

2.6 Handling Unbalanced Data: If the classes are unbalanced, the dataset can be balanced using methods like oversampling, undersampling, or artificial data generation.

2.7 Standardization/Normalization: By adjusting the features' scales to a similar range, one feature can be kept from outweighing the others while learning.

3. Machine Learning Algorithm:

3.1 Naive Bays Algorithm:

Based on the Bayes theorem, the Naive Bayes automated learning method is used to address a number of categorization issues. In this piece, we'll go into greater detail on the Naive Bayes method to clear up any misunderstandings.

Recompile and Prepare Data: Gather an input vector- and class-labeled training dataset.

Calculate each class's preliminary probability, P(Yi), using the occurrences from the formation table. For calculating a class's preliminary probability, use the following formula:

$$\mathrm{P(Yi\ =\ yt)} = \frac{\mathrm{count(Yi\ =\ yt)}}{N}$$

Where count(Yi = yt) is the number of instances of the class yt in the series of formation instances, and N is the total number of formation instances.

Calculate the residual probabilities: Calculate the posterior probabilities of the entry vector P(Y|X) for each specific class using the Bayes theorem:

$$\mathrm{P(Yi|Xi)} = \frac{\left(\mathrm{P(Xi|Yi)} * \mathrm{P(Yi)}\right)}{\mathrm{P(Xii)}}$$

The class with the highest likelihood of returning must be chosen in order to predict the name of the class for an impending and unknown vector of entry.

3.2 Support Vector Machine:

It is renowned for its capacity to effectively handle big datasets and is particularly excellent in addressing both linear and nonlinear classification issues.To handle numerous concerns like routing, localization, fault detection, congestion control, and communication problems, SVM has also been employed in a variety of fields, including Wireless Sensor Networks (WSNs).

A hyperplane with the greatest margin of separation between the two classes must be found using SVM. You can represent this hyperplane by:

$$\mathrm{W} * \mathrm{X} + \mathrm{B} = 0$$

SVM's decision meaning is described as follows:

$$\mathrm{F(x)} = \mathrm{sign(W} * \mathrm{X} + \mathrm{B\,)}$$

SVM minimises the classification error while maximising the margin between the classes. As a result, the optimisation problem is formulated as follows:

$$\text{minimize: } \tfrac{1}{2}\|\mathrm{w}\|^2 + \mathrm{C}\Sigma_\mathrm{i}\, \xi_\mathrm{i},$$

$$\text{subject to: } \mathrm{y_i}(\mathrm{w} \cdot \mathrm{x_i} + \mathrm{b}) \geq 1 - \xi_\mathrm{i},$$

3.3 Decision Tree:

The decision tree, also known as the classification tree and the regression tree, is a method of guided learning that predicts categorised variables for continuously categorised entry and exit variables. Because of its visual representation, human interpretation is straightforward and supports decision-making.

A splitting rule, also known as feature collection measure, is a heuristic used to choose the optimum criterion for data partitioning in order to produce the most efficient data separation. It aids in identifying the tuple breakpoints at a specific node. Each feature (or attribute) is given a rank or score by the attribute selecting measure based on how well it can describe the provided dataset. The splitting attribute is determined by the attribute with the greatest score. Split points are also chosen for qualities with continuous values in order to define each branch.

1. S is the algorithm's focal point.
2. The algorithm analyses each non-used group attribute, S, and calculates the associated entropy (H) and information gain (IG).

$$Entropy\ E(S) = \sum_{j=1}^{c} - Pi\ log_2 Pi$$

And Information Gain Calculate as:

$$Information\ Gain\ (IG) = Entrophy\ (before) - \sum_{j=1}^{c} Entrophy(j, after)$$

3. The algorithm chooses the property with the highest information gain or lowest entropy.
4. Depending on the chosen characteristic, the given S is divided.

5. The algorithm then iteratively applies to each subassembly, focusing only on characteristics that weren't previously chosen. Ginni Index Calculated as:

$$Ginni\ (Gi) = 1 - \sum_{k=1}^{j} (Pi)^2$$

6. To do this, the algorithm gradually divides the data base into various subgroups according to entropy or information quantity, and does so until a pause requirement is met.

7. Calculating Variance: Steps

$$Variance\ (Vi) = \frac{\sum \sqrt{(X - Xi)}}{n}$$

3.4 K- Nearest Neighbour:

The KNN is the simple algorithm in terms of implementation in machine learning methods is the k-NN (k-Nearest Neighbours) technique. In order to develop the model, the training dataset which acts as the reference data must be stored. The technique locates the closest data points, referred to as the "nearest neighbours," inside the training dataset when producing a estimate for a new statistics point.

The k-NN algorithm can be summed up mathematically as follows:

- Every point of data in the dataset used for training should be measured against the new data point (X).
- Based on the estimated distances, choose the k closest neighbours.
- When performing classification tasks, choose the dominating class among the k closest neighbours and set it as the predicted class for the newly generated data point.
- When completing regression activities, determine the average or weighted mean of the k nearest neighbour's goal values and utilise that quantity as the forecast value for the new data point.

Different proximity metrics, such as the distance from Manhattan or the distance calculated by Euclid, can be used to determine the separation among two data points based on the type of information and the issue at hand.

4. Performance Metrics:

The accuracy (ACC) is calculated as the percentage of correctly classified instances, whether they are normal or attacks, and is determined by the following formula:

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

The formula for calculating precision (P), which is the proportion of pertinent instances among the identified instances:

$$P = \frac{TP}{(TP + FP)}$$

Recall (R) is calculated as the ratio of the number of relevant instances over the total number of relevant instances discovered:

$$\mathrm{R} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}$$

The F1-Score is a metric that combines recall and precision into one number. It can be calculated using the formula below as the weighted average of recall and precision:

$$\mathrm{F1Score} = \frac{(2 * \mathrm{P} * \mathrm{R})}{(\mathrm{P} + \mathrm{R})}$$

In particular, when α = 1, the formula for the F1-Score simplifies. Overall, these formulas allow us to calculate accuracy, precision, recall, and the F1-Score, which are commonly used metrics for evaluating classification performance.

V. Result And Discussion

This study used a variety of machine learning techniques to create and assess a framework for predicting diabetes. Age, BMI, diabetic pedigree function, skin thickness, blood pressure, blood glucose levels, and insulin levels were among the nine factors in the sample, which contained 768 records. The performance of four classifiers Naive Bayes (NB), Decision Tree (DT), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM) was assessed using metrics such as accuracy, precision, recall, and F1-score.

Table 1: Performance metric comparison for different method

Method	**Accuracy in (%)**	**Precision in (%)**	**Recall in (%)**	**F1-Score in (%)**
NB	98.12	97.23	98.87	98.88
DT	98.87	98.01	99.1	98.72
KNN	99.11	98.01	98.66	98.81
SVM	99.65	97.33	98.44	98.76

One of the classifiers, SVM, demonstrated the best accuracy of 99.65%, proving its ability to accurately predict outcomes for both diabetes and non-diabetic individuals. SVM also managed to achieve a respectable precision of 97.33%, proving its ability to correctly distinguish real positive cases from all predicted positive ones. Additionally, SVM's strong recall rate of 98.44% showed that it could successfully identify a sizeable portion of the dataset's true positive cases.

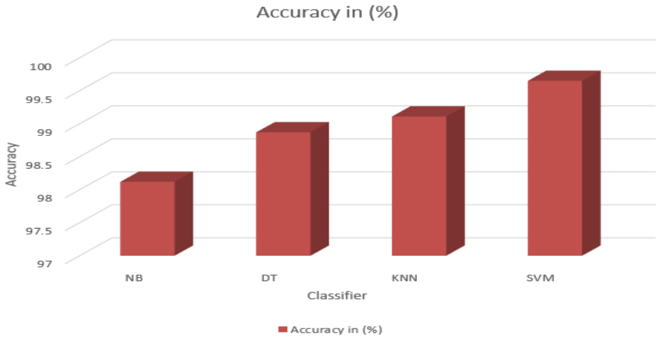

Fig. 3: Accuracy comparison of different methods

The Decision Tree classifier, which had an accuracy of 98.87%, came in second place to SVM. It demonstrated a balanced performance with a precision of 98.01%, a recall of 99.10%, and an F1-score of 98.72%. The accuracy, precision, recall, and F1-score of KNN were all excellent, with a combined score of 99.11%, 98.01%, 98.66%, and 98.81%.

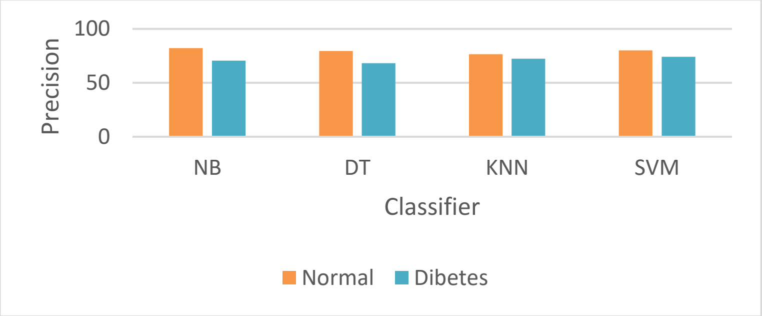

Fig. 4: Precision comparison of different method

The Naive Bayes classifier performed well overall, achieving 98.12% accuracy, 97.23% precision, 98.87% recall, and 98.88% F1-score. It showed high predictive powers for diabetes identification although performing somewhat worse than the other classifiers.SVM, one of the classifiers, showed the best accuracy of 99.65%, demonstrating its capacity to predict outcomes correctly for both diabetic and non-diabetic people. A reasonable precision of 97.33% was also attained by SVM, demonstrating its capacity to correctly identify genuine positive cases among all predicted positive cases. Additionally, SVM had a high recall rate of 98.44%, demonstrating its capacity to catch a significant fraction of the dataset's genuine positive cases.

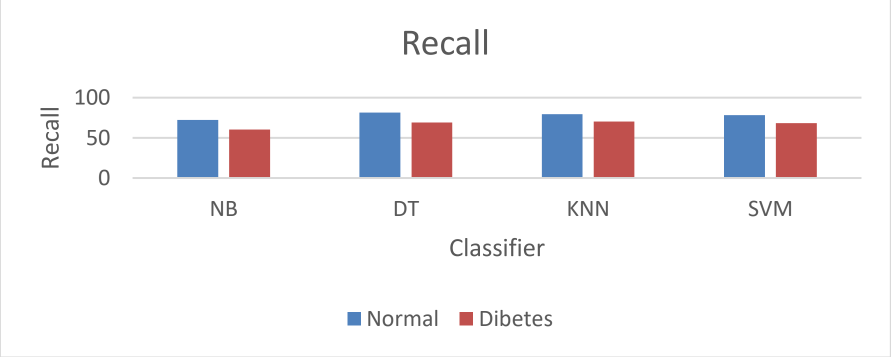

Fig. 5: Recall comparison of different method

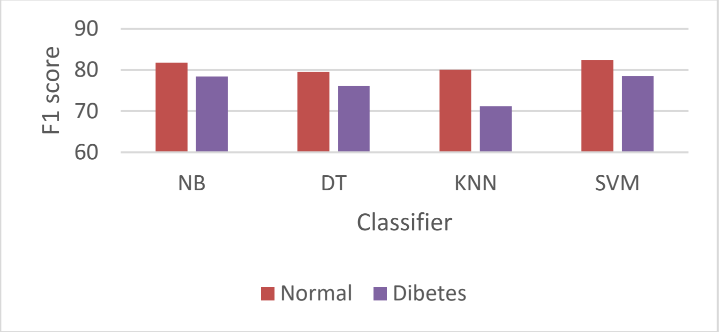

Fig. 6: F1 Score comparison of different method

Analysis of the data revealed that the classifiers' accuracy in differentiating between cases of normal and diabetes occurrence varied. In both normal and diabetic situations, SVM demonstrated the highest accuracy, obtaining 82.44% accuracy in normal cases and 78.54% accuracy in diabetic ones. This implies that SVM performed equally well in classifying both classes.Following closely, NB demonstrated competitive results when compared to the other classifiers, f1 score accuracy of 81.81% for normal cases and 78.44% for diabetic cases.Compared to SVM and NB, the Decision Tree classifier had a slightly lower accuracy of 79.55% for normal instances and 76.12% for diabetic cases, suggesting its potential to distinguish between the two classes as shown in figure 6.

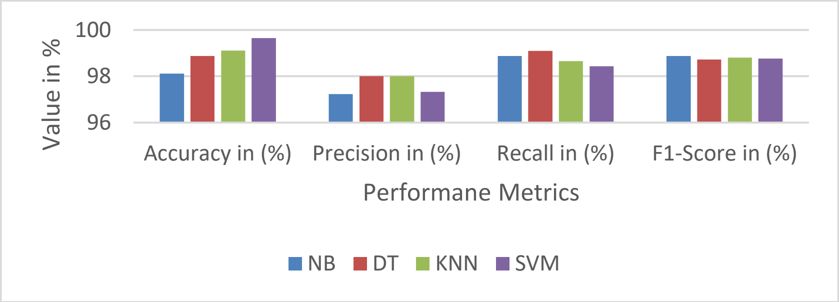

Fig. 7: Performance metric comparison for different method

The study was effective in creating a framework for predicting diabetes using machine learning algorithms and assessing its efficacy according to several measures. The findings suggest that SVM, DT, and KNN classifiers, each of which exhibits strengths in distinct performance areas, are particularly good in identifying diabetes. These results offer insightful information for the creation and application of a strong diabetes prediction system.

VI. Conclusion

Using medical data and disease influence variables, this study investigated the potential of machine learning algorithms for diabetes prediction. The dataset included 768 records with nine parameters including age, BMI, diabetes pedigree function, skin thickness, blood pressure, glucose levels, and insulin levels. The accuracy, precision, recall, and F1-score of four classifiers Naive Bayes (NB), Decision Tree (DT), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM) were assessed.The outcomes showed that SVM was the best-performing classifier, with a remarkable accuracy of 99.65%. It demonstrated outstanding prediction ability when separating cases with diabetes from those without diabetes. Further highlighting its capacity to accurately identify true positive cases and capture a sizeable part of actual positive cases in the dataset, SVM's high precision (97.33%) and recall (98.44%) were also very high.NB, DT, and KNN all shown great performance, albeit they each had different strengths in terms of diabetes prediction. While DT and KNN attained competitive accuracy rates of 98.87% and 99.11%, respectively, NB displayed robustness with an accuracy of 98.12% and excellent recall (98.87%).This study emphasizes the importance of machine learning in diabetes prediction and offers helpful information for the creation of effective diabetes detection systems. The findings imply that SVM is a potential contender for precise and trustworthy diabetes prediction, followed by NB, DT, and KNN. These findings will help academics and medical practitioners choose the best algorithms to help with early diabetes diagnosis and management, which will eventually enhance patient outcomes and public health.

References

[1] Moghissi, E.S.; Korytkowski, M.T.; DiNardo, M.; Einhorn, D.; Hellman, R.; Hirsch, I.B.; Inzucchi, S.E.; Ismail-Beigi, F.; Kirkman, M.S.; Umpierrez, G.E. (2009). "American Association of Clinical Endocrinologists and American Diabetes Association consensus statement on inpatient glycemic control." Diabetes Care, 32, 1119–1131.

[2] Borgharkar, S.S.; Das, S.S. (2019). "Real-world evidence of glycemic control among patients with type 2 diabetes mellitus in India: The TIGHT study." BMJ Open Diabetes Res. Care, 7, e000654.

[3] S. N. Ajani and S. Y. Amdani, "Probabilistic path planning using current obstacle position in static environment," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170727.

[4] S. Ajani and M. Wanjari, "An Efficient Approach for Clustering Uncertain Data Mining Based on Hash Indexing and Voronoi Clustering," 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 486-490, doi: 10.1109/CICN.2013.106.

[5] Fang, M.; Wang, D.; Coresh, J.; Selvin, E. (2021). "Trends in diabetes treatment and control in US adults, 1999–2018." New Engl. J. Med., 384, 2219–2228.

[6] Raveendran, A.V.; Chacko, E.C.; Pappachan, J.M. (2018). "Non-pharmacological treatment options in the management of diabetes mellitus." Eur. Endocrinol., 14, 31.

[7] American Diabetes Association. (2021). "Pharmacologic approaches to glycemic treatment: Standards of Medical Care in Diabetes—2021." Diabetes Care, 44(Suppl. S1), S111–S112.

[8] Garg, S.K.; Grunberger, G.; Weinstock, R.S.; Lawson, M.L.; Hirsch, I.B.; DiMeglio, L.A.; Pop-Busui, R.; Philis-Tsimikas, A.; Kipnes, M.S.; Lilenquist, D.R. (2023). "Improved Glycemia with Hybrid Closed-Loop (HCL) Versus Continuous Subcutaneous Insulin Infusion (CSII) Therapy: Results from a Randomized Controlled Trial." Diabetes Technol. Ther., 25, 1–12.

[9] Phillip, M.; Nimri, R.; Bergenstal, R.M.; Barnard-Kelly, K.; Danne, T.; Hovorka, R.; Kovatchev, B.P.; Messer, L.H.; Parkin, C.G.; Ambler-Osborn, L. (2022). "Consensus Recommendations for the Use of Automated Insulin Delivery (AID) Technologies in Clinical Practice." Endocr. Rev., 44, 254–280.

[10] Borkar, P., Wankhede, V.A., Mane, D.T. et al. Deep learning and image processing-based early detection of Alzheimer disease in cognitively normal individuals. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08615-w

[11] Ajani, S.N., Mulla, R.A., Limkar, S. et al. DLMBHCO: design of an augmented bioinspired deep learning-based multidomain body parameter analysis via heterogeneous correlative body organ analysis. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08613-y

[12] Battelino, T.; Alexander, C.M.; Amiel, S.A.; Arreaza-Rubin, G.; Beck, R.W.; Bergenstal, R.M.; Buckingham, B.A.; Carroll, J.; Ceriello, A.; Chow, E. (2022). "Continuous glucose monitoring and metrics for clinical trials: An international consensus statement." Lancet Diabetes Endocrinol., 11, 42–57.

[13] Pantalone, K.M.; Misra-Hebert, A.D.; Hobbs, T.M.; Wells, B.J.; Kong, S.X.; Chagin, K.; Dey, T.; Milinovich, A.; Weng, W.; Bauman, J.M. (2018). "Effect of glycemic control on the Diabetes Complications Severity Index score and development of complications in people with newly diagnosed type 2 diabetes." J. Diabetes, 10, 192–199.

[14] Pettus, J.H.; Zhou, F.L.; Shepherd, L.; Preblick, R.; Hunt, P.R.; Paranjape, S.; Miller, K.M.; Edelman, S.V. (2019). "Incidences of severe hypoglycemia and diabetic ketoacidosis and prevalence of microvascular complications stratified by age and glycemic control in US adult patients with type 1 diabetes: A real-world study." Diabetes Care, 42, 2220–2227.

[15] Basu, S.; Narayanaswamy, R. (2019). "A prediction model for uncontrolled type 2 diabetes mellitus incorporating area-level

social determinants of health." Med. Care, 57, 592–600.
[16] Chatterjee, R.; Yeh, H.C.; Edelman, D.; Brancati, F. (2011). "Potassium and risk of Type 2 diabetes." Expert Rev. Endocrinol. Metab., 6, 665–672.
[17] Jian, Y.; Pasquier, M.; Sagahyroon, A.; Aloul, F. (2021). "A Machine Learning Approach to Predicting Diabetes Complications." Healthcare, 9, 1712.
[18] Dinh, A.; Miertschin, S.; Young, A.; Mohanty, S.D. (2019). "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning." BMC Med. Inform. Decis. Mak., 19, 211.
[19] Zou, Q.; Qu, K.; Luo, Y.; Yin, D.; Ju, Y.; Tang, H. (2018). "Predicting diabetes mellitus with machine learning techniques." Front. Genet., 9, 515.
[20] Yang, L.; Gabriel, N.; Hernandez, I.; Winterstein, A.G.; Guo, J. (2021). "Using machine learning to identify diabetes patients with canagliflozin prescriptions at high-risk of lower extremity amputation using real-world data." Pharmacoepidemiol. Drug Saf., 30, 644–651.
[21] Del Parigi, A.; Tang, W.; Liu, D.; Lee, C.; Pratley, R. (2019). "Machine learning to identify predictors of glycemic control in type 2 diabetes: An analysis of target HbA1c reduction using empagliflozin/linagliptin data." Pharm. Med., 33, 209–217.
[22] Seo, W.; Lee, Y.-B.; Lee, S.; Jin, S.-M.; Park, S.-M. (2019). "A machine-learning approach to predict postprandial hypoglycemia." BMC Med. Inform. Decis. Mak., 19, 210.
[23] Hanson, R.L.; Imperatore, G.; Bennett, P.H.; Knowler, W.C. (2002). "Components of the “metabolic syndrome” and incidence of type 2 diabetes." Diabetes, 51, 3120–3127.
[24] Bhutto, A.R.; Abbasi, A.; Abro, A.H. (2019). "Correlation of hemoglobin A1c with red cell width distribution and other parameters of red blood cells in type II diabetes mellitus." Cureus, 11, e5533.
[25] All of Us Research Program Investigators. (2019). "The “All of Us” research program." N. Engl. J. Med., 381, 668–676.
[26] Ramirez, A.H.; Sulieman, L.; Schlueter, D.J.; Halvorson, A.; Qian, J.; Ratsimbazafy, F.; Loperena, R.; Mayo, K.; Basford, M.; Deflaux, N. (2022). "The All of Us Research Program: Data quality, utility, and diversity." Patterns, 3, 100570.
[27] R Core Team. (2022). "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing: Vienna, Austria. Available online: https://www.R-project.org/ (accessed on 13 April 2023).
[28] Menardi, G.; Torelli, N. (2014). "Training and assessing classification rules with imbalanced data." Data Min. Knowl. Discov., 28, 92–122.
[29] Fan, Y.; Long, E.; Cai, L.; Cao, Q.; Wu, X.; Tong, R. (2021). "Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent type 2 diabetes." Front. Pharmacol., 12, 1485.
[30] Motaib, I.; Aitlahbib, F.; Fadil, A.; Tlemcani, F.Z.R.; Elamari, S.; Laidi, S.; Chadli, A. (2022). "Predicting poor glycemic control during Ramadan among non-fasting patients with diabetes using artificial intelligence based machine learning models." Diabetes Res. Clin. Pract., 190, 109982.
[31] Tao, X.; Jiang, M.; Liu, Y.; Hu, Q.; Zhu, B.; Hu, J.; Guo, W.; Wu, X.; Xiong, Y.; Shi, X. (2022). "Predicting three-month fasting blood glucose and glycatedhemoglobin of patients with type 2 diabetes based on multiple machine learning algorithms." Research Square.
[32] Coregliano-Ring, L.; Goia-Nishide, K.; Rangel, É.B. (2022). "Hypokalemia in Diabetes Mellitus Setting." Medicina, 58, 431.
[33] Gulati, M. ., Yadav, R. K. ., & Tewari, G. . (2023). Physiological Conditions Monitoring System Based on IoT. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4s), 199–202. https://doi.org/10.17762/ijritcc.v11i4s.6514
[34] Prof. Parvaneh Basaligheh. (2017). Design and Implementation of High Speed Vedic Multiplier in SPARTAN 3 FPGA Device. International Journal of New Practices in Management and Engineering, 6(01), 14 - 19. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/51