A novel approach for detection of diabetes patients using machine learning techniques

P.G.Palanimani
Professor
Department of Mathematics
Erode Sengunthar Engineering College
Anna University
Tamil Nadu, India.
Palanimani@gmail.com

Dr Jenifer Jothi Mary A
Assistant Professor
Department of Computer Science
St. Joseph's College (Autonomous)
Affiliated to Bharathidasan University
Tiruchirappalli, Tamil Nadu, India.
ajenifer.jothi@gmail.com

Dr. V. Suresh Kumar
Assistant Professor (selection Grade)
Department of ECE
SRM Valliammai Engineering College
Kanchipuram, Tamil Nadu, India.
sureshkumarv.ece@srmvalliammai.ac.in

Dr. Sanket B. Kasturiwala
Department of Electronics & Telecommunication Engg.,
G.H.Raisoni Institute of Engineering and Technology
Nagpur, India.
sanket.kasturiwala@gmail.com

Shafaque Ahmareen
Computer Science Instructor
Department of Computer Science
Ministry of Education, UAE.
shafaqueahmareen7@gmail.com

Sneha Bohra
Assistant Professor
Department of Computer Science and Engineering
G H Raisoni University, Amravati
Maharashtra, India.
snehab30@gmail.com

***Abstract*—Supporting medical decisions with data mining techniques allows for more efficient detection and treatment of disease while reducing the burden on doctors. Using data mining methods to diagnose diabetes in advance. Diseases of the kidney, eye, and heart are all linked to diabetes mellitus, which has the fourth highest fatality rate of any disease in the world. As a result, many people today suffer from the effects of this illness. This is where the bulk of current study focuses. Using data mining techniques aids in making accurate medical diagnoses. Different parts of the body are impacted by the chronic disease diabetes mellitus. The ability to accurately predict the spread of illnesses early on has the potential to save lives and give us command of them. Historically, diabetes has been diagnosed using a battery of physical exams; however, these procedures have been shown to be inaccurate. This study predicts the condition by utilising a variety of Data Mining techniques for the purpose of both predicting and diagnosing diabetes mellitus, therefore overcoming this constraint. Naive Bayes, K-Nearest Neighbor, Support Vector Machine, and Decision Tree are the most popular data mining methods. Seven hundred and sixty-eight occurrences from the Pima Indian Diabetes Dataset were used to create this dataset. Diabetic risk is scored and assigned to one of three categories: mild, moderate, and severe. In addition, the effectiveness of various algorithms for diabetes diagnosis has been analysed using this data. The outcomes obtained demonstrate how effective our classification method.**

Keywords—Diabetic prediction, machine learning, feature selection and classification, Neural network, decision making.

I. Introduction

For decades, Diabetes mellitus has plagued the world's population. It's a collection of metabolic illnesses characterised by a chronic disease caused by excessive blood sugar, unhealthy eating, and inactivity. Types of diabetes include type1, type2, and gestational. Type1 diabetes develops in childhood, but type2 diabetes occurs at any age, primarily in people over 40. Gestational diabetes affects pregnant women. According to WHO statistics, 79% of diabetes-related deaths occurred in adults under 60. To handle the massive amount, speed, variety, validity, and estimation of information, a scalable infrastructure is needed. To address this constraint, use multiple Data Mining methods to predict and diagnose diabetes mellitus.

Diabetes is diagnosed with physical tests, but they're inaccurate. It failed with global cluster. It's slower.

Diabetics must continually maintain their blood sugar and nutrition, exercise, medicine, etc. However, diabetics struggle with this. To maintain their diet, they need management and family assistance. A fitness lifestyle. Thus, diabetics need a strategy to control their blood sugar. Natural healing ulcers are caused by loss of consciousness and high pressure feet. Four therapy methods to alleviate anterior head, heel, and foot pressure when walking are suggested in this study. A group of diabetics wore four insoles, including the project's computer model, during the stress-measured investigation. This study examines the effects of soot pressure on different insole soles in a recommended and random sample of early-stage diabetes.

It is important to realize the potential of implant devices for diabetics through diagnostic and diagnostic tools with minimal variance performance definitions such as ensuring adequate control of blood glucose levels. This task is to determine the appropriate blood glucose so that there is a minimum number of frequencies tracking taken to deliver the patient's blood glucose level with a good understanding of the model. In the whole community, the cost of diabetes monitoring is a very important issue, and these costs vary depending on the monitoring method and frequency.

Diabetic disease is an important health problem, a lot of people are affected by these diseases around the world. To diagnose the Diabetic as a result of the application is a classification based on a valid machine

2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC) | 978-1-6654-5687-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/IIHC55949.2022.10059675

learning to reduce the mortality. In order to improve the predictive ability of the pre-processing machine learning model data, standardization of data it is very important. The proposed first step is pre-processing for eliminate the unwanted records. The term feature selection refers to the process of selecting the optimal features (i.e., only the most relevant features). Then the proposed ACNN algorithm is used to classify the heart disease patients.

Artificial neural networks (ANNs) are distributed massively parallel processors made up of simple processing units that have the habit of accumulating experimental information and making it available for usage. The ANNs exhibit behaviour akin to that of biological neurons by assembling complex structures. Many millions of neurons in the human brain exchange information via chemical and electrical signals.

As a mathematical representation of neurons, ANNs are able to share information and gain knowledge through interaction. Synaptic weights, the strength of the connections between neurons, are where the brain keeps track of the information it gathers from its surroundings. The primary goal of neural network learning is to improve accuracy by decreasing training-error variance through recurrent weight changes. With the discovered relationship between the variables being unknown or complex, ANNs find widespread use in a variety of scientific problems based on regression, pattern recognition, classification, optimization, and signal processing. Any exceptionally big architecture may over fit the training set due to a surplus of information processing capability, and this might have a negative effect on the performance of the ANNs. Because of its limited ability to analyse data, a little architecture fails to achieve good results on the training set. Inadequate generalisations result from both over fitting and under fitting.

II. LITERATURE REVIEW

Effective component selection is crucial to the predicted performance of any machine learning system [13]. Predictive analytics and machine learning competition problems are important in facility selection [14]. For diabetes prediction, we employ the Least Absolute Shrinkage and Selection Operator (LASSO) technique, which selects the highest-efficiency feature. Predicting the spread of diabetes based on the correlation between individual environmental risk variables is possible with the use of a Bayesian Network (BN) [15]. Applying the BN model allows one to quantify the connection between direct and indirect peril.

In its broadest sense, diabetes mellitus (DM) refers to a cluster of metabolic diseases that threaten people all over the world. There is now a big database on various diabetes-related topics. Data about diabetes provide difficulties in the medical field (unstructured). It is crucial for the system to place an emphasis on these massive amounts of diabetic data in order to produce reliable prediction outcomes [16]. The primary responsibility of this forecast is to choose relevant features. To improve diabetes prognosis, feature selection algorithms are used [17]. [18]. Diabetes data in the health care industry are murky (unstructured). Therefore, the system must place an emphasis on these massive amounts of diabetic data in order to produce reliable prediction outcomes.

Selecting relevant attributes is the primary objective of this forecast. It is possible to get an accurate forecast for diabetes with the use of feature selection algorithms [19]. Monitoring Long-term diabetes pain and discomfort can be avoided with regular blood glucose monitoring if the patient does not use the currently available blood glucose monitoring equipment. The management of diabetes requires regular glucose monitoring. Breath acetone concentrations are abnormally low in diabetics, and reports also indicate a gradual elevation in blood glucose levels in these patients [20].

Specific indicators in elevated blood glucose levels are diagnostic of diabetes mellitus (BGLs). Therefore, analysing human breathing can help detect diabetes and forecast BGLs [21]. Documentary data classification of diabetes records is one form of computational diagnosis. Cardiac autonomic neuropathy (CAN) is a common diabetic consequence marked by a subclinical abnormality in the dynamics of the ventricular repolarization (VR) due to a progressive degeneration of autonomic nerve fibres. Rarely does a clinical diagnosis lead to a positive treatment outcome [22]. Many people with diabetes experience complications and preventable injuries despite the fact that there have been significant shifts in high-risk factors [23]. Consequences of increased identification are less common when risk variables are high.

Type 2 diabetes has been linked to a wide range of lifestyle shifts, and its prevalence has recently been confirmed to be on the rise. Diabetes is a disease in which the amount of sugar in the blood is persistently high [24]. In medicine, classification systems are employed frequently as predictive models for analysing patient data or as guidelines for making diagnoses. To improve diabetes categorization [25]. Multiple approaches are currently being developed to organise and put into practise data sets on diabetes.

III. MATERIALS AND METHODS

The suggested approach aids patients in determining whether or not they are suffering from Diabetes Mellitus without the need for multiple diagnostic procedures, such as blood testing, evaluating diastolic and systolic blood pressure, etc. These estimates are founded on the presence of certain physical factors and the symptoms seen in the onset of Diabetes Mellitus. Successful early disease prediction using data mining techniques has the potential to save lives and cut down on healthcare costs.

Data mining methods including Naive Bayes, Decision Tree, and Support Vector Machine (SVM) models for diabetes prediction using common risk factors are compared and contrasted. Best classification accuracy was achieved with the decision tree model, followed by Naive Bayes and the support vector machine. Patients are ranked according to how likely they are to develop diabetes on a scale from low to high. Improve the result's precision. It was a cost saver.

A. Data Preprocessing

The dataset that was utilised in this investigation is one that perhaps has some discrepancies. In order to eliminate these inconsistencies, the data must first be preprocessed. A supervised attribute filtering method was utilised throughout the data preprocessing stage of the project. In order to extract useful intervals of data, a discretize filter was applied.

In order to calculate the collective dataset, you must first compute For. (CdsI at the initialization at J feature) as the diabetic dataset. After that, you must check empty, null case, and relevant characteristics before filtering them. It has deleted all of the null and incorrect data from the dataset that had been used as input in this research.

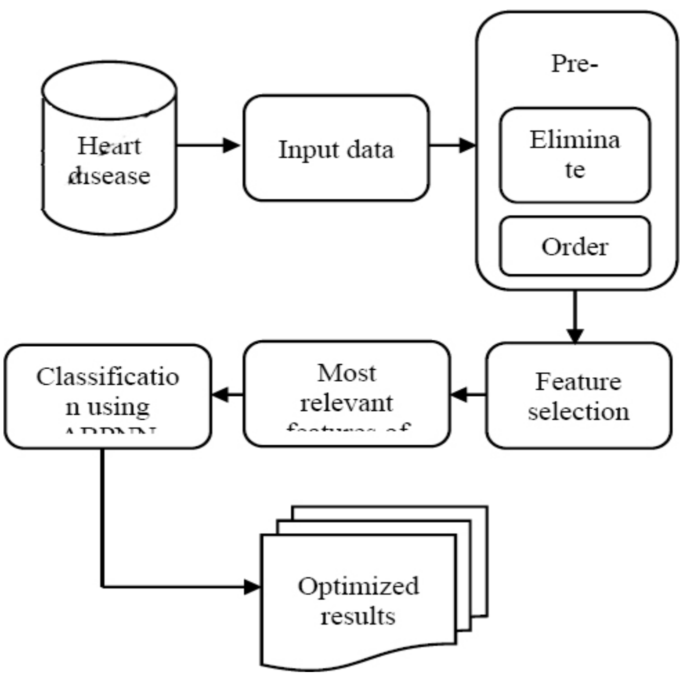

Fig 1. Architecture diagram of ABPNN

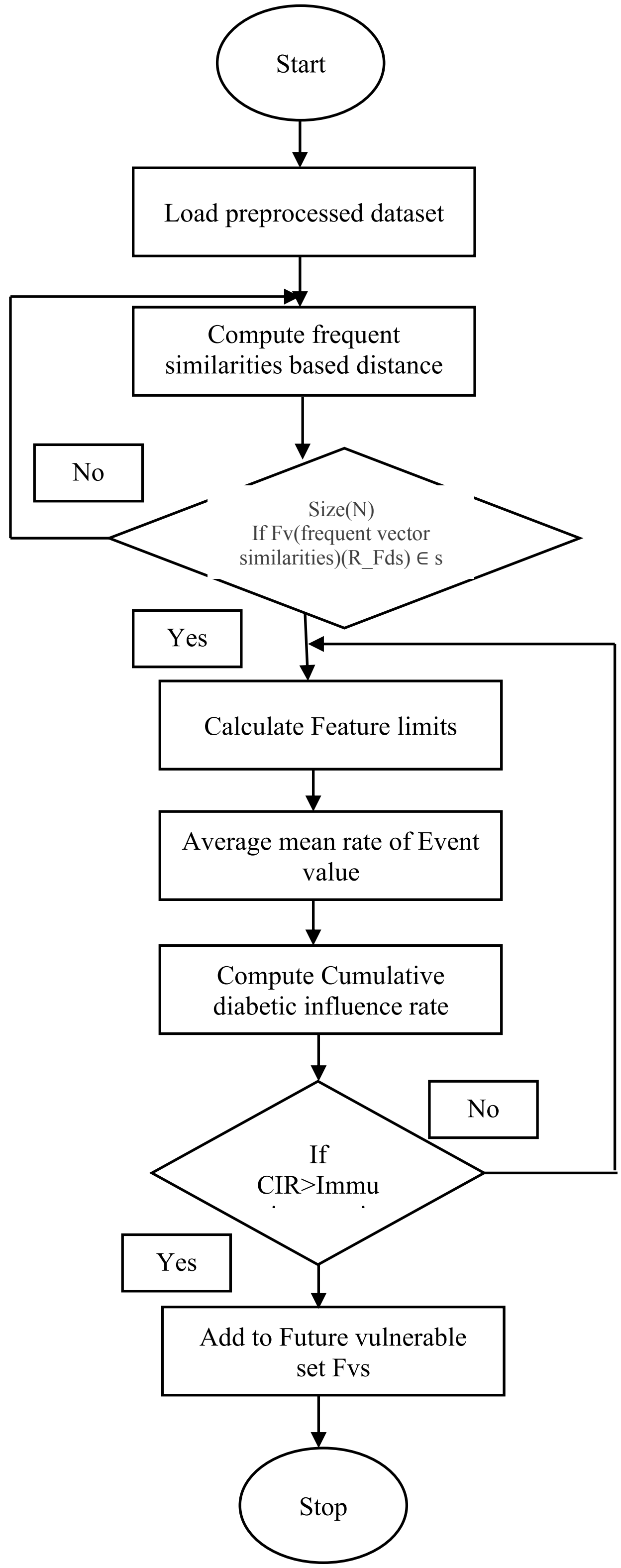

Fig 2. Work flow diagram for cumulative feature selection

B. Cumulative Behavioral diabetic margin rate

The selection of features is the most crucial part of In addition to the identification of the optimal feature set, there is a view that can lead to the contribution of this disease's classification manifold. In addition, this feature set can be of use to the specialists in the subject. In light of this, the primary objective of this research is to investigate the significance of the characteristic in question. A review of the performance comparison of the many methods of feature selection that are known is necessary for determining the optimal feature set.

Regarding every single PC pattern class

Compute the hidden layer neurons weight to c as

$$\text{set} = \int_{i=1}^{size(Apt)} \sum Apt(i).class = c$$

Closest pattern Pps = Closest pattern (Cset).

The relativity depends on the number of times the key terms have occurred. Once again, key terms also add up the most credible value for the unnecessary Diabetic events

For each closest pattern on the relative link, each pattern p.

The number of occurrences forming the feature relationships is identified using the frequency phenomenon of transformation feature weight.

By each similarity, features are classified as categories.

$$\text{Pfs} = \frac{\int_{i=1}^{size(p)} \sum P(i) == Scs(i)}{size(p)}$$

The framework should create a graph of the dataset by tracking sufficient binding weights. The amount of nodes in the input or data layer is set to selective features diabetic dataset be trained with decisive logical rules at the rate of margin events

$$\text{Compute cumulative PFS} = \frac{\sum_{i=1}^{size(pps)} Pfs}{size(pps)}$$

Behavior optimization Maximum values are Ps = PFs.

To achieve a consistent data collection, first the features build relational attribute form of predictions and negatively effect neural network categorization. They choose one of the most useful firefly subset feature selection models to provide the best selective case for classification accuracy characteristics.

C. Fuzzified Maximum subset Deterministic feature selection

This step measures maximal region similarity by encompassing scaling values from best pattern behavioural activity. Rough set groups create fuzzy rules for non-deterministic lower features. Multi-attributes are identical to nature, deemed unique, and similar to another trait in nature.

Algorithm

Step1: input PDs data initialization.

Step2: perfect logs PDsFor each Class Cl→Ts

Identify search term attribute for frequent query

Attribute For each Cl→A_i ofFv_i

Pattern compute data PCl = $\int_{i=1}^{N} \sum \left(Ai(Fvi) - Ai(Fv)\right)^2$

End

Ds(i) = $\sum Dsi + Cl$

End

End

Step3: identify each class Cl of data request set Ts

Ai →for each case attribute

Compute the Relational feature count

SC = $\int_{i=1}^{N} \sum Dsi(Ai) \geq STh$

End

Measure relative pattern case Dm = $\frac{Sc}{size(Cli)} \times 100$

End

Step4: read end

After that, each segment vector is used as an input vector, while the normal of each attribute is maximised in its features. During the training phase, the neural network is constructed, and then the network is trained by supplying input and output based on various features that have been chosen.

Adaptive Back propagation Artificial Neural Network

Due to their large scale and various sources, real-world databases have noisy, missing, and inconsistent data. The clustered dataset is entered into the classification model, which classifies diabetes risk as minimal, moderate, or severe.

Neural Networks include three layers: input, hidden, and output [20]. Each neuron in a layer is related to every neuron in the preceding layer. Back propagation network (BPN) algorithm is commonly used for ANN training. BPN algorithm uses gradient rule to find neuron weights. The weighted sum and bias term "bj" of a middle layer neuron "yj" are added to calculate its output (1).

$$y_j = f\left(\sum_{i=1}^{n} w_{ij} x_i + b_j\right) \qquad (j = 1,2,\ldots,n)$$

Where (xi, wij, bj) is the input signal, (f) is the non-linear activation function, (wij) is the weight from the output of the input unit to the successive output unit, and (bj) is the bias term [16,17]. The training procedure is measured by the sum square error (SSE). All training patterns and network outputs are computed through the equation (2).

$$E(x,w)=\frac{1}{2}\sum_{p=1}^{P}\sum_{m=1}^{M}e_{p,m}^{2}$$

When applying a pattern, p the training error at the output, m is defined by equation (3).

$$e_{p,m}=d_{p,m}-o_{p,m}$$

Training an ANN involves four stages: I gathering relevant training data; ii) creating a network object; iii) training the network; and iv) simulating the network's reaction to novel inputs. After proper training, the network is capable of associating nonlinear patterns between the target and input variables.

IV. Results And Discussion

The findings are incorporated into a data processing simulation tool used for machine learning. After being put through its paces on a wide variety of biological datasets, the proposed implementation algorithm has been shown to yield a higher detection rate by classifying results under a greater degree of classes. Analyzing test-results-parameters via a confusion matrix. The following matrix of possible meanings is given for your reference.

Table 1 displays all of the attributes used in this study, and the data set contains information for 650 diabetes patients across all age ranges.

TABLE I: THE ATTRIBUTES USED IN DATA SET

Attribute	Description	Type
Gender	Considered as Male=1 Female=0	Numeric
Insulin dependent	Considered as min=50and max=500	Numeric
Plasma	Considered as min=2 and max=11	Numeric
Systolic	blood pressure (Systolic)Considered as min=30 and Max=370	Numeric
Diastolic	blood pressure (Diastolic) Considered as min=60 and max=350	Numeric
Blood Group	Blood group Considered as 0= 'O',1= 'A',2 = 'B',3 = 'AB'	Nominal
Age	Considered as min=1 and max=125	Numeric

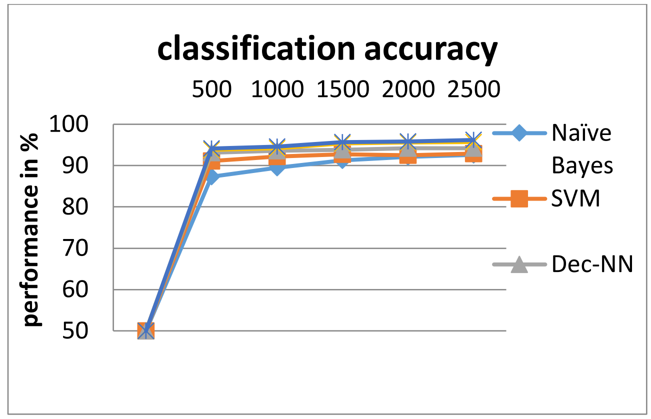

Fig 3. Impact of classification accuracy

One of the most common measures of a classifier's effectiveness is its accuracy at making classifications. Ratios are calculated by dividing the total number of cases by the number of positive and negative classifications obtained using the algorithms.

TABLE II. IMPACT OF CLASSIFICATION ACCURACY

Methods/Datasets	Impact of Classification Accuracy in %				
	Naïve Bayes	SVM	Dec-NN	SNC-MKFJC	AVMT-NN
500	87.3	91.1	93.1	93.8	94.1
1000	89.5	92.2	93.6	94.2	94.6
1500	91.3	92.7	93.8	95.4	95.7
2000	92.1	92.5	94.2	95.6	95.8
2500	92.6	92.9	94.2	95.7	96.2

Classification Accuracy Comparison Table I evaluates the performance of several approaches with regards to detection accuracy. The sensitivity of the classification method is another common criterion. Exact positive correlation with exact negative values at the margins. Classification.

$$\text{Sensitivity}=\frac{\text{TP}}{\text{TP}+\text{FN}}$$

Five very different data sets are used to estimate the sensitivity. The SSO classifier achieves 87.1% sensitivity whereas the AVMT-NN esteem achieves 94.1% sensitivity and a regular neural network achieves 88.3% sensitivity. We find that the proposed system has a more profound effect on sensitivity.

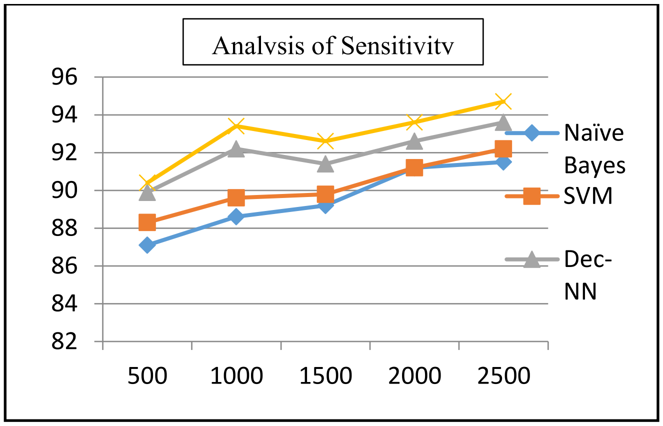

Fig 4. Impact of sensitivity analysis

As can be seen in Figure 4, the projected AVMT-NN method has provided a greater performance rate than additional existing ways when it comes to detecting and classifying data from a variety of logs from diverse datasets.

TABLE III. IMPACT OF SENSITIVITY ANALYSIS

Methods/Datasets	Impact of Sensitivity Analysis in %				
	Naïve Bayes	SVM	Dec-NN	SNC-MKFJC	AVMT-NN
500	87.1	88.3	89.9	90.4	91.6
1000	88.6	89.6	92.2	93.4	94.3
1500	89.2	89.8	91.4	92.6	94.7
2000	91.2	91.2	92.6	93.6	95.2
2500	91.5	92.2	93.6	94.7	95.7

The sensitivity analysis in Table III demonstrates that the proposed AVMT-NN technique has a greater performance ratio. By definition, confusion matrix defends classification with genuine negative split by false-positive values.Specificity $= \frac{\text{TN}}{\text{TN+FP}}$

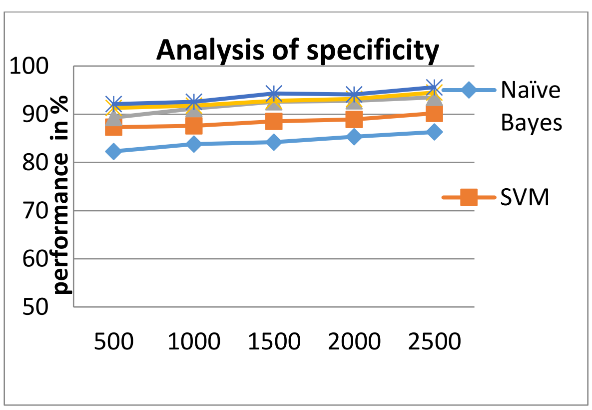

Fig 5. Impact of specificity

The projected AVMT-NN method has created higher-performing alternative ways, as seen in Figure 5. The figure compares the Specificity formed by the various approaches.

TABLE IV. IMPACT OF SPECIFICITY

Methods/Datasets	Impact of Specificity in %				
	Naïve Bayes	SVM	Dec-NN	SNC-MKFJC	AVMT-NN
500	82.3	87.3	89.3	91.3	92.1
1000	83.8	87.6	91.2	91.8	92.6
1500	84.2	88.5	92.6	92.8	94.3
2000	85.3	88.9	92.8	93.2	94.1
2500	86.3	90.2	93.5	94.5	95.6

Table IV compares the different measures of Specificity applied to different datasets, illustrating how the harmonic representation posed by true positives and false negatives relies on the accuracy and recall rates of the respective approaches. When calculating the rate of incorrect extraction, the ratio of unclassified regions to similarly-classified ones is used. (Fer) = $\sum_{k=0}^{k=n} \times \frac{TotalDataset\ FailedtoClassify\ (Fer)}{TotalnoofData\ (Fr)}$

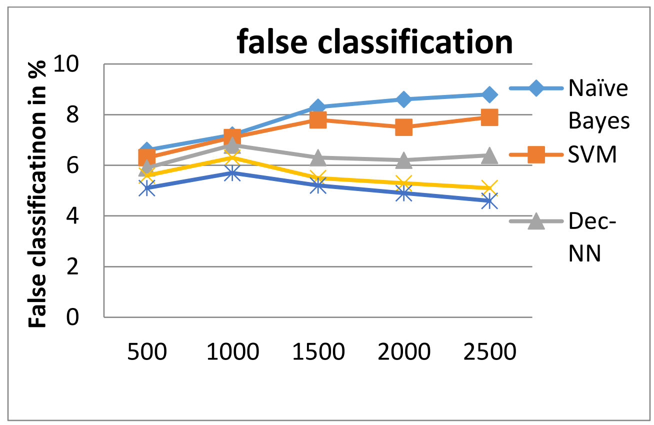

Fig 6. Impact of false classification

Look at Figure 6 to see that the projected technique has produced more accurate classifications than the rest of the methods combined.

TABLE V. IMPACT OF FALSE CLASSIFICATION

Methods/Datasets	Comparison of False Classification in %				
	Naïve Bayes	SVM	Dec-NN	SNC-MKFJC	AVMT-NN
500	6.6	6.3	5.9	5.6	5.1
1000	7.2	7.1	6.8	6.3	5.7
1500	8.3	7.8	6.3	5.5	5.2
2000	8.6	7.5	6.2	5.3	4.9
2500	8.8	7.9	6.4	5.1	4.6

Table V demonstrates the contrast of the false classification ratio and it shows that the proposed approach produces less false classification ratio. Time complexity (Tc) $=\sum_{k=0}^{k=n} \times \frac{\text{Total Features Handeled to Process in Dataset}}{\text{Time Taken(Ts)}}$

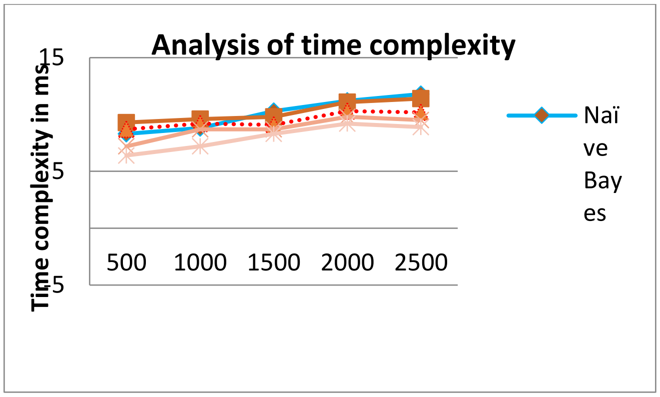

Fig 7. Impact of time complexity

Figure 7 demonstrates the contrast of time difficulty produced by different methods and shows that the proposed approach has produced less time complexity than other methods.

TABLE VI. IMPACT OF TIME COMPLEXITY

Methods/Datasets	Impact of Time Complexity in Milliseconds (ms)				
	Naïve Bayes	SVM	Dec-NN	SNC-MKFJC	AVMT-NN
500	8.3	9.3	8.7	7.2	6.4
1000	8.8	9.6	9.2	8.7	7.2
1500	10.3	9.8	9.1	8.7	8.3
2000	11.2	11.1	10.3	9.8	9.2
2500	11.8	11.4	10.2	9.5	8.9

Table VI shows how different methods' temporal complexity assessments differ, with the projected method resulting in a lower overall evaluation. The time complexity is measured by how long it takes to completely load the dataset in order to do the feature selection and classification.

CONCLUSION

To conclude theta the proposed diabetic prediction based on the diabetic prediction handles the influence rate to increase the performance of the proposed system. This implies the diabetic principle threshold rate formulated to improve the influence affected category of the important feature selection

This proposed system produce high performance compared to the other system as well in sensitivity produce 96.2 %, specificity rate produce 96.1 %, and classification accuracy produce 96.8 % to predict early risk level for premature treatment recommendation.

REFERENCES

[1] Eileen M. Weinheimer-Haus, Stefan Judex, William J. Ennis, Timothy J. Koh, “Low-intensity vibration improves Angiogenesis and wound healing in diabetic mice” 1-8 (2014)

[2] Caroline Oi-Ling Yu et al., “Low-magnitude high-frequency vibration accelerated the foot wound healing of n5-streptozotocin-induced diabetic rats by enhancing glucose transporter 4 and microcirculation” 1-12 (2017)

[3] R.A.Buda, M. Mohd.Addi, “A Portable non-invasive blood glucose monitoring device”, 964 – 969 (2014).

[4] T.Jayalakshmi, Dr.A.Santhakumaran, “A novel classification method for diagnosis of diabetes mellitus using Artificial Neural Network”, 159 – 162 (2010).

[5] Sierra-Sosa, D., Garcia-Zapirain, B., Castillo, C., Oleagordia, I., Nuno-Solinis, R., Urtaran-Laresgoiti, M., & Elmaghraby, A. (2019). Scalable Healthcare Assessment for Diabetic Patients using Deep Learning on Multiple GPUs. IEEE Transactions on Industrial Informatics, 1–1. doi:10.1109/tii.2019.2919168

[6] M. L., ALVA, et al. "The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84)," Diabetic Medicine, vol. 32, pp. 459-466, 2015

[7] Edurne, ALONSO-MORÁN, et al. "Prevalence and quality of care indicators of type 2 diabetes in the population of the Basque Country (Spain)," Avances en Diabetología, vol. 31, pp. 72-79, 2015

[8] Guo-li DU, et al. "Metabolic Risk Factors of Type 2 Diabetes Mellitus and Correlated Glycemic Control/Complications: A Cross-Sectional Study between Rural and Urban Uygur Residents in Xinjiang Uygur Autonomous Region," PloS one, vol. 11, 2016.

[9] A, A. B., & S, V. (2019). Enhanced optimal insulin regulation in post-operative diabetic patients: an adaptive cascade control compensation based approach with diabetic and hypertension. IEEE Access, 1–1. doi:10.1109/access.2019.2927248

[10] Sunil Kumar K N and Dr. Shivashankar, “Compression of PPG Signal through Joint Technique of Auto-encoder and Feature Selection”, International Journal of Health Care Information Systems and Informatics-ESCI Indexed ACM-Digital Library, Vol. 16, Issue-4, Article-23, ISSN:1555-3396, DOI:10.4018/IJHISI.

[11] Sunil Kumar K N and Dr. Shivashankar, “Bio-signals Compression Using Auto Encoder”, Journal of Electrical and Computer Engineering (Q2 Indexed), Institute of Advanced Engineering and Science publishers, ISSN: 2088-8708, Vol.11, No.1, pp. 424-433, http://doi.org/10.11591/ijece.v11i1.pp424-433. 2021.

[12] Sunil Kumar K N and Dr. Shivashankar, “Security Framework for Physiological Signals using Auto Encoder”, Journal of Advanced Research in Dynamical and Control Systems–(Q3 Indexed), Vol. 12, 01-Special Issue, pp. 583-592, https://doi.org/10.5373/jardcs/v12sp1/20201107, 2020.

[13] Sunil Kumar K N and Dr. Shivashankar, “Improving AODV Route Recovery Mechanism using PSO in WSN”, International Journal of Sensors, Wireless Communications and Control (Q4 Indexed), Bentham Science Publishers, ISSN: 2210-3287. Vol. 10, No. 10, https://doi.org/10.2174/221032790966619112611O 201, 2020.

[14] Khodakaramzadeh, S., Batmani, Y., Meskin, N.: ‘Automatic blood glucose control for type 1 diabetes: A trade-off between postprandial hyperglycemia and hypoglycemia’, Biomed. Signal Proc. Control, 2019, 54, p. 101603

[15] Lomte, R., Dagale, S., Bhosale, S., & Ghodake, S. (2018). Survey of Different Feature Selection Algorithms for Diabetes Mellitus Prediction. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). doi:10.1109/iccubea.2018.8697452

[16] Ansari, M. F., Sharma, P. K., & Dash, B. (2022). Prevention of phishing attacks using AI-based Cybersecurity Awareness Training. International Journal of Smart Sensor and Adhoc Network., 61–72. https://doi.org/10.47893/ijssan.2022.1221

[17] Dash, B., Ansari, M. F., & Sharma, P. (2022a). FUTURE READY BANKING WITH SMART CONTRACTS-CBDC AND IMPACT ON THE INDIAN ECONOMY.

[18] Ansari, M. F., Dash, B., Sharma, P., & Yathiraju, N. (2022a). The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review. International Journal of Advanced Research in Computer and Communication Engineering, 11(9), 81-90.

[19] Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022b). THREATS AND OPPORTUNITIES WITH AI-BASED CYBER SECURITY INTRUSION DETECTION: A Review. International Journal of Software Engineering & Applications, 13(5), 13-21.

[20] Dash, B., Sharma, P., Ansari, M. F., & Swayamsiddha, S. (2022). A review of ONDC's digital warfare in India taking on the e-commerce giants. International Journal of Engineering & Technology, 11(2), 96-99.