MANTECH Publications

Journal of Data Mining and Knowledge Engineering
Volume 7, Issue 2, July-December 2022

Methods for Diabetes Prediction Using Data Mining

Ujwal Rawat[1], Dr. S. K. Kachhawa[2]

Student[1], Professor[2]

Department of CSE

Matoshri College of Engineering & Research Centre

***Corresponding Author's Email:** -rrawatuujwal4@gmail.com[1]*

Abstract

Disease prediction in the healthcare business is greatly aided by data mining. Data mining techniques are often employed in early disease diagnosis. Diabetes is one of the world's most serious health issues. Diabetes is a common chronic illness. Diabetes prediction is a rapidly expanding science. Diabetes prediction in an early stage will lead to more effective treatment. Diabetes is vital to avoid, monitor, and raise diabetes awareness since it causes other health problems. Diabetes, whether type 1 or type 2, can cause heart problems, renal disease, and eye difficulties. This comprehensive study examines several techniques and data mining methodologies for predicting diverse diabetic problems at an early stage. Diabetes has caused it to become a chronic illness. Early detection of this condition will save patients' lives. Diabetes is averted and treatment rates are decreased by the use of data mining tools and methods. Among the most frequent and essential data mining technologies are association rule mining, classification, clustering, Random Forest, prediction, and the Artificial Neural Network (ANN). There are several data mining approaches accessible to avoid ailments such as heart disease, cancer, and renal disease, among others. This study looks at how data mining tools may be used to forecast numerous kinds of illnesses.

***Keywords**: Data Mining, Diabetes Diseases, Prediction of Diabetes, Data Mining Techniques*

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

INTRODUCTION

Data mining (DM) is a technique for creating and viewing information from data in a way that people can understand. It is a way of analysing large amounts of data. The implementation of data mining tactics in many areas, such as banking, education, and so on, is important in IT. [1]. DM may be effectively extended to illness prediction by utilising various data mining methodologies in the medical realm. DM appears to be expected and characterised by two primary goals. The prediction makes use of numerous variables or fields within the data set to anticipate important variables, either uncertain or potential values. The explanation focuses on the patterns in human perception. Before developing prediction models, it is critical to study the development and features of diabetes from outside sources. The objective is to employ data extraction methods to predict diabetes and characterise diabetes causes [2].

Diabetes, often known as high blood sugar, is a disorder that occurs when the blood glucose level in the body is too high. Diabetes, as previously stated, can result in serious cardiovascular issues. According to the WHO, there are 3.7 million deaths by the age of 70, with cardiovascular disease being the leading cause of death. Diabetes is mostly caused by uncontrolled blood glucose levels. A chronic ailment is a continuing sickness or condition with long-term repercussions. This sickness has a major negative impact on one's quality of life. Diabetes is one of the most serious and widespread diseases. This chronic condition is a leading cause of adult mortality globally. The expenditures are also linked to chronic illnesses. Chronic illnesses consume a considerable portion of the government and public budgets [3,4]. According to diabetes data, around 382 million individuals worldwide suffer from this condition [5].

Diabetes is a chronic, lifelong disease that affects the human body by reducing the amount of glucose-carrying insulin in the blood cells. This elevates blood sugar levels, which leads to a variety of problems, including stroke, heart illness, blindness, renal failure, and death [6]. Diabetes can be a chronic disease with no treatment if it insists on maintaining a normal blood glucose level without causing hypoglycemia. Diet, exercise, and the use of appropriate pharmaceutical medicines are all doable. Diabetes Mellitus occurs all around the world, but it is most prevalent in underdeveloped countries. The rise in rates in industrialised nations is the result of growing urbanisation, which

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

includes changes in lifestyle, such as a "Western" diet. This is due to a lack of comprehension. Diabetes diagnosis for quantitative analysis is a difficult task. Certain criteria, including A1c, fructosamine, white blood cell count, fibrinogen, and so on, are ineffective owing to low values. [7].

Data Mining Application in Healthcare Sector

Data mining is a relatively new notion for extracting knowledge from large amounts of data. Mining is the process of using and processing data to make judgments. The DM phase involves the investigation of models in big data sets using methodologies at the intersection of machine learning, statistics, and database systems. DM is an interdisciplinary subfield of computer science and statistics with the ultimate goal of extracting knowledge from a data set and converting it into a meaningful structure. Data mining has evolved to meet individual needs and can assist patients in recognising relationship patterns and making predictions based on pre-determined rules and programme requirements (Eapen, 2004). Data mining aids in the examination of patterns, including the classification of data through cluster studies, the identification of abnormal records, often known as anomalies, and the associated rules or dependencies. [8,9]. I Different DM techniques may be applied in the field of medical DM to properly anticipate illnesses. The proclivity of DM is to foresee and designate two primary objectives. The prediction employs numerous variables or fields from the data collection to anticipate such important variables' uncertain or probable values. The goal of the definition is to establish trends that detail human details. Before developing prediction models, it is critical to study the development and features of diabetes from outside sources. The objective is to employ data extraction methods to predict diabetes and characterise diabetes causes [10].

Diagnosis and Prediction of Diseases

In terms of social security providers, we must conclude and expect suffering, which is one of the most important reasons for using knowledge for social insurance. The use of human resources information-gathering has enabled professionals to better regulate their health. They cannot sit idle and make money by selecting a patient who will also harm the patient's health.

Ranking of Various Hospitals

In terms of social security providers, we must conclude and expect suffering, which

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

is one of the most important reasons for using knowledge for social insurance. The use of human resources information-gathering has enabled professionals to better regulate their health.

Better Treatment Techniques

Using data mining tools, the expert and patient will select the most appropriate care choice among all treatment plans. They will choose the appropriateness and cost of healthcare treatments. They can also identify medication interactions through information mining, lowering the danger for patients.

Effective Treatments

In comparison, medical expertise mining is being used to separate drug adequacy with components such as causes, signs, symptoms, and cost. For example, the effects of different medicines used to treat different patients who had a common illness were not Along these lines, we will mention whether the patient is well enough and whether the fees are reasonable.

Infection Control in Hospitals

Doctoral problems often include a large number of patients, and the degree of sedate-safe contamination is rather high. Contamination is evaluated through information mining to find such unforeseen situations in disease control information. A skilled individual investigates these examples further for contamination control. This identification method was refined to the point where information extraction procedures were altered at the University of Alabama in order to locate elusive examples of information from a primary source.

Identifying High-Risk Patients

Through administration, American Health techniques enable diabetic clinics to enhance efficiency and reduce diabetes patients' costs. American Health methods employed an unique monitoring methodology to differentiate between high- and low-risk individuals. Using an impactful display strategy, healthcare practitioners distinguished high-risk patients who required greater attention to their well-being.

Medical Device Industry

The social insurance sector cannot function without medical devices. Remote bio-sensors, whether portable or inexpensive, appear to be the most significant component of flexible medications, a safe strategy that emphasises patients' critical indications [11].

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

OVERVIEW OF DIABETES DISEASE

Diabetes is a chronic disease that affects people all over the world. It occurs when the body is unable to generate enough insulin. The pancreas produces insulin, one of the most important hormones in the body for maintaining glucose levels. Diabetes may be controlled with hormone injections, a nutritious diet, and regular exercise. Diabetes also causes blindness, high blood pressure, heart failure, and renal failure.

Diabetes is the most common illness. Diabetes is vital to avoid, monitor, and raise diabetes awareness since it causes other health problems. Diabetes is divided into two types: type 1 diabetes, which increases the need for insulin injections via therapy and vaccinations, and type 2 diabetes, which does not. Diabetes type 2 affects the majority of diabetic people. Diabetes was formerly a major worry among middle-aged individuals, but it now affects youngsters due to lifestyle changes. External stimuli that lead to the death of insulin-producing cells in the body are unavoidable in type 1 diabetes. Changing diets to obtain a specific physical weight will aid in the prevention of type 2 diabetes. Diabetes is a severe but treatable chronic disease. High blood glucose levels as a result of insulin deficiency, insulin action, or both. 1,2 1,2 Diabetes rates worldwide were 15.1 million in 2003, and it is anticipated to rise to 36.6 million by 2030. Adults account for 90-95% of people with type 2 diabetes. Diabetes has an impact on both men and women, with over 12 million men and 11.5 million women living with the disease. As a result, manually forecasting diabetes appears to be a waste of time and money. Diabetes therapy is based on controlling blood sugar levels in order to avoid various diet and exercise symptoms and consequences [12].

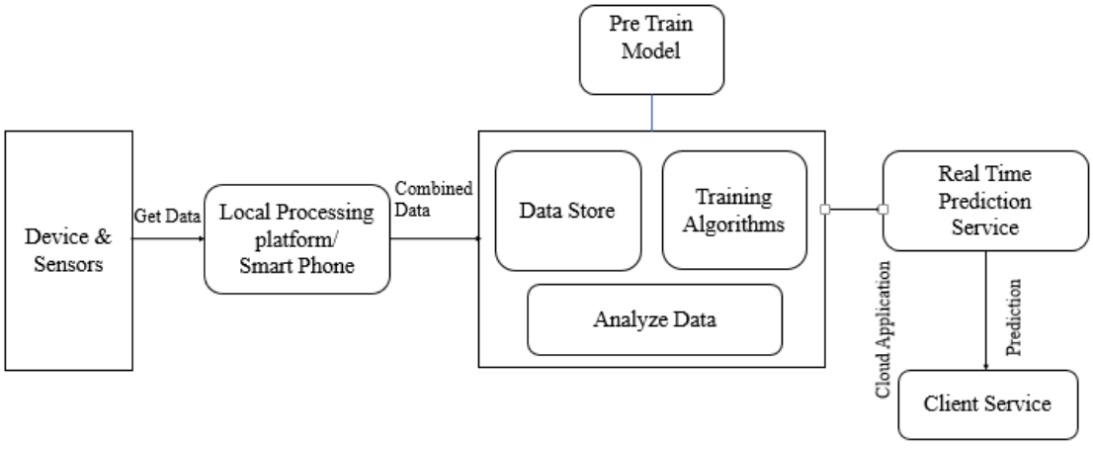


Figure 1: Diabetes Prediction Model

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

Types of Diabetes.

1. Type 1 DM

As a result of Type 1 diabetes, the pancreas does not produce enough insulin. Previously, the terms "insulin-induced diabetes mellitus" (IDDM) or "juvenile diabetes" were used. The cause behind this is unknown. Diabetes affects young people under the age of 20. Type 1 will impact pancreatic cells and cause them to fail. Type 1 diabetics have been suffering their entire lives and are dependent on insulin injections due to a lack of insulin production. Type 1 diabetic patients should engage in regular physical activity and eat a well-balanced diet.

2. Type 2 DM

Type 2 diabetes begins with insulin resistance, which occurs when cells do not respond appropriately to insulin. As the condition worsens, an insulin deficit may develop. Previously, the phrases "non-insulin-dependent diabetes mellitus" or "adult-induced diabetes" were used. The most prevalent reason is being overweight and not exercising. It occurs when insulin levels are insufficient to meet the body's needs. Obesity increases the risk of Type 2 diabetes based on family history and age. It is usually around the age of four.

3. Gestational Diabetes

This is the third primary type, which occurs when pregnant women acquire elevated blood sugar levels without a history of diabetes. According to the most recent data on diabetes, around 18 percent of pregnant women have diabetes. Increased gestational diabetes at an older age may be a risk factor. The third major type is frequently induced by elevated levels of blood sugar in pregnant women's bodies.

4. Presentational diabetes

is an insulin-dependent diabetes that occurs before pregnancy. It is a stage prior to type 2 diabetes in which the individual's glucose level is higher than usual but not at type 2 levels. In such situations or measurements, a man with prediabetes is more likely to score 2 [13–18].

B. Effects of diabetes

Diabetes is affected by the various parts of the body:

1. Loss of vision

Retinopathy is a condition in which the optic nerve of the eye is damaged. Contact of mind may be weakened due to difficulties with night vision deficit and swelling in the region of the retina. A diabetic individual can manage their eye

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

sight at the outset with a few tests or pharmaceuticals. The process involves a visual sharpness test, tonometry, student extension, and optical intelligibility tomography. Anti-VEGF infusion therapy, the macular laser focal/lattice surgical method, and a corticosteroid are among the medications used.

2. Kidney Neuropathy

Higher blood sugar levels damage kidney arteries, resulting in chronic kidney disease or diabetic neuropathy. The kidney is useful for transporting waste and water into the bloodstream. Overheads can disinfect the blood due to raised blood pressure and sugar levels in the kidney, which can lead to renal deception or subsequent blood dialysis. The operation will require kidney, pancreatic, or kidney replacement.

3. Liver Problems

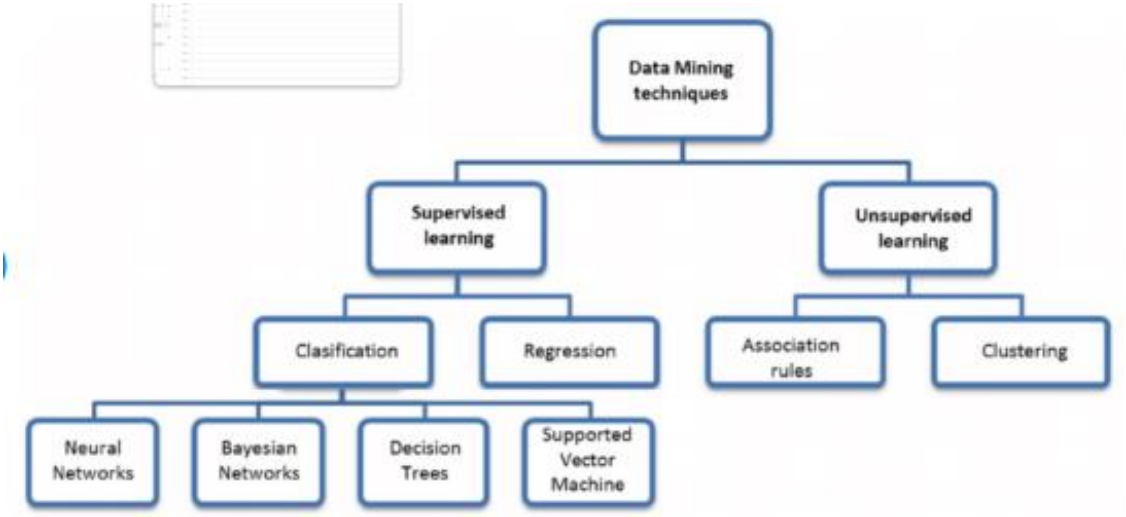
The liver performs a vital job in adjusting the quantity of blood glucose in the blood by digesting starch via glucogenesis or glycogenosis. Type 2 diabetes increases the risk of liver problems. A fatty liver takes over the production of a liver tumour. These concerns include renal weakness, altered metabolism, insulin resistance or excessive blood sugar levels, and malnutrition. Affected people tend to take numerous anti-toxin therapies or liver administration, which includes several dissertations, such as dietary modification, pharmacology, insulin secretagogues, biguanides, -glucosidase inhibitors, TZD, and weight loss.

4. Heart Problems Cardiovascular Ailment

In the United States, 68% of the population suffers the negative repercussions of cardiac attacks, atherosclerosis, or blocked supply pathways, exhaustion, and heart stress, which cause them to die. Because of the increased volume of sugar, blood transmits with a broader thickness, sticks to the veins, and the routes of delivery and veins put greater pressure on advancing. Continuous damage to arteries and nerves leads to the deception of the circulator, or organ frame. Hypertension, high cholesterol and triglycerides, obesity, and a lack of physical activity are all risk factors for cardiovascular disease. Multiple clinical characteristics, such as poor glycemic control and diabetes with insulin resistance, have an impact on heart issues [19].

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

Prediction of Diabetes Using Dm Technique



Diabetes prediction in an early stage will lead to more effective treatment. Prediction of diabetes using several diabetes management techniques Classification, clustering, prediction, NB, DT, and other techniques are frequently used for diabetes data extraction and prediction. Data mining techniques are utilized in a variety of applications, including the intriguing and educational areas of banking and so on.

1. Classification

Classification is one way used to predict diabetes. Classification is the most important mining activity. A large amount of corporate or medical data is often categorized. Data classification is a data mining function that may be assigned to target categories for items in a database. As a preprocessing step, this approach may be employed before placing data in the classification model. Clusters of similar components To collect everything from its properties, we must cluster the data. And add them in the order of their similarity.

Clustering is a type of segmentation. It is used to categories raw case groups based on a range of factors. cases within a comparable category with more or less balanced attribute values. Clustering is a single data mining task. This indicates that the preparation phase isn't defined by a single characteristic. Both input characteristics are treated in the same way. Before clustering, the values of characteristics must be codified in order to eliminate the higher values that affect the lower value attributes.

2. Naive Bayes

The Bayes theorem is founded on NB and predictor guesswork. NB facilitates the

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

quick building of models that provide estimates, hence giving a tool for data navigation and interpretation. When building a regression model using NB for predictive analysis, this process may be expanded. All input data must have comparative independence. An NB model has no iterative parameters yet is simple to construct. Naive Bayes might be a useful predictor. This approach is appropriate for certain extremely large data sets. NB starts before and after the lowering of characteristics with comparable development times. The NB prediction is simple to grasp.

3. Decision tree

The most prevalent is DT technology. DT is an important and simple-to-implement classification model. A decision tree is used as a statistical model for data processing in DT. In this work, DT is used to estimate illnesses and classify techniques based on patient data. DT can be quickly developed and translated. The forecast based on DT is well-organized. It handles a massive amount of data. It is perfect for studying scientific investigation. Finally, the Decision Tree findings are easy to explain and interpret.

4. Knowledge discovery database

KDD is a data and information collection strategy. This strategy is also considered a data mining methodology. It entails data processing, data selection, data preparation, and information establishment on data sets, as well as the interpretation of optimal ways based on observed results. Data mining is an important stage in KDD. It is used to retrieve database data. KDD consists of an iterative data integration sequence, DM pattern recognition, and information introduction.

5. Regression

In diabetes, regression is a commonly used function. Regression is rated the same way. It is used to forecast an infinite number of variables, such as age, weight, temperature, and sickness. All of these may be forecasted using regression techniques. Regression activities can help to tackle a variety of medical problems. Linear regression and logistic regression are the two most popular types of regression.

6. Association

The most important DM property that determines the likelihood of items is association. The rules of the association depict the relationships between working together and the items. association, which is also known as market-based research. In terms of association, an object is a value

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

or, more specifically, every attribute or value-pair. The job of association has two goals: one is to recognise often occurring things, and the other is to discover association rules. Relationship simulation has been shown to be an efficient approach in other domains. The organisation is also involved in other domains of application, such as bio-information technology, medical diagnostics, web mining, and the analysis of scientific evidence. The association rules are created to analyse data on recurrent patterns and categorise the most comprehensive associations with the help of support and confidence stages. Support is one of the most visible items in the database. Confidence is the number of times assertions have been confirmed to be correct.

7. K- means clustering

Clustering [20] is a grouping strategy based on similar artefacts. It is an unsupervised learning approach that computes the natural grouping of examples for unlabeled data. The kinds are interchangeable. However, the things in one cluster differ from the objects in other clusters. Objects' intracluster similarity is high, while their intercluster similarity is low. There are numerous classifications, such as partitions and hierarchical clusters, but in this investigation, the K-means clustering strategy was chosen. K-Means clustering, which employs numerical data and describes K as a cluster core, is extremely straightforward and understandable. Each instance of the cluster is assigned based on the distance of each data point from the centre, and the cluster centres are transferred using all of the cluster's data points.

8. Artificial Neural Network

An ANN is a data processing device modelled on the operation of the human brain. NNs are often organised into layers with several linked nodes and an activation function. The network receives patterns from the input layers, which link to one or more hidden layers, where real processing is handled via a system of weighted ties. For outcomes, hidden layers are connected to an output layer.

9. Random forest

RF is a tree-based collection classifier in which the randomly assigned vectors are distributed totally differently, with the input X serving as the unit for each tree. A random vector is created, and a tree used for training is isolated from earlier random vectors of the same distribution. The random forest method is a scalable, easy, and straightforward methodology for combining tree predictors. The results of

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

the random forest are often excellent. Its performance is difficult to improve, and it can operate with a variety of integer, binary, and nominal data.

10. Naive Bayes Classifiers

The Bayesian network is commonly used for classification problems. Naive Bayesian Networks (NBN) are the most fundamental Bayesian networks, consisting of direct acyclic graphics with one parent and only a few children (related to the nodes seen), with a strong premise of autonomy amongst children in their parents' environment.

11. Bayesian Networks

BNs are visual models that are used to describe and quantify the likelihood or uncertainty of relationships between events and concepts. Some of BN's primary uses are knowledge extraction, input or recognition software-based predictions. As a probabilistic model, a Bayesian network defines the variables and their relationships.

12. Support Vector Machines (SVMs)

SVMs are supervised learning algorithms that are used to identify, regress, or detect forms. There are several advantages to using SVM: ii) it uses a subset of decision-making training points (called support vectors), so memory effectiveness is also possible, and iii) it is scalable, because it has various functions of a kernel that can be defined for the decision [21, 22,23,24].

LITERATURE REVIEW

W. Chen et al. (2017) This research proposes a hybrid prediction model for the diagnosis of type 2 diabetes. K-means will be employed as a classification classifier for data reduction using the J48 decision tree in the created framework. The results show that the suggested model is more accurate than other earlier investigations included in the literature. Based on the findings [25], it has been shown that the suggested method is helpful for diagnosing type 2 diabetes.

A. Mir and S. N. Dhage (2018) The goal of this project is to develop a model for diabetes categorization utilising a WEKA tool and the NB, SVM, RF, and Basic CART algorithms. Based on effective data, the study seeks to provide the best algorithm for illness prediction. Each dataset algorithm was examined and recommended for execution. The SVM is shown to be the most successful in illness estimate [26].

A. M. Rajeswari et al. (2018) The proposed method defines prediabetes risk

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

factors using associative classification technology based on fuzzy logic. The proposed technique revealed a more precise set of prediabetes (outliers) among aspects relevant to the negative diabetes outcome. [27]

K. VijiyaKumar et al. (2019) This study would provide a strategy for predicting diabetes early using RF algorithms for a highly accurate patient using ML. The degree of precision is higher, especially when compared to other methods. The suggested model provides the best diabetes forecasting outcomes, and the data shows that the prediction approach may reliably and, more importantly, rapidly anticipate diabetes illness [28].

N. L. Fitriyani et al. (2020) To identify and delete external information and balance data distribution, the proposed framework combines forest isolation with the oversampling-Tomek minority synthetic link approach. They use 10-fold cross-validation to accomplish four outcome measures' precision, recall, f-measure, and accuracy. The test results show that the proposed model outperformed other models with precision of up to 93.18%, 98.87%, and 96.09% for datasets I, II, and III, respectively. When type 2 diabetes is detected, the early diabetes prediction is designed to urge patients to take precautions [29].

S. M. Mahedy et al. (2020) In this paper, the Adaptive Boosting Algorithm (AB) is utilised to develop a tree-driven ML model for the classification of a Pima Indian diabetes data set. In the study of experimental findings, the Extra Tree (ET) as a basis estimator of AdaBoost classification provides the highest precision of 90.5 percent. Our proposed tree-based ML model will also aid healthcare practitioners in the detection of diabetes. [30].

G. Tripathi and R. Kumar (2020) The proposed approach provides the closest findings in terms of empirical outcomes, allowing for better patient diagnosis. There are four ML algorithms used in the prediction analysis of early-stage diabetes: Linear Diskriminant Analysis (LDA), KNN, SVM, and RF. These classification algorithms put their output to the test using a variety of statistical criteria such as sensitivity (recall), precision, accuracy, F-score, and precision. Precision is the computation of correct and incorrect case classification. The test results indicate that RF provides 87,66% maximum accuracy, which has been achieved in previous classification methods [31].

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

K. L. Priya et al. (2020) In this recommended experimental procedure, the Nave Bayes Classification is frequently used to anticipate diabetes. Information mining is a technique for retrieving data from a dataset and converting it into a lawful system for future usage. Diabetes may be predicted with a better accuracy criterion (0.96) than is now standard in the current innovative strategy. In this suggested technique employing the Nave Bayes Classification system, the output would be a Web Interface with the result of diabetes or not, whether the input data together with insulin and age are also taken into account. This enhances the accuracy of the system [32].

CONCLUSION

This study expands on several data mining tools and strategies utilised in early diabetes prediction. DM is a method that helps you comprehend more by extracting important knowledge from a large volume of data. By using data mining methods and technology, diabetes can be avoided and treatment expenses reduced. This research makes an attempt to forecast diabetes using DM methods. Diabetes is a disease that affects a large number of people every day. Using various classifiers, numerous automated information systems for diabetes prevention and diagnosis have been built. More and more homes are unfair as the prevalence of diabetes mellitus rises. Many people with diabetes are unaware of their pre-diagnosis risk factors. Data mining aids in the detection of patient ailments. It has played a significant role in diabetes research. Diabetes is a disease that affects a large number of people every day. The majority of research is conducted in this sector.

REFERENCES

1. Falvo D, Holland BE. Medical and psychosocial aspects of chronic illness and disability. Jones & Bartlett Learning; 2017.
2. Tao Z, Shi A, Zhao J. Epidemiological perspectives of diabetes. Cell Biochem Biophys 2015;73:181–5.
3. Skyler JS, Bakris GL, Bonifacio E, Darsow T, Eckel RH, Groop L, et al. Differentiation of diabetes by pathophysiology, natural history, and prognosis. Diabetes 2017;66:241–55.
4. Sathya Chandrasekaran, K.Dharmarajan," Survey on Data Mining Classification Techniques to Predict Diabetes", Elysium journal, A u g u s t 2 0 1 7, V o l u m e - 4, I s s u e – 4

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

5. Karnika Dwivedi, Dr Hari Om Sharan," Review on Prediction of Diabetes Mellitus using Data Mining Technique", International Journal of Engineering and Technical Research (IJETR), ISSN: 2321-0869 (O) 2454-4698 (P) Volume-8, Issue-12, December 2018
6. Acharya, Rajendra, U and Yu, Wenwei, (2010).Data Mining Techniques in Medical Informatics. The Open Medical Informatics Journal, PMCID: PMC2916206.
7. Chauraisa V., and Pal, S.,(2013). 'Data Mining Approach to Detect Heart Diseases', International Journal of Advanced Computer Science and Information Technology (IJACSIT), 2, (4), pp 56-66.
8. S.Vijiyarani S.Sudha," Disease Prediction in Data Mining Technique"– A Survey, International Journal of Computer Applications & Information Technology Vol. II, Issue I, January 2013 (ISSN: 2278-7720)
9. Anusha N, Rajashree, Srikanth Bhat K," Survey on Medical Data by using Data Mining Techniques", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 7, Issue 1, January 2018, ISSN: 2278 -7798
10. Desmond Bala Bisandu," Diabetes Prediction Using Data Mining Techniques", International Journal of Research and Innovation in Applied Science (IJRIAS) | Volume IV, Issue VI, June 2019|ISSN 2454-6194
11. P. Suresh Kumar and V. Umatejaswi," Diagnosing Diabetes using Data Mining Techniques", International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017 705 ISSN 2250-3153
12. P. Suresh Kumar and V. Umatejaswi, " Diagnosing Diabetes using Data Mining Techniques", International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017.
13. M. Sharma, G. Singh, R. Singh, "Stark Assessment of Lifestyle Based Human Disorders Using Data Mining Based Learning Techniques", Elsevier, vol. 5, pp. 202-222, 2017.
14. Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang, "Type 2 diabetes mellitus

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

prediction model based on data mining", ScienceDirect, Vol. 11, issue 3, pp. 12-23, 2018.

15. Yan Luo, Charles Ling, PhD, Jody Schuurman, Robert Petrella, MD, "GlucoGuide: An Intelligent Type-2 Diabetes Solution Using Data Mining and Mobile Computing", 2014 IEEE International Conference on Data Mining, Vol. 9, issue 8, pp. 12-23, 2014.
16. Isha Vashi, Prof. Shailendra Mishra," A Comparative Study of Classification Algorithms for Disease Prediction in Health Care", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 9, September 2016.
17. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: analysis and implementation. IEEE transactions on pattern analysis & machine intelligence. 2002. p. 881–92.
18. N. Sneha1 and Tarun Gangil," Analysis of diabetes mellitus for early prediction using optimal features selection", Springer open, 2019, page 1-19
19. K.Priyadarshini," A Survey on Prediction of Diabetes Using Data Mining Technique", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Special Issue 11, September 2017
20. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Commun. Surv. Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016, DOI: 10.1109/COMST.2015.2494502.
21. I. Muhammad and Z. Yan, "Supervised Machine Learning Approaches: a Survey," ICTACT J. Soft Comput., vol. 05, no. 03, pp. 946–952, 2015, DOI: 10.21917/ijsc.2015.0133.
22. N. F. Haq, A. R. Onik, A. K. Hridoy, M. Rafni, F. M. Shah, and D. Farid, "Application of Machine Learning Approaches in Intrusion Detection System: A Survey," Int. J. Adv. Res. Artif. Intell., vol. 4, no. 3, pp. 9–18, 2015.
23. W. Chen, S. Chen, H. Zhang and T. Wu, "A hybrid prediction model for type 2 diabetes using K-means and decision tree," 2017 8th IEEE International Conference on

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved

Software Engineering and Service Science (ICSESS), Beijing, China, 2017, pp. 386-390, DOI: 10.1109/ICSESS.2017.8342938.

24. A. Mir and S. N. Dhage, "Diabetes Disease Prediction Using Machine Learning on Big Data of Healthcare," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, DOI: 10.1109/ICCUBEA.2018.8697439.
25. A. M. Rajeswari, M. S. Sidhika, M. Kalaivani and C. Deisy, "Prediction of Prediabetes using Fuzzy Logic based Association Classification," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 782-787, DOI: 10.1109/ICICCT.2018.8473159.
26. K. VijiyaKumar, B. Lavanya, I. Nirmala and S. S. Caroline, "Random Forest Algorithm for the Prediction of Diabetes," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019, pp. 1-5, DOI: 10.1109/ICSCAN.2019.8878802.
27. N. L. Fitriyani, M. Syafrudin, G. Alfian, A. Fatwanto, S. L. Qolbiyani and J. Rhee, "Prediction Model for Type 2 Diabetes using Stacked Ensemble Classifiers," 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 2020, pp. 399-402, DOI: 10.1109/DASA51403.2020.9317090.

Page 49- 64 © MANTECH PUBLICATIONS 2022. All Rights Reserved