C. R. Rao. et.al., Vol 1(1), pp 50-60
https://doi.org/10.55306/CJIESN.2025.010105
e-ISSN: XXXX.XXXX
Ci-STEM Journal of Intelligent Engineering Systems and Networks

Diabetes Prediction Using Different Classification Algorithms on Data Mining – A Survey

Raja Rao Chatla
Board of Practical Training Eastern Region, Kolkata, India.
email: crrao@bopter.gov.in
Soumitra Kumar Mandal
Department of Electrical Engineering
National Institute of Technical Teachers Training and Research, Kolkata, India
email: skmandal@nitttrkol.ac.in

Open Access
Review Article

Received : 25/04/2025
Accepted : 13/08/2025
Published : 26/10/2025
Corresponding author email:
crrao@bopter.gov.in
Citation:
C. R. Rao., et. al., "Diabetes Prediction Using Different Classification Algorithms on Data Mining – A Survey," Ci-STEM Journal of Intelligent Engineering Systems and NetworksDigital Technologies and Expert Systems, Vol. 1(1), pp. 50-60, 2025, doi: 10.55306/CJIESN.2025.010105
Copyright:
***©2025 C. R. Rao**. et. al.,*
This is an open-access article distributed under the terms of the Creative Commons Attribution License which grants the right to use, distribute, and reproduce the material in any medium, provided that proper attribution is given to the original author and source, in accordance with the terms outlined by the license.
(https://creativecommons.org/licenses/by/4.0/).
Published By:
Ci-STEM Global Services Foundation, India.

Abstract:
Healthcare generates volume of data daily in different formats such as notes, reports, images and a numbers pool and so forth. But there is no scientific instrument in medicine to study this information. The data from this point on may be mined for information that can be utilized by media experts to predict future steps in the process. Cardiopathy is the most common cause of death for the general population. Early identification and risk perception is essential for the patients' drugs and analysis of specialists. Data mining is a process that extracts information from the collected data and structures it for further use. In the present study, we pay attention to such medical decision learning design based on diabetes data and establish a smart therapeutic choice emotional supporting network for doctors. The primary objective of this study is to develop a intelligent diabetic disease prediction for analyzing diabetes malady by using database of diabetes patients. Health data are by its nature unpredictable and always changing, which makes it extremely hard to deal with. In order to tackle the challenges mentioned above, some studies have presented a range of ML approaches for detecting and prognosis of illness. This article compares a number of diabetes prediction models to find an approach for diagnosing the disease. The research aims to shed light on different methods of diagnosis for the disease, enabling patients to get treated more quickly. The prediction of the product, glucose level in blood is also predicted with advance technology, Variety of Machine Learning techniques e.g., Neural Network (NN), SVM classifier, Data mining Techniques etc which are used here for predicting result. The study hopes to find a faster and more efficient way of diagnosing the condition, so that patients can receive treatment earlier.
Keywords: *Data Mining, Intelligent Diabetic Disease Prediction, Machine Learning.*

1. INTRODUCTION

Diabetes refers to any one of a large number of disease conditions where too much glucose, frequently referred to diagnostically as, diagnostic glucose occurs within the body. The primary source of energy is blood sugar derived from the food we eat. Diabetes occurs when the pancreas does not produce enough of the hormone insulin, experts say. Insulin hormone which pushes sugar from the blood into various cells to be used as energy. Due to lacking the necessary amount of insulin, it disrupts the body's internal ability to produce and utilize insulin in a proper way. Therefore, lots of glucose is lost in the urine. Data mining is the process of extracting knowledge from data to make it understandable by humans. It's a way of sifting through vast amounts of information. This Data mining system essentially is an effective use of information systems in industries such as, banking, educational sector and many other functional areas like stock market etc. By using different methods (approaches) of data mining, one can apply data mining in the medical domain to predict disease effectively. 2. Prediction and description are two main goals of data mining. Prediction is estimation under prediction of uncertain or future quantities, on the basis of some parameters or quantities from a data collection. The purpose of description is to reveal patterns in development that are understandable to people. A fundamental

knowledge of growth and associated foreign diabetes related variables is essential prior to developing predictive models.

The aim is to predict the diabetes (ordering of sick people by priority) and how it works. Datamining tools could be used to help identify an early diagnosis of disease saving lives and reducing treatment costs. 382 million people in the world have diabetes, according to the International Diabetes Federation. It will reach 592 million by 2035. This study also investigates application of several DM approaches to predict the diabetes. Some of the basic data mining techniques such as classification, clustering and predictions are the most important and popular ones that have been analyzed to predict disease diabetes. Mining tools are utilized in diverse sectors including education and banking.

Etymology Diabetes mellitus is a disease taken from the greek diabetes meaning siphon - to pass through and the latin mellitus meaning honeyed or sweet. Diabetes is considered a disease in which too much sugar circulates in the blood and urin. Diabetes occurs when glucose is too high in people. Humans derive the most energy consumption from glucose-rich meals. Insulin, which is released by the pancreas, helps glucose from food to enter cells where it can be used as energy. The most prevalent type of diabetes are type 1 and 2, along with diabetes mellitus. Body has no ability to produce insulin, with type 1 diabetes. The immune system attacks and kills the cells in the pancreas that generate insulin. Type 1 diabetes most often develops in children and young adults, though it can start at any age. Type 1 diabetics need daily insulin to live. The body doesn’t produce or use insulin effectively in type 2 diabetes. Type 2 diabetes can befall anyone, young or old. It is most common among people in their 40s or 50s. Type 2 diabetes is the most common form. There is also gestational diabetes, which some women experience during their pregnancy. Most of the times, this type of diabetes is cured that once the baby is delivered. You have a higher risk of type 2 diabetes later and also immediately after pregnancy if you developed gestational diabetes. It may be that diabetes diagnosed in pregnancy really is type 2. One in four people over 65 has diabetes. Type 2 diabetes is the form found in adults and it makes up 90 to 95 percent of all cases.

Data mining is described as the process of discovering and extracting useful information from a large quantity of raw data. Basically, obtained data from Data Mining is used for predicting the clandestine patterns, future trends, behaviours and also enables people to make inferences. In essence, data mining is quite a bit like helping decision makers analyse data from many angles (or dimensions, viewpoints) and summarise them into useful categories. Classification and prediction in Data Mining is increasingly popular with growth of daily incoming data, that requires a proper framework for processing and organizing it. In this field, it's key to use different combinations of clustering methods together with classification algorithms. Classification and clustering are used to group objects into one or more classes based on a characteristic. Classification is the task of assigning input instances to given class labels; and Clustering (without supervision) is the process of forming groups based on similarity of occurrences without any supervisory information. k-means is a popular clustering method and Nave Bayes combined with Support Vector Machine are among the widely used classifying algorithms.

Diabetes is a chronic condition that happens when the body fails to produce insulin, a hormone that helps control blood sugar. The damaging effects of the condition dealt to different organs, nerves and blood vessels can be slow-acting and destructive. Failure to diagnose the symptoms may result in the patient's life being placed at risk. Therefore, some work on the prediction of illness has been carried out. Neural Networks (NN) and SVM classifiers [4] & data mining techniques are some of the machine learning methods employed for prediction of diabetes like symptoms. Synthetically, a NN is a computational model with multiple processing elements which take input(s) and give output(s) by their action (Note: EANs are not limited to that behaviour). We use such a vectors of features and train a Neural Network (NN) to predict future blood sugar values. Classification methods are really a powerful and common means for classification in many real-world applications, such as diabetes patients’ detection. For the discrimination of diabetic patients, different methods such as SVM and Naive Byes are presented. The full form is supervised machine learning, which can be employed to resolve classification problems.

And, in order to predict the change of the range value, we perform classification with an hyper-plane separating both classes. The Naive bayes problem, or algorithms that classified an object or instance under observation to a one of several pre-specified set using attribute values, it is also a well known

model of classification problem. They intend to calculate the conditional posterior distribution and then predict in a probabilistic sense into the more likely class. Data mining is the process of finding usable information by parsing through huge amounts of data. Random Forest and Decision Tree are two of the most important and popular data mining techniques. RF is a universal machine learning method for predicting and estimating the regression part. The classification cases are being classified by means of decision tree. It can also be thought of as a set of if-then rules in feature and class space, or conditional probability distributions. The main objective of this study is to compare various methods for predicting diabetes more accurate and efficient.

2. RELATED WORKS

2.1 Diabetes

Diabetes is a disease that occurs when your blood sugar, also known as blood glucose, is too high. Yang et al.,The brain and other organs cannot tolerate anemia The main source of energy (glucose) in the blood is derived from the meat we eat. Diabetes occurs when the pancreas does not make enough of the hormone insulin, experts say. Insulin that moves sugar from the blood into cells, where it is used as a source of energy. With lack of insulin, functioning of the body's own ability to produce and use insulin is interrupted. Thus, excess glucose is excreted in the urine. Untreated diabetes can cause liver failure, cardiovascular disease and other organs affected in the long term.

Kumari et al, suggested that following factors were used to ascertain the diabetes and it was classified in positive and negative category (age, insulin, cigarette smoking, starting smoking age etc). Diabetes is a chronic condition that impacts millions of people world-wide. - One of the most important blood hormones is insulin, which is secreted by the pancreas and plays a significant role in regulating glucose levels. Insulin shots, a healthy diet and regular exercise can all help to manage diabetes. Blindness, high blood pressure, heart disease, kidney disease and other complications can result from diabetes. There are four types of diabetes.

Type 1 Diabetes: When the pancreas is not able to produce insulin, this disease develops. Insulin is the hormone that the pancreas produces. Diabetes type 1 can strike at any age. It will most frequently affect children and teenagers.

Type 2 Diabetes: It occurs when there is not enough insulin to meet the body's needs. Obesity increases the risk of type 2 diabetes as a result of family history, old age and obesity. This is most common among those over 40 years of age.

(i) Gestational Diabetes: It's the third most common kind, and it mostly affects pregnant women owing to high blood sugar levels in the body.

(ii) Pregestational Diabetes: Pregnant women with insulin-dependent diabetes are said to have pregestational diabetes. Diabetes Prediction is a crucial part of the data mining process. For diabetes prediction, a variety of data mining techniques are used. These would be the data mining strategies that have been used to forecast diabetes, as listed below.

2.2 Functional Requirement

Theis et al., gives a functional requirement specifies what the system should be able to perform. A functional requirement also describes the operational activities a system must be capable of. Specifications of data to be loaded into memory should be included in functional requirements. Description of the system's process, as well as reporting and other outputs, the following is some of the proposed system's functional requirements: i. the developed scheme will also provide a platform for analysing datasets for new patients. ii. The suggested system will check the correctness of the dataset.

3. DATA MINING UNDER HEALTHCARE

Data mining is a process by which useful information is obtained from large data sets, and it has become more significant in the health-care industry. Khan et al., describes that diabetes might benefit from data mining techniques because they can uncover hidden knowledge from large amounts of diabetes-related data. Data mining is a technique for extracting usable data and patterns from large datasets as shown in fig 1.

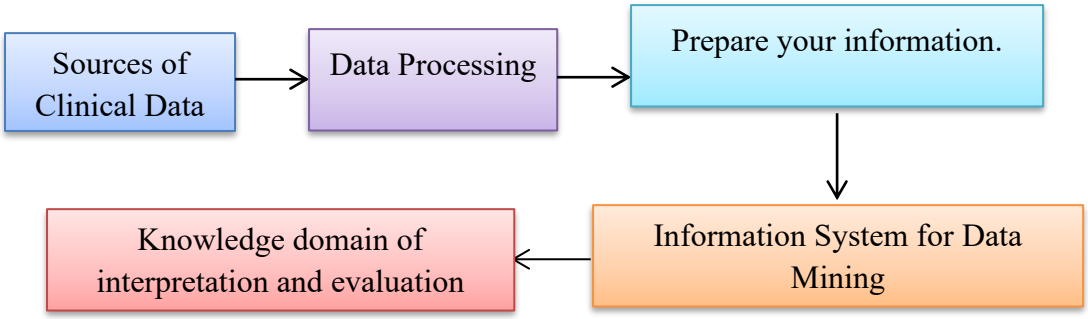

Figure 1: Steps in the progress of clinical data mining.

Ahmad et al., proposed the amount of datasets in health care, in particular, is enormous and changing, making statistical predictions difficult. Many data mining strategies have been proposed in the clinical sector. In the field of clinical data mining, the most common and widely used approaches include statistical analysis, prediction and classification. Given the large volume of health care data from various sources, it is increasingly essential to develop techniques with more robust characteristics for analysis, interpretation and decision-making.

Data cleansing, also known as data cleansing, is the act of identifying and removing noisy, unnecessary, and missing elements from the collected data gives by kalagotla et al. The next step in the cleaning process is data integration, which involves combining data from many sources into a single source format. Collecting unknown and secret predictive intelligence from large dynamic databases is known as clinical data mining. It is recognized as a crucial and novel technique with enormous potential for focusing on the most significant data inside the clinical data sets depicted by Khanam et al. This study examines by Sourij et al., that the different data mining methodologies and tools for managing clinical operations. Data mining may greatly aid diabetes research and, as a result, increase the quality of diabetic patient treatment. The nontrivial extract of unknown new and possibly relevant information from databases and data is referred to as this procedure. The basic phases in the information mining procedure utilising clinical data. Clinical Data Sources: Clinical data is gathered from a variety of sources, including sensors, web repositories, and manually generated synthetic datasets.

4. CLASSIFICATION

A method used to predict diabetes is classification and suggested by Bhoi et al. The biggest data mining task is classification. Classification is common in large amounts of corporate and medical data sets. Classification is a data-mining function that divides a collection of things into desired categories. Classification is the process of dividing situations into groups based on a known characteristic. Each case has a set of attributes, the first being the class attribute, which is also known as the predictable attribute. Creating a model that defines the attribute as both a function of applied attributes is part of the work as shown in fig 2.

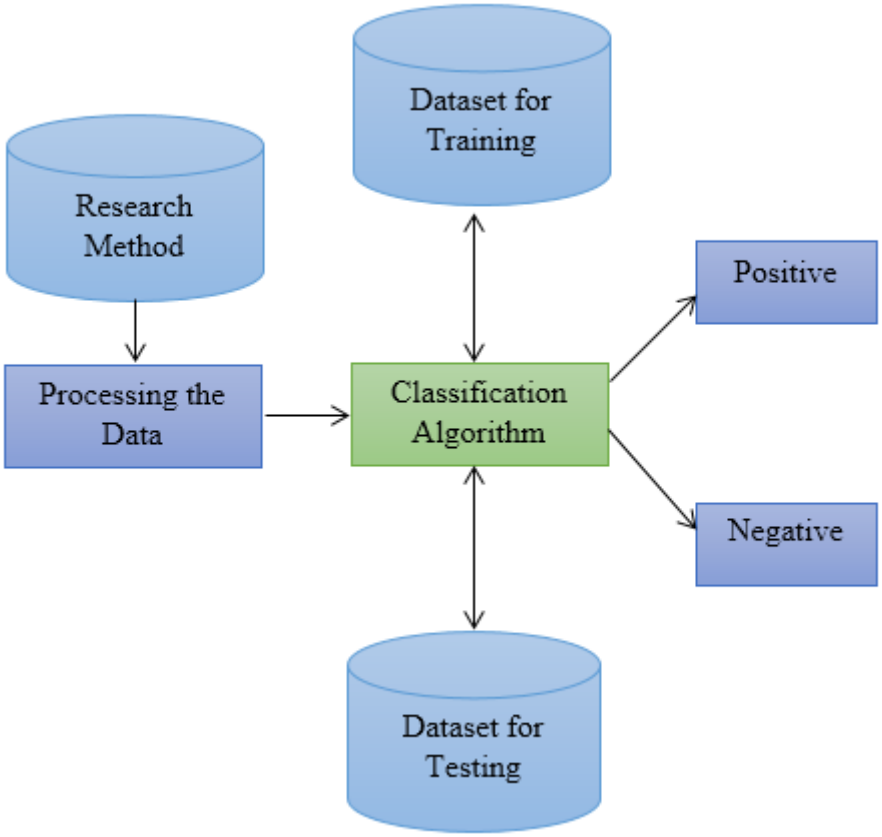

Figure 2: Classification Algorithm for Predicting Diabetes

Classification in tools is concerned with identifying the problem by differentiating illness characteristics amongst patients and diagnosing or predicting which algorithm performs best suggested by Azad et al. Classification's fundamental goal is to accurately anticipate the classifier for every example in the data. And grouping them together based on their commonalities. Segmentation is another term for clustering. It's used to find unprocessed case groups based on the set of criteria. Cases with more or less similar attribute values inside the same category. Clustering is a data mining job that is done on its own. This means that the training process is not modelled using just one characteristic. All input characteristics are handled equally. Before clustering, the attribute values must be codified to prevent higher-value attributes from impacting lower-value attributes.

4.1 Naive Bayes

Naive Bayes uses the Bayes theorem, that makes assumption between predicators that is tested by He et al. Its so easy to use that it helps you develop models fast both in term of its predictive ability and a provides fresh approach towards perceiving and understanding your data. This technique can be adapted for a predictive model (when using Naive Bayes to develop the predictive model). For this to be effective though, all the input features need to similarly be independent. Feature 2: NB Based on the feature of categorical sequence composition, an easy construction and no complex iterative parameter for Naive Bayesian system can get. Naive Bayes might be a good predictor. For extremely large sets of data, this approach is highly favorable. On features before and after dimension reduction, NB gives the same construction time. Naive Bayes visualizations were easy to understand.

4.2 Decision tree

Decision tree is the most common method used for data mining. The general and simple approach for constructing a decision tree is one of the most important learners. A prediction model called a decision tree is employed in data mining. In the proposed investigation made by Abdollahi et al., prediction of the disease is performed based on patient data in which a decision tree and classification algorithm are used. Decision trees are easy to construct and comprehend. Decision tree-based prediction is well-

ordered. It can process a great amount of extent data. It is more suitable for knowledge discovery searches. Finally, results will be simple to understand and can be interpreted.

4.3 Knowledge discovery database

The process of obtaining knowledge and information from a set of data is known as knowledge discovery in databases (KDD). It is sometimes called data extraction. It involves processing, selecting and purifying the data, as well as understanding the datasets and determining the appropriate solutions from the observed results which provides by Aggarwal et al. Data mining is indeed a crucial phase in the KDD process. It's used to extract information from a database. Data integration, information retrieval recognition, and knowledge presentation are all part of the KDD process.

4.4 Regression

A common data mining function is regression. Classification and regression are two terms that are used interchangeably. Regression is a technique for predicting unending quantities such as age, weight, temperature, and sickness. Regression methods may be used to predict all of these. Gupta et al., proposed many medical domain issues may be solved using regression tasks. The most often used regression methods are regression analysis and logistic regression.

4.5 Association

One of the most important data mining functions is association, which invents the likelihood of the elements in a collection. Associative rules are used to show the links between coexisting as well as the objects which is proposed by Gupta et al. Marketplace analysis is another association name. Every value, or more generally every attribute or combination of values, is considered as an element in terms of association. The work of association has two objectives: finding itemsets and discovering association rules. Association modelling has become a common method in a variety of fields. Bioinformatics, medical diagnostics, Web mining, and scientific data analysis are some of the additional application industries that are linked to association. The most extensive links are recognised by reviewing data for repeating patterns and utilising the measurement for strong support to build association rules. The frequency of things viewable in the database is represented by support. The number of incidents the propositions have been determined to be true is represented by confidence.

4.6 Neural Network [NN]:

A Neural Network (NN) is a mathematical model of the human brain architecture that reflects its "learning" and "generalisation" capacities. As a result, NNs fall under the AI umbrella. NNs are commonly used techniques because they can represent highly nonlinear systems with uncertain or complicated relationships between variables.

A NN consists of several layers, i.e., an input nodes, an output layer or one convolutional units (as suggested by Wu et al. The neurons of hidden layers are computing the seven equations throughout this research, and the input layer of output neurons could be linear summation. The current G (t) and the last blood sugar G (t-N*PH) are network inputs as shown in Table 1. The weights and the network biases are randomly “skipped” according to Levenberg-Marquardt drive. Actual and predicted variations of blood glucose is backs propagated by each layer of NN that make the best weight for minimum error. We have also used our network ($N \approx 193$, $H > N$) to Stochastic Resonance: neural-like networks for binary patterns Table 2. It is the aim of NN to predict future readings based on previous Nos. The estimated value will then be added with the past N-1 values for futurics prediction and so on Since the activation function of each neuron is non-linear, therefore these estimates were gradually, adaptively and 3.4 Non-linear predicted values of -NN The predicted instead of previous_linear.

Table 1: Parameters that were utilized to make the forecast

Sl. No	Different Factors	Summary	Values that are permitted
(i)	Years	The Subject's Age	Number that is Discrete
(ii)	Insulin should be taken.	Take any medication or injections that can help you avoid developing diabetes.	Y (or) N
(iii)	About Smoke	Whether or whether the subject is a cigarette smoker	Y (or) N
(iv)	When did you first start smoking?	The subject's age determines whether or not he or she smokes.	number that is discrete

4.7 Application of Machine Learning in Disease Prediction

Pandey et al., are concentrating their machine learning efforts on three disorders. Breast cancer is a relatively frequent condition among women, as are heart disorders, which are the main cause of mortality in the United States, and diabetes, which is characterised by high blood glucose as well as blood sugar levels. The datasets are available for download through the UCI machine learning community. There are four phases in the suggested technique. Exploration of the dataset, Dataset Collection, Feature Selection, and Model Fitting and Validation are the steps involved. The initial phase entails a python environment examination of the dataset. It tries to find out whether there is a missing value in the Data Mining process. If any values are absent, the mean value is used in the case of continuous data and the mode value in the case of categorical values. The purpose of feature selection is to account for multi-collinearity and to eliminate any redundant characteristics that were already highly correlated with one another, hence increasing the model's performance. For feature selection, the backward selection approach is utilised. The characteristics with a p-value larger than 0.05 were removed from the model and the remaining variables were re-fitted. This technique was repeated until each of the model's existing variables reached a meaningful level. The classification algorithm was given the specified characteristics as input. Algorithms for classification are Classification methods include logistic regression, making decision tree, random forest, example of potential, and adaptive boosting. The dataset was split into two parts: a training set and a testing set, with 90% of the data used for training and 10% for testing. Fazakis et al., suggested method's logistic regression achieves 87.1 percent prediction accuracy in the detecting for cardiovascular diseases, 85.71 % in the prevention of type 2 diabetes using the Help Vectors Model (regular kernels), as well as 98.57 percent in the Breast Cancer Classifier using AdaBoost.

4.8 Analysing Feature Importance for Diabetes Prediction using Machine Learning

The analysis of the diabetes predictive characteristic of the data set was proposed by Bukhari et al. The characteristics that are utilised to make predictions are by far the most important aspect of an algorithm, as well as certain features have a huge impact on the prediction. In a ratio of 67 percent train to 33 percent test, the dataset was partitioned into training set and testing. To confirm the presence of significantly associated traits, the relationship between each property was determined, and thresholds were established at 0.7. The important characteristics selected from association graphs were diabetes, BMI, age, insulin, and glucose. The collected characteristics were supplied as input to three classification algorithms: logistic regression, SVM, and random forest. The algorithms for diabetes prediction are Random Forest, which has an accuracy of around 84 percent.

4.9 Levels of Testing

Unit testing was used to test the programmes that make up the system. Program testing is another name which are proposed by Deng et al. This phase of testing relies on the modules as a separate entity. The goal of unit test is to ensure that the individual modules are operating properly. We initially used the code testing technique for unit testing, which looked at the logic of the application. All syntax problems and other issues were discovered throughout the development process. For this, we created a test case

that resulted in the programme or module executing every instruction, i.e., every path through the programme was tested. Unit testing entails the exact specification of test cases, as well as the application of testing criteria and the administration of test cases.

User Input- In a user interface, data is entered via a graphical user interface (GUI) and then tested. Each element is checked for valid and incorrect data ranges. Error Handling- We attempted to handle any errors that happened when executing the GUI forms in this system. Reading an empty record and showing a compiler warning are two typical issues we've seen.

System Testing- Once we're sure that all of the modules are working well and that there are no issues, we look at how the system will operate or perform once all of the modules have been assembled. The primary goal is to identify inconsistencies between the system and also its original goal, current requirements, and system documentation. Analysts look for moulds that were created with various standards, which might lead to incompatibility. At this point, the system is being checked to determine that all of the user's needs are met. At this phase, various layers of testing are carried out to guarantee that the network is fault-free. Testing is fundamental to the success of the system. Testing the system requires that all components of a system are functioning properly. The system was first provided to the customer for data entry, with validation supplied at each level to ensure that the user did not enter irrelevant data. The user receives instruction on how to enter data. During the implementation of the system, it was discovered that the user was first resistant to the change; nevertheless, because the system was urgent and user-friendly, the fear was resolved. Ahead of the actual day-to-day transaction, entering live data from previous months' records was a touch tiresome. The greatest test of the system was to see if it produced the desired results. All of the outputs were double-checked and determined to be correct. Before the acceptance test, feedback sessions were held, and the user's proposed adjustments were implemented. Finally, the system is accepted and put into operation with real-time data. System tests are used to ensure that a fully created system fits its requirements by validating it.

Table 2: Table of accuracy comparisons

Designs	Approximation in Percent
DT	84.9
NN	83.5
NB	80.3
Bagging	83.2
Boosting	83.4
RF	84.3
SVM	86.6
SVM+	93.3
Fuzzy	91.2
ABC	90.2
PCA	89.8
Genetic	87.5
FW-SVM	82.6
Modified ABC	80.3

The table 2 displays the percentage of accuracy attained by each algorithm; the genetic as well as customized ABC algorithms have better accuracy values than the others as shown in fig 3.

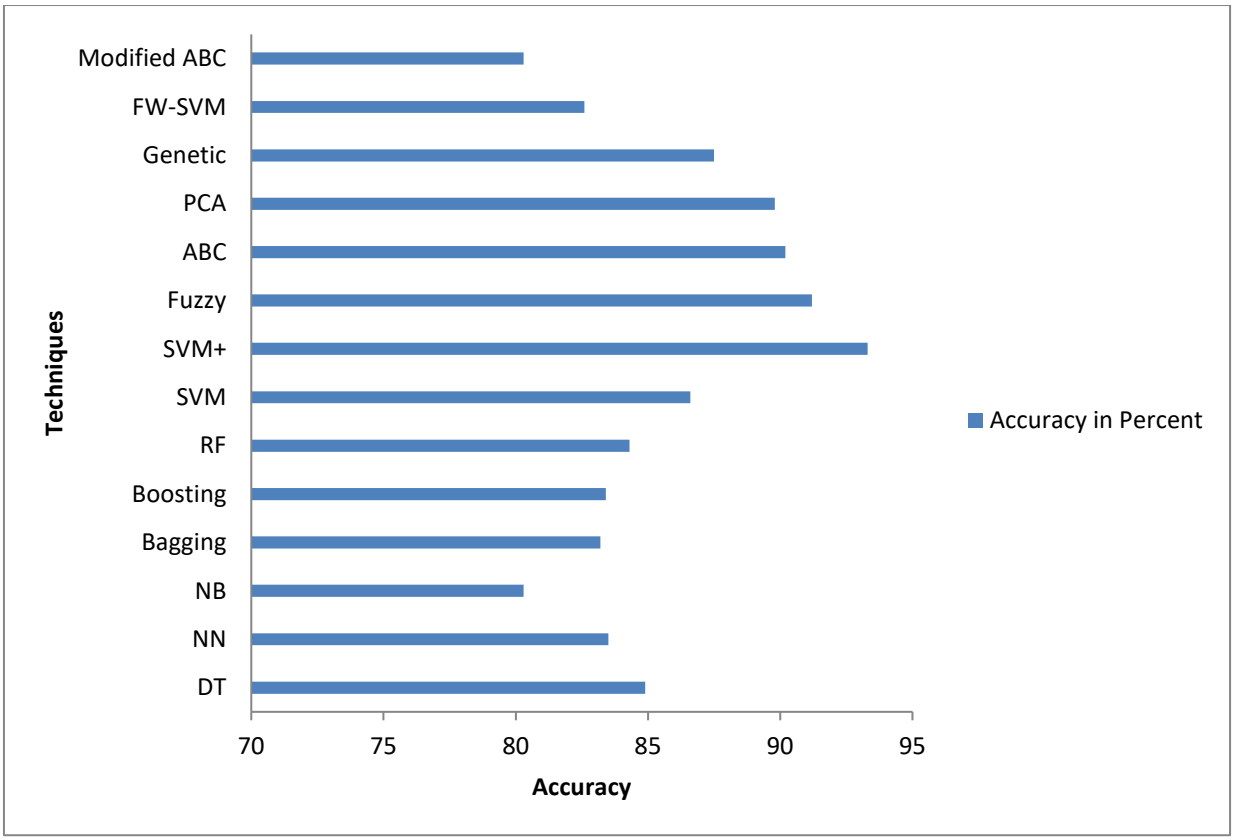

Figure 3: Accuracy comparisons for Different Algorithms

Existing techniques and tools for the diagnosis and prognosis of diabetes are supervised procedures that require increasingly specific training samples. Training images should be collected and included in the types of test data, which implies that the classifier should have been taken into account. As previously stated, correct training samples are used to determine classification accuracy. However, certain strategies have been developed to uncover hidden characteristics in a dataset that has not been thoroughly examined. As a result, complete training samples are required for a successful classifier. To address data gathering difficulties, a few research employed semi-supervised classification and unsupervised learning processes. This may help with the supervised learning issue, but it also creates a problem with accuracy and class imbalance. The primary goal of a classifier is to minimise iterations and over-fitting issues. Such problems may affect the tree-based method. We compared many algorithms and strategies for dealing with clinical datasets in this study. While comparing the methodologies, various flaws were discovered and identified, along with a set of undesirable characteristics.

5. CONCLUSION

In the present trend, clinical activity analysis is quite significant. The most essential challenge in a real-time environment is detecting and analysing clinical activity, because of lack of qualified sample and appropriate data makes these operations extremely difficult. Effective data mining tools and procedures may be used to accomplish this clinical data analysis. Diagnosing and pronging diabetes mellitus can be done in a variety of ways. This review examines several data mining strategies for resolving the diabetic condition diagnostic challenge. We identify various issues and findings in the clinical datasets handling procedure as a result of the investigation. The many strategies employed in the medical field diabetes prediction are addressed in this survey study. Diabetes was predicted using neural network. These systems are distributed in many ways and can be deployed in parallel. Among the most efficient and extensively used algorithms for diagnosis and detection is the decision tree as well as Random Forest. The models that do better in classifying diabetes individuals. With an accuracy of 76.30 percent, the Bayesian Classification algorithms findings indicate the suitability of the

constructed system. SVM is effective for a diabetic classification model that generates good classification results, indicating that applying the information science technique is feasible.

DECLARATIONS:

Acknowledgments	:	Not applicable.
Conflict of Interest	:	The author declares that there is no actual or potential conflict of interest about this article.
Consent to Publish	:	The authors agree to publish the paper in the Global Research Journal of Social Sciences and Management.
Ethical Approval	:	Not applicable.
Funding	:	Author claims no funding was received.
Author Contribution	:	Both the authors confirms their responsibility for the study, conception, design, data collection, and manuscript preparation.
Data Availability Statement	:	The data presented in this study are available upon request from the corresponding author.

REFERENCES

[1] Abdollahi, J., & Nouri-Moghaddam, B. (2021). Hybrid stacked ensemble combined with genetic algorithms for Prediction of Diabetes. *arXiv preprint arXiv:2103.08186*.

[2] Aggarwal, Y., Das, J., Mazumder, P. M., Kumar, R., & Sinha, R. K. (2021). Heart rate variability time domain features in automated prediction of diabetes in rat. *Physical and Engineering Sciences in Medicine*, *44*(1), 45-52.

[3] Ahmad, H. F., Mukhtar, H., Alaqail, H., Seliaman, M., & Alhumam, A. (2021). Investigating health-related features and their impact on the prediction of diabetes using machine learning. *Applied Sciences*, *11*(3), 1173.

[4] Azad, C., Bhushan, B., Sharma, R., Shankar, A., Singh, K. K., & Khamparia, A. (2021). Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus. *Multimedia Systems*, 1-19.

[5] Bhoi, S. K. (2021). Prediction of diabetes in females of pima Indian heritage: a complete supervised learning approach. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(10), 3074-3084.

[6] Bukhari, M. M., Alkhamees, B. F., Hussain, S., Gumaei, A., Assiri, A., & Ullah, S. S. (2021). An improved artificial neural network model for effective diabetes prediction. *Complexity*, *2021*.

[7] Deng, Y., Lu, L., Aponte, L., Angelidi, A. M., Novak, V., Karniadakis, G. E., & Mantzoros, C. S. (2021). Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. *NPJ Digital Medicine*, *4*(1), 1-13.

[8] Fazakis, N., Kocsis, O., Dritsas, E., Alexiou, S., Fakotakis, N., & Moustakas, K. (2021). Machine learning tools for long-term type 2 diabetes risk prediction. *IEEE Access*, *9*, 103737-103757.

[9] Gupta, H., Varshney, H., Sharma, T. K., Pachauri, N., & Verma, O. P. (2021). Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. *Complex & Intelligent Systems*, 1-15.

[10] He, Y., Lakhani, C. M., Rasooly, D., Manrai, A. K., Tzoulaki, I., & Patel, C. J. (2021). Comparisons of polyexposure, polygenic, and clinical risk scores in risk prediction of type 2 diabetes. *Diabetes Care*, *44*(4), 935-943.

[11] Jaiswal, V., Negi, A., & Pal, T. (2021). A review on current advances in machine learning based diabetes prediction. *Primary Care Diabetes*, *15*(3), 435-443.

[12] Joglekar, M. V., Wong, W. K., Ema, F. K., Georgiou, H. M., Shub, A., Hardikar, A. A., & Lappas, M. (2021). Postpartum circulating microRNA enhances prediction of future type 2 diabetes in women with previous gestational diabetes. *Diabetologia*, *64*(7), 1516-1526.

[13] Kalagotla, S. K., Gangashetty, S. V., & Giridhar, K. (2021). A novel stacking technique for prediction of diabetes. *Computers in Biology and Medicine*, *135*, 104554.

[14] Khan, F. A., Zeb, K., Alrakhami, M., Derhab, A., & Bukhari, S. A. C. (2021). Detection and prediction of diabetes using data mining: a comprehensive review. *IEEE Access*.
[15] Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. *ICT Express*, *7*(4), 432-439.
[16] Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, *2*, 40-46.
[17] Li, J., Chen, Q., Hu, X., Yuan, P., Cui, L., Tu, L., ... & Xu, J. (2021). Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques. *International Journal of Medical Informatics*, *149*, 104429.
[18] Pandey, A., Vaduganathan, M., Patel, K. V., Ayers, C., Ballantyne, C. M., Kosiborod, M. N., ... & Everett, B. M. (2021). Biomarker-based risk prediction of incident heart failure in pre-diabetes and diabetes. *Heart Failure*, *9*(3), 215-223.
[19] Sharma, A., Guleria, K., & Goyal, N. (2021). Prediction of Diabetes Disease Using Machine Learning Model. In *International Conference on Communication, Computing and Electronics Systems* (pp. 683-692). Springer, Singapore.
[20] Sourij, H., Aziz, F., Bräuer, A., Ciardi, C., Clodi, M., Fasching, P., ... & COVID-19 in diabetes in Austria study group. (2021). COVID-19 fatality prediction in people with diabetes and prediabetes using a simple score upon hospital admission. *Diabetes, Obesity and Metabolism*, *23*(2), 589-598.
[21] Theis, J., Galanter, W., Boyd, A., & Darabi, H. (2021). Improving the In-Hospital Mortality Prediction of Diabetes ICU Patients Using a Process Mining/Deep Learning Architecture. *IEEE Journal of Biomedical and Health Informatics*.
[22] Wu, Y. T., Zhang, C. J., Mol, B. W., Kawai, A., Li, C., Chen, L., ... & Huang, H. F. (2021). Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. *The Journal of Clinical Endocrinology & Metabolism*, *106*(3), e1191-e1205.
[23] Yang, H., Luo, Y., Ren, X., Wu, M., He, X., Peng, B., ... & Lin, H. (2021). Risk prediction of diabetes: big data mining with fusion of multifarious physical examination indicators. *Information Fusion*, *75*, 140-149.

Authors

Mr. C. Raja Rao obtained his B.E in ECE from Andhra University in 1995 and M.Tech in ECE from JNTU Hyderabad in 2001. He started his career in 1999 as Assistant Professor of Electronics and Communication Engineering Department with SVITS Mahbubnagar. He worked in various Engineering colleges and continued in teaching up to Dec 2009. In December 2009 he has been appointed as Deputy Director in Board of Practical Training (Eastern Region) Kolkata (An Autonomous body under Ministry of Education Government of India). He has over 10 years of teaching experience and about 15 years of experience in implementation of Apprentices Act 1961. He is also a life member of ISTE, IETE and IE(I). His research interest includes Signal and Image Processing, Low power VLSI design, Vocational and Technical Education and Training.

Dr. Soumitra Kumar Mandal has obtained his B.E. from Bengal Engineering College (Now IIEST), Shibpur, M.Tech from Institute of Technology, Banaras Hindu University, Varanasi and Ph.D. from Punjab University, Chandigarh all in Electrical Engineering. He started his career as Lecturer at SSGM Engineering College, Shegaon, and then moved to Punjab Engineering College, Chandigarh. In February 2004, he has been appointed as Assistant Professor of Electrical Engineering in National Institute of Technical Teachers' Training and Research (NITTTR), Kolkata. He is now serving as Professor of Electrical Engineering in the same institute. Throughout his academic career, he has published about 45 research papers in National and International Journals and presented many papers in conferences. He has also published 8 Textbooks for undergraduate and Post Graduate Students of Electrical Engineering. His research interests include Microprocessor and Microcontroller based System Design, Embedded System Design, Computer Controlled Drives, Neuro-fuzzy Computing, Signal Processing and VLSI design. He is also a life member of ISTE and a member of IE.