Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 9s (2025)

A Machine Learning Approach Using Feature Selection and Scaling with Hyper Parameter Tuning Method for Early Prediction of Diabetes

[1]Jyoti N. Dangat (Shendage) , [2]Dr. Santosh P. Shrikhande, [3]Dr Vijaya Kumbhar

[1]School of Computer Studies, Sri Balaji University, Pune (SBUP) (MS)
Email: dangatjyoti@gmail.com
School of Technology
[2]S.R.T.M. University Nanded, Sub-Campus, Latur (MS)
Email: santoshshrikhande@gmail.com

[3]School of Computer Studies, Sri Balaji University, Pune (SBUP) (MS)
Email: veejeya.kumbhar@gmail.com

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract:

Introduction: Diabetes mellitus is an enduring condition characterized by raised blood glucose levels and has become a significant global health concern. Early and accurate diagnosis is crucial and that can help to prevent or delay of problems like cardiovascular diseases, kidney complications, nerve impairment, and vision diminishing due to diabetes. Diabetes is a chronic metabolic disorder that affects millions of people worldwide, leading to severe health complications if not diagnosed and managed in its early stages. Early prediction of diabetes is crucial for timely intervention and personalized treatment, reducing the risk of long-term complications. Traditional diagnostic approaches rely on clinical tests, which may not always be efficient in identifying high-risk individuals before the onset of the disease.

With advancements in artificial intelligence (AI) and machine learning (ML), predictive models have gained prominence in healthcare applications, offering improved accuracy and efficiency in disease diagnosis. However, the performance of these models heavily depends on the quality of input features, data preprocessing techniques, and hyperparameter tuning strategies.

Objectives: The main objective of this Research work is to predict diabetes at an early stage so that any severe complications may avoid.

Methods: The most significant and robust features of the dataset are chosen using the attribute selection tool and correlation attribute estimation method by using the WEKA software tool. Then, features form the dataset are scaled using the standardization feature scaling technique and different ML classification algorithms such as LR, KNN, Naïve Bayes, Support Vector Machine, Decision Tree and Random Forest are used for experimenting with the above machine algorithm on the PIMA Indian diabetes dataset in PythonIn the preprocessing method, identification and removal of null and duplicate values have been replaced with the mean values.

Results: By applying six different machine learning algorithms on PIMA Indian diabetes Dataset have shown that the K-Nearest Neighbor and Naïve Bayes both classifiers reported the maximum prediction accuracy of **81.82%,** followed by LR, SVM, and RF with accuracies of 79.87%, 79.22%, and 77.27% respectively.

Conclusions: By appropriate feature Selection and by hyperparameter tunning increase the diabetes prediction accuracy at an early stage.

Keywords: Diabetes Prediction, Feature Selection, Features Scaling, Hyper Parameter Tuning, Machine Learning Algorithms.

1. Introduction

Diabetes Mellitus or simply diabetes is one of the most widespread diseases in the world with the high mortality rate. Approximately 830 million people global have diabetes disease, with the majority living in the low and middle-income countries [1], [2]. Each year, diabetes is directly responsible for the millions of deaths worldwide. As per the International Diabetes Federation (IDF) projections indicates a concerning growth in the diabetes and approximately 1 in 8 adults will be living with diabetes by 2045 [3]. Diabetes is the condition considered by unusually maximum level of glucose in the blood because human body does not properly process the food for making energy. Diabetic human body does not produce enough amount of insulin from the pancreas and therefore glucose cannot be converted into energy. The general symptoms of diabetes are recurrent urination, weight loss, rise in thirst, growth in hunger, slowly recovery of injuries and dizziness etc. [4], [5]. If diabetes is not undergone proper treatments then it effects on other part of human body such as cardiovascular disease i.e. heart attacks, kidney diseases, blindness, nerve impairment etc. [6]. Hence, initial-stage diagnosis of diabetes is essential, as several cases become severe due to the delayed detection and medical treatments. Now a day, machine learning algorithm based approaches have gained prominence for their effectiveness in early prediction of diabetes. Therefore, machine learning algorithms are mostly used recently for early prediction of diabetes so that treatments can be taken quickly and avoid further complications. The proposed research work methodology for early prediction of the diabetes disease using different machine learning algorithms with feature selection, feature scaling and hyper parameter tuning techniques on the Indian diabetes PIMA dataset. The main contributions of this proposed methodology are: (i) select the robust and significant feature from the diabetes dataset using attribute selection tool and correlation attribute estimation method in the WEKA. (ii) Apply the standard scalar features scaling techniques for scaling the significant features (iii) Then, different machine learning algorithms with hyper parameter tuning techniques are used for classification of the diabetes dataset and improving the diabetes forecast results. The performance of the proposed methodology is verified using perdition accuracy percentage, confusion matrix, precision, recall, and F1 score measures.

2. Objectives

The primary objectives of this study are as follows:

1. **Develop an Early Prediction Model for Diabetes** – Utilize machine learning techniques to create an efficient and accurate model for the early detection of diabetes.

2. **Implement Feature Selection Techniques** – Identify the most relevant features that contribute to diabetes prediction, reducing computational complexity and improving model interpretability.

3. **Apply Data Scaling Methods** – Standardize and normalize data to enhance model performance by ensuring consistent feature contribution and reducing biases.

4. **Optimize Model Performance Using Hyperparameter Tuning** – Fine-tune machine learning algorithms to achieve the best predictive accuracy and generalization.

5. **Compare Machine Learning Algorithms** – To determine the most effective approach, evaluate and compare different ML models, such as Decision Trees, Support Vector Machines, Random Forest, and Deep Learning models.

6. **Enhance Predictive Accuracy and Efficiency** – Improve the overall performance of diabetes prediction models by integrating advanced data preprocessing and optimization techniques.

3. Methods

In the proposed research work, a machine learning approach is developed with different preprocessing techniques for significant feature selection, features scaling and hyper parameter tuning method for improving classifier performance. The significant features from the dataset play an important role in the accurate prediction of the diabetes disease. Therefore, most significant features from the database are selected using attribute selection tool and correlation attribute estimation method in the WEKA software tool. Then, selected significant features form the database are scaled using standardization feature scaling technique. Lastly, the different ML algorithms such as LR, KNN, NB, SVM, DT and RF algorithms and hyper parameter tuning technique are used for classification of dataset into diabetes classes. The Python Jupiter Notebook is used for implementing the code of classification algorithms of proposed method. Overall work flow of the proposed research ML model using preprocessing, feature selection, feature Scaling, hyperparameter tuning method is shown in the following Figure-1.

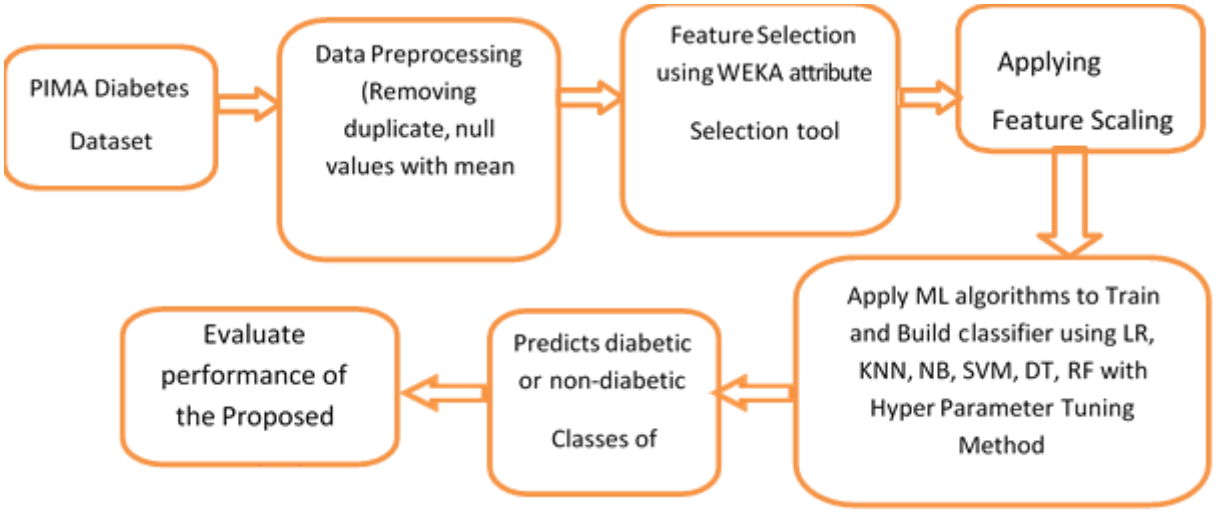

Figure 1. Proposed machine learning model for diabetes prediction

a. Data Set selection

The proposed approach using machine learning algorithms utilizes the PIMA Indian diabetes dataset, obtained from UCI Machine Learning Repository [22]. This dataset contains 768 records, including 268 individuals detected with diabetes and 500 with no diabetes condition. There are nine attributes available for every individual related to their health such as Pregnancy, BMI, Insulin level, Age, Blood pressure, Skin thickness, Glucose, Diabetes pedigree function, and Outcome. The "Outcome" attribute has two values: "0" for non-diabetes and "1" for diabetes. The following Table 1 shows the feature from the diabetes dataset.

Table 1. PIMA Indian diabetes data set with all nine features

Sr. No.	Feature/Attribute	Data Type	Description
1	**Pregnancies**	Integer	Number of times pregnancies
2	**Glucose**	Integer	Glucose level in Blood
3	**Blood Pressure**	Integer	Diastolic blood pressure measurement

4	**Skin Thickness**	Integer	Triceps skin thickness
5	**Insulin**	Integer	2 Hour Serum Insulin level in blood
6	**BMI**	Float	Body Mass Index
7	**Diabetes Pedigree Function**	Float	Diabetes percentage
8	**Age**	Integer	Patients Age
9	**Outcome**	Integer	Class 0-for Negative, 1-for Positive

b. Pre-Processing

The pre-processing techniques play a vital job in transforming dataset into different forms for better prediction accuracy in the diabetes forecast using different classification algorithms. Following task are achieved in the pre-processing of the PIMA diabetes set of data for the proposed research work.

i.Replacing Duplicate and Null Values

The duplicate and null values from the diabetes dataset are initially identified and replaced with the mean values in the Python using the functions available for the entire dataset.

ii.Feature Selection

The Pearson's correlation matrix and WEKA software tool is been used for the significant feature selection from the diabetes dataset. The correlation coefficient is considered in the proposed model that relates with input and output features. By applying feature selection technique on dataset, three features i.e. Skin Thickness, BMI, and Diabetes Pedigree Function are cancelled by checking its values in the correlation matrix table. Following Figure 2(a) and Figure 2(b) shows the correlation matrix before and after feature selection. In WEKA software, correlation filter was applied to determine coefficient correlation between all the attributes, with the results displayed in Table 2. A cut-off value of 2 was used to identify relevant features, leading to exclusion of three features from the dataset: Skin Thickness, BP and Diabetes Pedigree Function. The final model has selected five highly correlated attributes such as Glucose, BMI, Insulin, Pregnancies, and Age as the most related features for diabetes prediction.

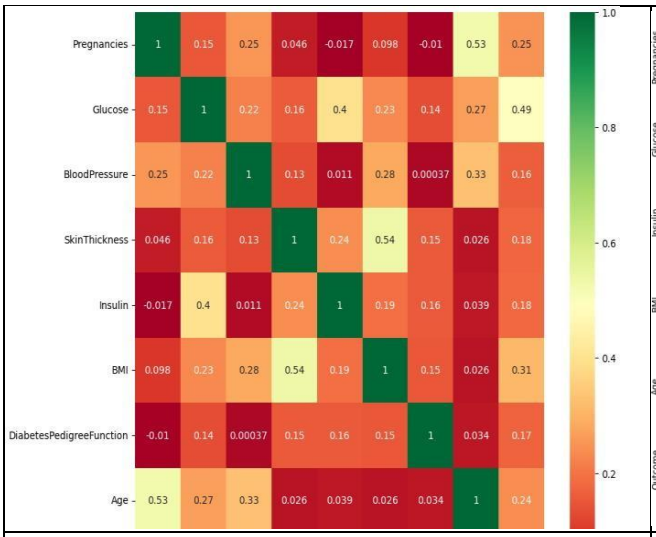

Figure 2(a). Before feature selection correlation between attribute.

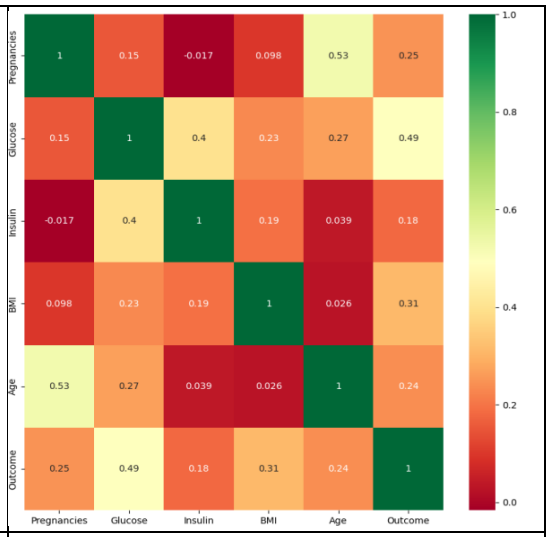

Figure 2 (b) . After feature selection correlation between attributes.

Table 2. The relation between input and output features using WEKA tool.

Attributes	Correlation Coefficient
Glucose	0.48
BMI	0.32
Insulin	0.26
Pregnancies	0.23
Age	0.22
Skin Thickness	0.19
Blood Pressure	0.18
Diabetes Pedigree Function	0.18

From above table-2, it has been observed that these three features are not so much correlated to diabetes prediction, and hence not significant in the diabetes prediction. Therefore, only significant correlated features are selected and used in the diabetes prediction. Now, dataset has five independent features like Pregnancies, Glucose, Blood Pressure, Insulin, Age and one dependent feature i.e. Outcome. The 'Glucose' and 'Outcome' attributes are having 0.49 correlation coefficient value and hence these are highly correlated. The following Table 3 shows selected significant features for the proposed approach.

Table 3. PIMA Indian diabetes dataset after significant feature selection

S. N.	Feature/Attribute	Data Type	Description
1	**Pregnancies**	Numeric	Number of times pregnancies
2	**Glucose**	Numeric	Glucose level in Blood
3	**Blood Pressure**	Numeric	Diastolic blood pressure measurement
4	**Insulin**	Numeric	2 Hour Serum Insulin level in blood
5	**Age**	Numeric	Patients Age
6	**Outcome**	Numeric	Class 0-for Negative, 1-for Positive

After applying the different pre-processing techniques on the PIMA diabetes dataset following observations are illustrated using histogram of the ranges and distribution of features which is shown in the following Figure 3.

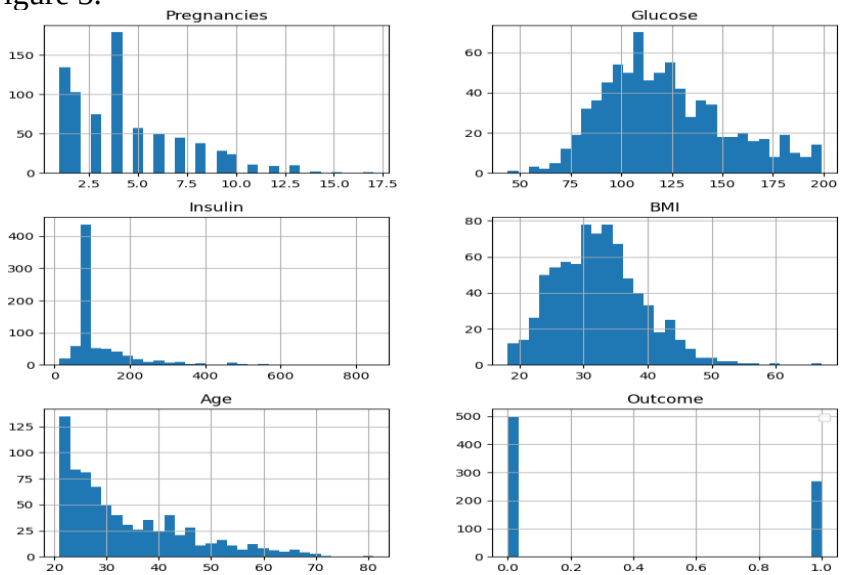

Figure 3. Histogram of attribute ranges and distribution after pre-processing

c. Standard Feature Scaling Technique

The diabetes dataset used for training the machine learning model generally contains uneven range of the values that affects the prediction accuracy. The machine learning algorithms work on numeric values and do not know what that numbers represents. The feature scaling techniques are used for solving this problem and by adjusting the values in same range and produce best predictions results. Therefore, in this research work Standard Feature Scaling technique is used for scaling dataset that boosted the diabetic prediction accuracy. The standard scalar is also known as z-score standardization. It is a scaling method where the values are centered on the mean with a unit standard deviation. This implies that the mean of the attribute is shifted to zero and the resulting spreading has a standard divergence of one. The mathematical representation of standard scalar is as below [7], [9].

$$Xnew = \frac{Xi - Xmean}{Standard\ Deviation} \quad (1)$$

d. Classification Algorithms

After applying the pre-processing techniques and selecting the significant features, the dataset has been prepared to train and test using machine learning algorithms. In the proposed approach, six different ML algorithms are applied to the PIMA diabetes dataset for classification of dataset into two predefined classes such as diabetic and non-diabetic.

4. Results

In proposed methodology, experiment is conducted using different ML algorithms with feature scaling, features selection and hyper parameter tuning method on PIMA diabetes dataset for classification of the diabetes. The hyper parameter tuning method is applied on the classifiers by selecting appropriate parameters that improves the diabetes prediction accuracy. In the hyper parameter tuning for K-NN algorithm kth value is set to 21. The experiments of the proposed approach are conducted in the Python's Jupyter Notebook environment. The diabetes prediction results using proposed approach on the PIMA diabetes set of data is given in the following table-4.

Table 4: Prediction accuracy before and after applying feature selection, feature scaling and hyperparameter tuning using PIMA dataset.

ML Technique	Prediction Accuracy BeforeFeature Selection, Scaling (with All 8 Features)	Prediction Accuracy AfterFeature Selection, Scaling with Hyper Parameter tuning (Selected 5 Features)
Logistic Regression (LR)	75.97 %	**79.87 %**
K-Nearest Neighbor (KNN)	72.72 %	**81.82 %**
Naïve Bayes (NB)	77.27 %	**81.82 %**
Support Vector Machine (SVM)	77.27 %	**79.22 %**
Decision Tree (DT)	69.48 %	**75.98 %**
Random Forest (RF)	75.97 %	**77.27 %**

From the above table is has been observed that the prediction results of after applying features

selection, feature scaling and hyper parameter tuning method are better than without applying pre-processing techniques. In case of Logistic Regression classifier prediction accuracy is increased from 75.97% to 79.87%, KNN accuracy increased from 72.72% to 81.82%, Naive Bayes accuracy is increased from 77.27 % to 81.82% that shows the significant transformation in diabetes prediction accuracy. In the proposed model the KNN and NB classifiers have shown 81.82 % as the highest prediction accuracy as compare to other classifiers. The following figures shows prediction accuracy before and after applying the feature selection, feature scaling and hyper parameter tuning techniques on the PIMA Dataset.

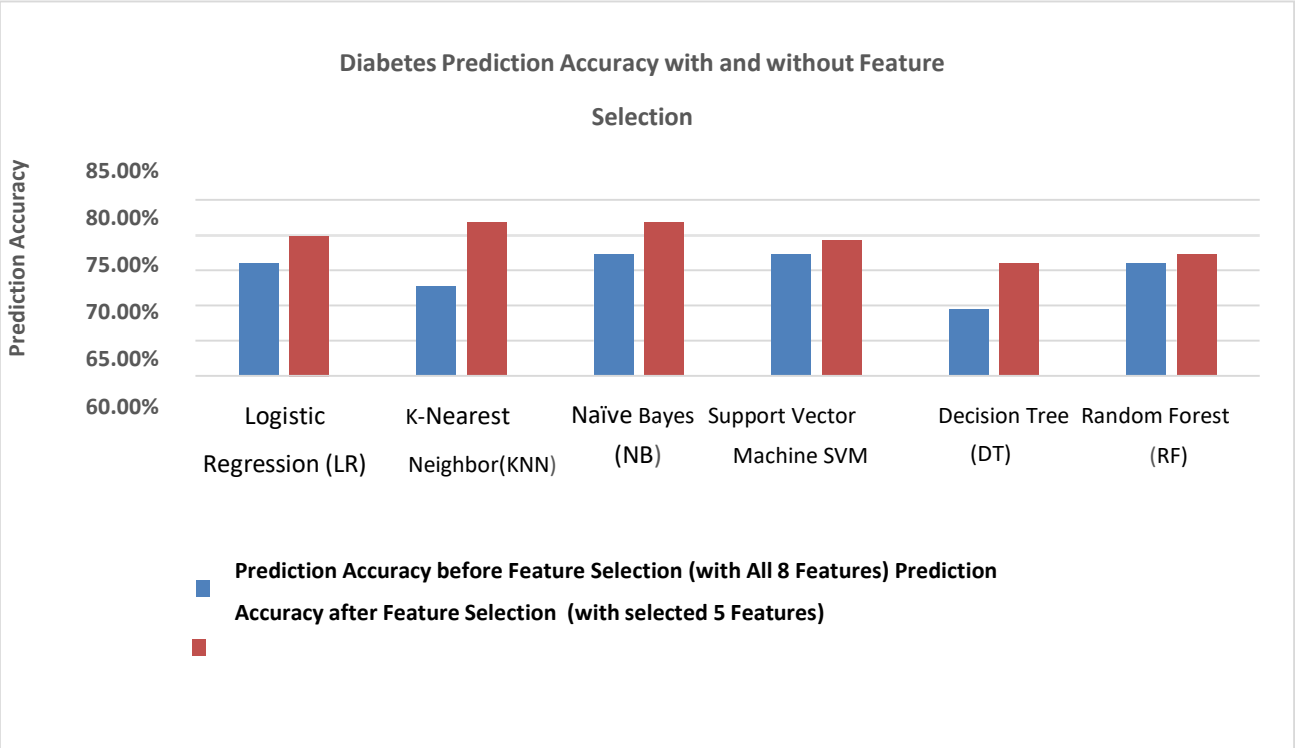

Figure 4. Diabetes prediction accuracy with and without pre-processing

The performance evaluation and classification report with confusion matrix of the proposed machine learning approach using PIMA diabetes dataset is shown in the following table.

Table-5. Performance evaluation and classification report of the proposed machine learning approach

ML Algorithms	Classification Report	Confusion Matrix
Logistic Regression	Classification Report of Logistic Regression: precision recall f1-score support 0 0.8224 0.8800 0.8502 100 1 0.7447 0.6481 0.6931 54 accuracy 0.7987 154 macro avg 0.7836 0.7641 0.7717 154 weighted avg 0.7952 0.7987 0.7951 154	Predicted values 0 1 All Actual Values 0 88 12 100 1 19 35 54 All 107 47 154

K-Nearest Neighbor	Classification Report of KNN: precision recall f1-score support 0 0.8158 0.9300 0.8692 100 1 0.8250 0.6111 0.7021 54 accuracy 0.8182 154 macro avg 0.8204 0.7706 0.7856 154 weighted avg 0.8190 0.8182 0.8106 154	Predicted values 0 1 All Actual Values 0 93 7 100 1 21 33 54 All 114 40 154
Naïve Bayes	Classification Report of NB: precision recall f1-score support 0 0.82 0.93 0.87 100 1 0.82 0.61 0.70 54 accuracy 0.82 154 macro avg 0.82 0.77 0.79 154 weighted avg 0.82 0.82 0.81 154	Predicted values 0 1 All Actual Values 0 93 7 100 1 21 33 54 All 114 40 154
Support Vector Machine	Classification Report of SVM: precision recall f1-score support 0 0.8091 0.8900 0.8476 100 1 0.7500 0.6111 0.6735 54 accuracy 0.7922 154 macro avg 0.7795 0.7506 0.7605 154 weighted avg 0.7884 0.7922 0.7866 154	Predicted values 0 1 All Actual Values 0 89 11 100 1 21 33 54 All 110 44 154
Decision Tree	Classification Report of Decision Tree: precision recall f1-score support 0 0.78 0.87 0.82 100 1 0.70 0.56 0.62 54 accuracy 0.76 154 macro avg 0.74 0.71 0.72 154 weighted avg 0.75 0.76 0.75 154	Predicted values 0 1 All Actual Values 0 87 13 100 1 24 30 54 All 111 43 154
Random Forest	Classification Report of Random Forest: precision recall f1-score support 0 0.7982 0.8700 0.8325 100 1 0.7111 0.5926 0.6465 54 accuracy 0.7727 154 macro avg 0.7546 0.7313 0.7395 154 weighted avg 0.7676 0.7727 0.7673 154	Predicted values 0 1 All Actual Values 0 87 13 100 1 22 32 54 All 109 45 154

The performance of the proposed methodology is compared with performances of different techniques form the literature. Following table shows diabetes prediction accuracies achieved by various researchers using different machine learning classification methods and compared with accuracy of proposed methodology.

Table- 6 . Diabetes prediction accuracy comparison of proposed method with existing methods

Ref. No.	Technique Used	LR	K –NN	NaiveBayes	SVM	Decision Tree	Random Forest
Gangani Dharmarathne [9]	DT, KNN, SVC, XGB	-	77.0 %	--	77.0 %	76.0 %	--

Muhamad Exell Febriana [10]	KNN, Naive Bayes	-	73.33%	76.07 %	--	--	--
Hakim El Massari[11]	SVM, KNN, ANN, NB, LR, DT, Ontology	77.2 %	70.2 %	79.0 %	77.3 %	--	73.8 %
Jobeda Jamal [12]	DT, RF, NB, LR, KNN, AB, SVM	78.86 %	79.42 %	78.28 %	77.71 %	73.14 %	77.34 %
Neha P.Tiggaa [13]	LR,KNN,SVM, NB, RF	74.4 %	70.8 %	68.9 %	74.4 %	69.7 %	75.0 %
Gaurav Tripathi [14]	LDA, KNN, SVM, RF	-	79.24 %	--	80.85 %	--	87.66%
KM Jyoti [17]	KNN, LR, SVM	78.0 %	78.0%	--	77.0 %	--	--
Priyanka Sonar [18]	DT, ANN, NB,SVM	-	--	77.0 %	77.3%	85.0%	--
Proposed Method Results	LR, KNN, NB, SVM, DT, RF	79.87 %	81.82 %	81.82%	79.22 %	75.98 %	77.27 %

From the above predication accuracy table-6 it was clear that the proposed ML approach using different algorithms performed acceptable prediction accuracy in comparison with other methodologies form the literature. The better diabetes prediction accuracy result obtained by KNN and Naïve Bayes is 81.82 % as compare to the other classifiers.

5. Discussion

In this research work a novel approach was developed using machine learning algorithms with features selection, features scaling and hyper parameter tuning technique for diabetes disease prediction. The significant and robust features from dataset are selected using attribute selection tool and correlation attribute estimation method in WEKA software. Afterwards, the standard features scaling technique is used for the scaling of selected significant features. Then, ML algorithms such as LR, KNN, Naive Bayes, SVM, DT, and RF along with hyper parameter tuning technique are used for the classification purpose. After conducting experiments on PIMA Indian dataset with selected significant features and above classification algorithms with hyper parameter tuning, KNN and NB classifiers have shown the highest prediction accuracy of **81.82%.** Then, Logistic Regression, Support Vector Machine and Random Forest have shown 79.87%, 79.22%, and 77.27% of prediction accuracy respectively. In the forthcoming research, we aim to develop a robust methodology using machine learning algorithms for diabetes prediction that would be focusing on better prediction accuracy using different techniques.

Refrences

[1] World Health Organization's Health Topics, https://www.who.int/health-topics/diabetes#tab=tab_1, Accessed on January 2024.

[2] Guidance on global monitoring for diabetes prevention and control: framework, indicators and application, Geneva: World Health Organization; 2024, https://iris.who.int/bitstream/handle/10665/379529/9789240102248-eng.pdf?sequence=1

[3] International Diabetes Federation (IDF), Diabetes Atlas, Brussels, Belgium: 2023, Online Available at https://diabetesatlas.org/atlas/diabetes-and-kidney-diseases/

[4] American Diabetes Association: Diagnosis and Classification of Diabetes Mellitus In: Diabetes Care, Volume 3, Supplement 1, pp. s81- s90, January 2014.

[5] Zidian Xie, Olga Nikolayeva, Jiebo Luo: Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques In Preventing Chronic Disease, Public Health Research, Practice, and Policy, Vol. 16, E-130, pp.1-6, September-2019

[6] Santosh P. Shrikhande, Prashant P. Agnihotri, "Comparative Study of Various Data Mining techniques for Early Prediction of Diabetes Disease", International Journal of Scientific research in Computer Science, Engineering and Information Technology, Volume 8, Issue 1, pp. 287-295, 2022.

[7] Ong yee hang, Virgianti Wiwied, Rosly Rosaida, "Diabetes prediction using Machine Learning", Journal of Advanced Research in Applied Sciences and Engineering Technology, 37, Issue 1, 2024

[8] Mazen Alzyoud, Raed Alazaidah, Mohammad Aljaidi, Ghassan Samara, Mais Haj Qasem, Muhammad Khalid, and Najah AlShanableh, "Diagnosing diabetes mellitus using machine learning techniques", International Journal of Data and Network Science 8, pp. 179–188, 2024

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

[10] Muhammad Exell Febrian, Fransiskus Xaverius Ferdinana, Gustian Paul Sendani, Kristein Margi, Suryanigrum, Rezki Yunanda, "Diabetes prediction using supervised machine learning", Science Direct,7th International Conference on Computer Science and Computational Intelligence, Procedia Computer Science 216, pp. 21–30, 2023

[11] Hakim El Massari, Zineb Sabouri, Sajida Mhammedi, Noreddine Gherabi, "Diabetes Prediction Using Machine Learning Algorithms and Ontology", Journal of ICT Standardization, Vol. 10 2, pp. 319–338, 2022.

[12] Jobeda J. Khanam, Simon Y. Foo, "A comparison of machine learning algorithms for diabetes prediction", ICT Express Vol-7, No. 4, pp. 432-439, Published by Elsevier, pp.1-8, 2021.

[13] Neha Prerna Tiggaa, Shruti Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods", International Conference on Computational Intelligence and Data Science (ICCIDS), 2019

[14] Gaurav Tripathi,, Rakesh Kumar, "Early Prediction of Diabetes Mellitus Using Machine Learning", 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Amity University, Noida, India, 2020.

[15] Sivaranjani S, Ananya S, Aravinth J, Karthika R., "Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction", 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021.

[16] MD. Kamrul Hasan, MD. Ashraful Alam, Dola Das, Eklas Hossain, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers", Digital Object Identifier 10.1109/. DOI 10.1109/ACCESS.2020.2989857, IEEE Access, 2020

[17] KM Jyoti, "Diabetes Prediction Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology 64718, 2020.

[18] Priyanka Sonar, Prof. K. Jaya Malin, "Diabetes Prediction Using Different Machine Learning Approach", Proceedings of third International Conference on Computing Methodologies and

Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4.

[19] Aishwarya Mujumdar, Dr. Vaidehi V, “Diabetes Prediction using Machine Learning Algorithms”, Science Direct, International Conference on Recent Trends in Advance Computing, 2019.

[20] Mariwan Ahmed Hama Saeed, “Diabetes type 2 classification using machine learning algorithms with up-sam technique”, Springer open access, journal of Electrical Systems and Inf. Technology (2023).

[21] Victor Chang, Jozeene Bailey, Qianwen Ariel Xu, Zhili Sun, ”Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms”, Neural Computing and Applications (2023)35:16157-16173.

[22] Pima Indian Diabetes Database-Predict the onset of diabetes based on diagnostic-https://www.kaggle.com/uciml/pima-indians-diabetes-database.

[23] Henock M. Deberneh and Intaek Kim, “Pridiction of Type 2 Diabetes Based on Machine Learning Algorithm”, International Journal of Environmental Research and Public Health 2021.

[24] Md. Maniruzzaman, Md. Jahanur Rahman, Benojir Ahammed, Md. Menhazul Abedin, “Classification and prediction of diabetes disease using machine learning paradigm”, Health Information Science and Systems (Springer Nature), pp. 1-14, 2020.

[25] Amani Yahyaoui , Akhtar Jamil, Jawad Rasheed, Mirsat Yesiltpr, “A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques” 2019 IEEE.

[26] Muhammad Azeem Sarwar, Nasir Kamal, Wajeeha Hamid, Munam Ali Shah, “Prediction of Diabetes Using Machine Learning Algorithms in Healthcare”, Proceedings of the 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2018.