Data Mining Techniques Based Diabetes Prediction

Aditya Saxena, Megha Jain, Prashant Shrivastava

Abstract: Data mining plays an important part in the healthcare sector disease prediction. Techniques of data mining are commonly used in early disease detection. Diabetes is one of the world's greatest health challenges. A widespread chronic condition is a diabetes. Diabetes prediction is a science that is increasingly growing. Diabetes prediction at an early stage will lead to better therapy. It is necessary to avoid, monitor and increase diabetes consciousness because it causes other health issues. Diabetes of type 1 or type 2 can lead to heart disorders, kidney diseases or complications with the eye. This survey paper reflects on numerous approaches and data mining strategies used to forecast multiple diabetes disorders at an early stage. Become a chronic disease because of diabetes. The patient lives will be spared by an early prediction of this disease. By the use of data mining tools and processes, diabetes is avoided and treatment rates are reduced. The association rule mining, classification, clustering, Random Forest, Prediction as well as the Artificial Neural Network (ANN) are among the most common and important data mining technology. Different data mining methods are available to avoid diseases such as cardiac disease, cancer including kidney etc. This study examines the use of data mining methods to predict multiple disease types.

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

I. INTRODUCTION

Data mining (DM) is used to invent and view information from data in a readily interpreted condition for humans. It is a method for reviewing vast quantities of data. The application of data mining strategies in different industries such as finance, education and so on plays a significant role in IT. [1]. DM can be successfully extended to disease prediction using different data mining approaches in the area of the medical domain. DM appears to be expected and defined in two dominant objectives. The prediction uses several variables or fields within the data set to forecast such variables of importance, either uncertain or potential values. The focus of the explanation is on the trends of human perception.

Aditya Saxena, Student, Department of computer science and Engineering, Lakshmi Narain College of Technology Excellence Bhopal, India. Email: Aditya01oist@gmail.com

Megha Jain*, Assistant professor, Department of computer science and Engineering, Lakshmi Narain College of Technology Excellence Bhopal, India. Email: meghaj@lnct.ac.in

Prashant Shrivastava, Assistant professor, Department of computer science and Engineering, Lakshmi Narain College of Technology Excellence Bhopal, India. Email: prashant.shri1@gmail.com

Before designing predictive models, it is very important to understand the development and characteristics of impacting diabetes from external sources. The idea is to foresee diabetes & to classify diabetes causes by using data extraction methods [2].The word diabetes is a condition occurring when the blood glucose in the body is too high and is also known as blood sugar. As described above, diabetes can lead to other significant cardiovascular complications. According to the WHO, there are 3.7 million deaths by age 70 years and cardiovascular disease are responsible for this high death rate. The key cause behind diabetes is uncontrolled blood glucose levels. A chronic condition is a continuous disease or condition whose consequences are permanent. This illness forms have a significantly detrimental effect on the quality of life. Diabetes is one of the most acute and worldwide illnesses. This chronic disease is a significant cause of death of adults worldwide. The costs are also correlated with chronic conditions. Governments and people expend a large part of the budget on chronic diseases [3,4]. Worldwide diabetes statistics showed that about 382 million people worldwide suffered from this disease [5].

Diabetes is a chronic lifelong condition that affects the human body by the reduction of glucose-carrying insulin into the blood cells. This raises the level of sugar in the body which leads to multiple complications, including stroke, cardiac disease, blindness, kidney failure and death [6]. Diabetes can be a chronic sickness with no illustrious cure if it insists on maintaining a traditional level of blood glucose without inducing hypoglycemia. Diet, exercise and use of relevant medicinal products are manageable. Diabetes Mellitus takes place around the world, and in developing worlds it's a big deal. The rise in rates in developed countries is a result of the movement towards rapid urbanization including changes in lifestyles, such as a "West" diet. This is attributed to a lack of understanding. Diabetes diagnosis for quantitative analysis is a complex problem. Certain parameters such as A1c, fructosamine, the number of white blood cells, fibrinogen, etc., are inefficient due to small levels. [9].

II. DATA MINING APPLICATION IN HEALTHCARE SECTOR

Data mining is a very recent concept for knowledge extraction from a vast variety of data. Mining| involves the use and treatment of the data available to make decisions. The phase of DM is an exploration of models in large data sets involving approaches at crossroads of machine learning, statistics & database systems. DM is an interdisciplinary subfield of computer science & statistics with the ultimate purpose of extracting knowledge (with an improvement of existing) from a data set as well as converting it into a meaningful structure.

29

Data mining has developed based on individual needs and can help patients recognise relationship patterns and predictions based on pre-determined guidelines and programme stipulations (Eapen, 2004).

Data mining assists in the analysis of patterns including the categorization of data by doing cluster studies, recognising irregular records often called anomalies detection and associated rules or dependency. [8,9].

In the field of medical dm, different DM approaches can be used to accurately forecast diseases. DM propensity is to predict and define two predominant targets. The prediction uses several variables or fields within the data collection to forecast such variables of importance, either uncertain or potential values. The definition aims to define trends that detail the human details. Before designing predictive models, it is very important to understand the development and characteristics of impacting diabetes from external sources. The idea is to foresee diabetes & to classify diabetes causes by using data extraction methods [10].

1. Diagnosis and Prediction of Diseases

In terms of social security providers, we must conclude and expect suffering, which is one of the most significant factors behind the use of knowledge for social insurance. The use of human resources information-digging has allowed experts to improve the control of their health. They cannot sit idle and have a cash base by selecting a patient who will also damage the health of the patient.

2. Ranking of Various Hospitals

The data mining approach serves to focus on each of the points of concern of the various health centres, taking into account the final purpose of classifying them. Associations identify various health centres as possessing the potential to handle people of real illnesses, i.e., higher-ranking healing centres are ideally equipped to treatment for high-risk patients when they are more in need, however, this is not the case in decommissioning clinics since they may not recognise the risk factors substantially.

3. Better Treatment Techniques

The professional and patient will choose the most suitable care option from all treatment schemes with the aid of data mining techniques. They will choose both the adequacy and expense of the right healthcare services. They can also detect the reactions of various drugs by information mining and thereby reduce the risk for patients.

4. Effective Treatments

By comparison, medical expertise mining is being used to split the adequacy of drugs with components such as causes, signs, symptoms as well as cost. For example, the effects of different patients who have experienced a common illness have been treated with different medicines. Along these lines, we will point out if the patient's well enough and costs are convincing.

5. Infection Control In Hospitals

Doctoral disorders consistently involve a significant number of patients, as well as the level of sedate-safe contamination, is relatively high. To identify such unpredictable instances in the disease control information, contamination is examined by information mining. These examples are further investigated by a trained person for contamination control. This system of identification to the

degree that information extraction techniques were revised in the University of Alabama to identify elusive examples for information from a primary.

6. Identifying High-Risk Patients

American Health approaches allow diabetic clinics to improve efficiencies and minimize the expenses of diabetic patients through administration. American Health approaches used an innovative monitor technique to distinguish high- and low- chance patients. Using an influential display technique, the providers of healthcare care differentiated high-chance patients who needed more attention to their well-being.

7. Medical Device Industry

The social insurance industry does not survive without medicinal gadgets. Portable or economic remote bio-sensors seem to be the most important component of flexible drugs, a safe technique that emphasises patients' imperative indications [11].

III. OVERVIEW OF DIABETES DISEASE

Diabetes is a condition that lasts a long time and affects people all over the world. It develops when the body can't produce adequate insulin. Insulin, one of the most essential hormones in the body that is required to retain glucose levels, is activated by the pancreas. The use of hormone injections, a healthy diet & frequent exercise may regulate diabetes. Diabetes affects several more problems, including blindness, blood pressure, heart and kidney failure.
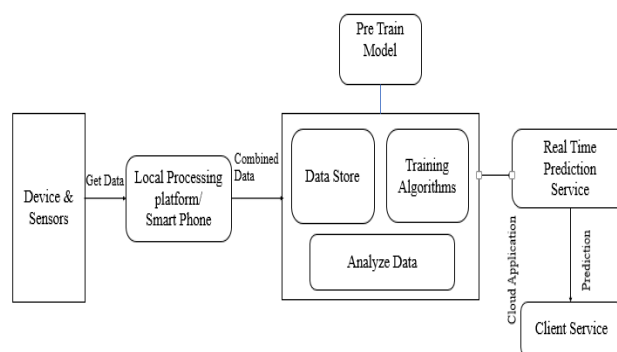

Figure 1: Diabetes Prediction Model

Diabetes is the disease most prevalent. It is necessary to avoid, monitor and increase diabetes consciousness because it causes other health issues. diabetes is classified into two distinct types: diabetes type 1 strengthens the need for insulin to be injected artificially by treatment and vaccines, and diabetes type 2. Most diabetes patients are affected by diabetes type 2. Diabetes was a prevalent concern among adults of middle age but diabetes still affects children because of changes in lifestyles. The numerous external stimulants that contribute to the death of the cells producing insulin in the body are unpreventable in Type 1 diabetes. Changing diets to achieve a particular physical weight will help to prevent type 2 diabetes from increasing. Diabetes is a debilitating but preventable chronic health condition.

High blood glucose levels due to insulin development deficiencies, insulin action, or both. 1,2 1,2 In 2003 diabetes rates globally amounted to 15.1 million, and by 2030 it is expected to grow to 36.6 million diabetes worldwide. Of all, 90-95% of people with type 2 diabetes are adults. influence of diabetes is relative to males and females, with over 12 million men or 11.5 women living with diabetes. So it also seems not objective to manually forecast diabetes, because it takes a lot of time and resources. Diabetes treatment relies on blood sugar levels management to avoid different diet and exercise and exercise symptoms and complications [12].

A. Types of Diabetes.

1. Type 1 DM

The pancreas does not contain enough insulin as a result of Type 1 DM. The word "insulin-induced diabetes mellitus" (IDDM) or "juvenile diabetes" was used in earlier years. The reason is not known. Youth and under 20 years of age are impacted by diabetes. Type 1 will affect the pancreatic cells and would not function. Type 1 diabetic patients have been suffering throughout their lives and are reliant on insulin injections due to null insulin secretion. Diabetic patients with type 1 should consistently follow physical exercises and balanced diets.

2. Type 2 DM

Type 2 DM starts with insulin resistance, which means that cells do not respond adequately to insulin. A deficiency of insulin may evolve as the disease progresses. The term 'non-insulin-based diabetes mellitus' has been used before or adult-induced diabetes.' Overweight and poor exercise is the most common cause. It happens when insulin levels are not adequate for the body's requirements. Obesity raises the risk of Type 2 diabetes as a function of a family's history, old age. It is typically at the age of 4 years.

3. Gestational Diabetes

This is the 3rd primary type that happens whenever pregnant women develop higher blood sugar levels without a prior history of diabetes. It has been observed that about 18 percent of women with pregnancies have diabetes, as per the latest report on diabetes. Increased gestational diabetes during an older age may be at risk. The third major type is often caused by high amounts of blood sugar in the body of pregnant women.

4. Pregestational Diabetes

Pregestational diabetes takes place before pregnancy with insulin-dependent diabetes. It is a period before type 2 diabetes, where the individual's glucose level was higher than average but not at type 2 levels. A man with prediabetes is more likely to get 2 in such conditions or measures [13-18].

B. Effects of diabetes

Diabetes is affected by the various parts of the body:

1. Loss of vision

Retinopathy is a disorder in which the eye, the optic nerve, is impaired. Owing to problems with vision deficiency at night, swelling in the area of the retina, contact of mind may be reduced. In a few tests or pharmaceutical, a diabetic person can cope with eye vision at the onset. The procedure includes visual sharpness test, tonometry, student extension and tomography of optical intelligibility. Various drugs include anti-VEGF infusion treatment, macular laser focal/lattice surgical technique, a corticosteroid.

2. Kidney Neuropathy

The higher blood sugar levels destroy kidney arteries, which is chronic kidney or diabetic neuropathy. The kidney is good for channelling waste and loads of water into the blood. Due to elevated blood pressure and sugar levels in the kidney, overheads can disinfect the blood and this can lead to kidney deception or successive blood dialysis. The procedure will require replacement for the kidney, kidney or pancreas.

3. Liver Problems

The liver takes an invaluable function to adjust the amount of blood glucose in the blood through the digestion of starch through glucogenesis or glycogenosis. The risk of liver complications is raised by type 2 diabetes. A fatty liver takes over the task to produce a liver tumour. These issues include renal weakness, changed metabolism, opposition to insulin or high blood sugar levels, malnutrition. Affect people tend to take multiple anti-toxin treatments or liver administration incorporates several dissertations, such as dietary modification, pharmacology, insulin secretagogues, biguanides, inhibitors of α-glucosidase, TZD, weights to be smaller.

4. Heart Problems Cardiovascular Ailment

In the United States, 68% of the population suffer the adverse consequences of cardiac attacks that lead even to death, heart attack, atherosclerosis or solidify supply routes, fatigue and heart stress cause them to die. Owing to the higher amount & sugar, blood transmits a wider thickness, adheres to the veins, the routes of delivery & veins bring more pressure on advancing. It is continuous harm to vessels and nerves which in person contributes to the deception of the circulator or organ frame. Hypertension, exceptional cholesterol and elevated triglycerides, corpulence, lack of physical exercise, are considered a danger of cardiovascular disease. The results of multiple clinical parameters, such as weak glycaemic regulation and diabetes insulin opposition, affect cardiac problems [19].

IV. PREDICTION OF DIABETES USING DM TECHNIQUE

Diabetes prediction at an early stage will lead to better therapy. Prediction of diabetes with multiple DM strategies. Classification, clustering, prediction, NB, DT & so on techniques are widely recognized for data extraction and prediction for diabetes. For various applications such as the interesting and educational domain, finance, etc., data mining techniques are used.
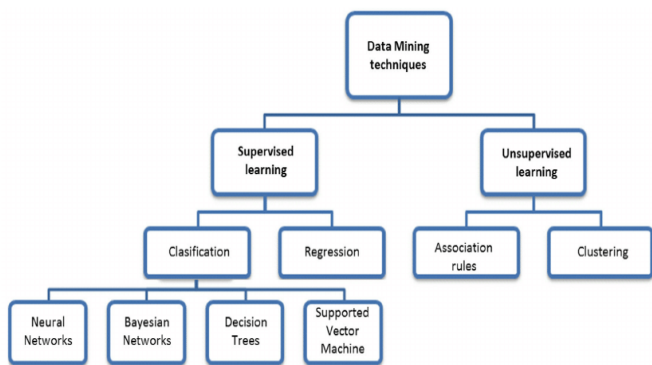
31

Figure 2: Data Mining Techniques

1. Classification

One of the methods applied to predict diabetes is classification. The most significant mining activities are classification. Typically a great deal of business or medical data is classified. Classification is a function for data mining and can be allocated to target categories for objects in a database. This procedure can be used before storing data in the classification model as a preprocessing step. Clusters of the same elements. We have to cluster to find a group of data to capture everything from its characteristics. And add them in line with their similarities. As segmentation, clustering is also known. It is applied to classify unprocessed case groupings depending on a variety of attributes. Cases within a similar category of values of attributes more or less balanced. Clustering is a single task for data mining. This means that the preparation phase is not formed by a single attribute. Both input attributes are similarly processed. The values of attributes must be formalized to remove the higher values that affect the lower value attributes before clustering.

2. Naive Bayes

Bayes theorem is based on NB with predictor speculation. NB helps the rapid development of models delivering estimation and thus providing a tool for data traversing and comprehension. This methodology can be extended when constructing a regression model with NB for predictive analysis. This includes comparative independence in all input data. An NB model has no specific iterative parameters but is easy to build. Naive Bayes may be a good indicator. For some very large data sets, this analysis is suitable. NB begins before & after the reduction of attributes of similar development time representation. The NB prediction is easy to understand.

3. Decision tree

The DT technology is the most common. DT is one of the most significant and easy to implement a classification model. As a statistical model for data processing, DT utilizes a decision tree. DT is used for the estimation of diseases and classification methods in this study from patient data. DT can be designed rapidly and translated easily. The DT-based forecast is well organized. It processes a tremendous volume of data. It is ideally suited to study the exploration of science. Finally, it is easier to explain and read findings from the Decision Tree.

4. Knowledge discovery database

KDD is the discovery method for data and information collection. This method is also regarded as a methodology for data mining. It involves data processing, data selection, data preparation and information establishment on data sets & interpretation of best approaches from observed findings. Data mining is a key KDD stage. It is used for the retrieval of database data. KDD comprises an iterative data integration sequence, DM pattern recognition & introduction of information.

5. Regression

Regression is a frequently used function in DM. Regression is graded identically. it is used for predicting endless numbers, age, weight, temperature, disease. These can all be forecast using strategies of regression. Regression tasks can solve several challenges in the medical field. The most common forms of regression are linear regression & logistic regression.

6. Association

Association is the most significant DM features that decide the probability of objects. The relations between working together and the objects are shown as rules of the association. Often known as market-based research is Association. A value or more commonly every attribute or value-pair is known as an object in terms of association. The role of association has two objectives: one is to identify frequent objects and another to find association rules. The simulation of relationships has also been an efficient process for other fields. The association also has links with other fields of use, including bio-information technology, medical diagnostics, web mining and the study of scientific evidence. The association rules are developed to analyse data on recurring trends and to classify the most comprehensive relationships using support and confidence steps. Support is the commonly visible objects in the database. Confidence is the amount of instances declarations are proven to be valid.

7. K- means clustering

Clustering [20] is the grouping mechanism centred on identical artefacts. It is an unsupervised learning strategy in which the natural grouping of instances for unlabeled data was calculated. The types are equivalent. But one cluster's objects vary from all cluster objects. The intra clustering similarity of objects is strong and the intercluster similarity of objects is low. There are several classifications, such as partitions and hierarchical clusters, but the K means clustering approach has been used in this analysis. It is very simple or comprehensible to use K-Means clustering which uses numerical data in which K is described as a cluster core. By taking the distance of each data point from the centre, each of the instances of the cluster is allocated to it, and the cluster centres are transferred using all of the data points of the cluster.

8. Artificial Neural Network

ANN is an information processing device based on human brain functionality. NNs are typically structured into layers composed of many interconnected nodes with an activation function.

32

The input layers present patterns to the network that connect with one or more hidden layers, where actual processing is performed by a system of weighted ties. secret layers are then connected to an output layer for results.

9. Random forest

RF is a tree-based collection classifier that the random vectors independently allocated are distributed completely differently, with the input X being the unit for every tree. A random vector is generated as well as a tree used for the training test is separated from the prior random vectors of the same distribution. The random forest approach is a scalable, easy and simple algorithm that combines tree predictors. Random forest findings are generally satisfactory. Its performance is hard to boost and it can even work with a number, binary, and nominal data of various kinds.

10. Naive Bayes Classifiers

Classification tasks are typically found in the Bayesian network. Naive Bayesian Networks (NBN) are very basic Bayesian networks consisting of direct acyclic graphics with one parent and only multiple children (relating to the nodes observed), with an important presumption of autonomy between kids in their parents' context.

11. Bayesian Networks

BN are graphic models used to explain relations among events and ideas and assess their probability or uncertainty. Knowledge extraction, input or recognition software-based predictions are some of BN's key applications. A Bayesian network describes the variables as well as the relations with them as a probabilistic model.

12. Support Vector Machines (SVMs)

SVMs are a group of supervised learning techniques used to identify, regress or detect shapes. For the use of SVM, there are several advantages: i) Large feature space is efficient, ii) Utilizes the sub-set of decision-making training points (named support vectors), so memory effectiveness is also possible and iii) is scalable, since it has various functions of a kernel can be defined for the decision [21, 22,23,24].

V. LITERATURE REVIEW

W. Chen et al. (2017) A hybrid prediction model for the diagnosis of type 2 diabetes is being suggested in this paper. In the developed framework, K-means will be used as a classification classifier for data reduction with the J48 decision tree. The findings indicate that the model proposed is more accurate than other previous studies listed in the works. Due to the findings, the proposed model is proven useful for the diagnosis of type 2 diabetes [25].

A. Mir and S. N. Dhage (2018) this study aims to create a model for the classification of Diabetes using a WEKA tool by using NB, SVM and RF and Basic CART algorithms. The study hopes to propose the best algorithm for the prediction of diseases based on effective data. Analyzed advisable to conduct of each dataset algorithm. It is found that the SVM is most effective in the estimation of the disease [26].

A. M. Rajeswari et al. (2018) The method suggested defines prediabetes risk factors using the associative classification technology fuzzy logic - based. The suggested technique showed that a more exact set of prediabetes (outliers) was established among aspects relevant to the negative diabetes outcome. [27].

K. VijiyaKumar et al. (2019) This paper would build a method that can forecast diabetes early with the use of RF algorithms for a highly accurate patient using ML. Especially in comparison with other algorithms, the degree of precision is higher. The model proposed offers the best diabetes forecasting outcomes, and the findings demonstrate that the prediction method can accurately, and particularly instantly predict diabetes disease [28].

N. L. Fitriyani et al. (2020) The suggested framework combines forest isolation as well as the oversampling-Tomek minority synthetic link technique to detect and eliminate the external information and to balance the distribution of data. They apply 10-fold cross-validations to achieve the precision, recall, f-measure or accuracy of four outcome measures. The test data demonstrate that the proposed model outperformed additional models with precision up to 93.18%, 98.87% & 96.09% respectively for datasets I, II and III. The early diabetes forecast is supposed to make people take care when type 2 diabetes is identified [29].

S. M. Mahedy et al. (2020) The Adaptive Boosting Algorithm (AB) is used in this study to build a tree-driven ML model for classification of Pima Indian diabetes data set. The Extra Tree (ET), as basis estimator of AdaBoost classification, gives the maximum precision of 90.5 per cent in the analysis of experimental results. Our suggested tree-based ML model will also help healthcare practitioners detect diabetes. [30].

G. Tripathi and R. Kumar (2020) The suggested framework offers the nearest results in terms of empirical outcomes which allows improving patient diagnosis. In the predictive analysis of early-phase diabetes, there are four ML algorithms: Linear Diskriminant Analysis (LDA), KNN, SVM & RF. These classification algorithms test their output using various statistical metrics including sensitivity (recall), precision, accuracy, F-score or precision. Precision is the calculation of the proper and inaccurate classification of cases. The test results suggest that RF offers 87,66 percent of maximum accuracy and has been reached in other classification algorithms [31].

K. L. Priya et al. (2020) The Naïve Bayes Classification is often used to forecast diabetes in this suggested experimental method. Information mining is a method by which data from a dataset is retrieved as well as changed into a legitimate system for further use. Diabetes with higher accuracy standard (0.96) than current ordinarily can be predicted in the current novel Strategy. The output would be a Web Interface with the result of diabetes or not, whether the input values along with insulin, age, are also taken in this proposed method using the Naïve Bayes Classification system. This improves the system's accuracy [32].

33

VI. CONCLUSION

This study expands on different approaches and techniques of data mining used in early diabetes prediction. DM is a process used to derive valuable knowledge from a vast amount of data that helps you to understand more. By the use of data mining tools and technologies, diabetes is avoided and treatment costs are minimized. This paper attempts to predict diabetes by using the techniques of DM. Diabetes is a condition that affects a lot of people every day. Different automated information systems have been developed for the prevention and diagnosis of diabetes using various classifiers. Due to the increasingly rising rate of diabetes mellitus, more and more households are unjust. Many diabetes is little aware of their prediagnosis risk factor. Data mining helps detect illnesses of the patient. In diabetes research, it has played a major role. Diabetes is an illness that affects a lot of people a day. In this field, most research takes place.

REFERENCES

1. P. Radha, Dr B. Srinivasan, "Predicting Diabetes by cosequencing the various Data Mining Classification Techniques", IJISET - International Journal of Innovative Science, Engineering & Technology Vol. 1 Issue 6, August 2014

2. K.Priyadarshini, Dr I.Lakshmi "A Survey on Prediction of Diabetes Using Data Mining Technique" International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 6, Special Issue 11, September 2017.

3. Falvo D, Holland BE. Medical and psychosocial aspects of chronic illness and disability. Jones & Bartlett Learning; 2017.

4. 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. [CrossRef]

5. Tao Z, Shi A, Zhao J. Epidemiological perspectives of diabetes. Cell Biochem Biophys 2015;73:181–5. [CrossRef]

6. 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

7. 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 [CrossRef]

8. Acharya, Rajendra, U and Yu, Wenwei, (2010).Data Mining Techniques in Medical Informatics. The Open Medical Informatics Journal, PMCID: PMC2916206. [CrossRef]

9. 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.

10. 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)

11. 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

12. 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

13. 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

14. 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.

15. 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.

16. Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang, "Type 2 diabetes mellitus prediction model based on data mining", ScienceDirect, Vol. 11, issue 3, pp. 12-23, 2018.

17. 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.

18. 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.

19. 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. [CrossRef]

20. N. Sneha1 and Tarun Gangil," Analysis of diabetes mellitus for early prediction using optimal features selection", Springer open, 2019, page 1-19 [CrossRef]

21. 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

22. 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. [CrossRef]

23. 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. [CrossRef]

24. 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. [CrossRef]

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

26. 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. [CrossRef]

27. 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. [CrossRef]

28. 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. [CrossRef]

29. 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. [CrossRef]

30. S. M. Mahedy Hasan, M. F. Rabbi, A. I. Champa and M. A. Zaman, "An Effective Diabetes Prediction System Using Machine Learning Techniques," 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), Dhaka, Bangladesh, 2020, pp. 23-28, DOI: 10.1109/ICAICT51780.2020.9333497. [CrossRef]

31. G. Tripathi and R. Kumar, "Early Prediction of Diabetes Mellitus Using Machine Learning," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 1009-1014, DOI: 10.1109/ICRITO48877.2020.9197832. [CrossRef]

34

32. K. L. Priya, M. S. Charan Reddy Kypa, M. M. Sudhan Reddy and G. R. Mohan Reddy, "A Novel Approach to Predict Diabetes by Using Naive Bayes Classifier," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 603-607, DOI: 10.1109/ICOEI48184.2020.9142959. [CrossRef]

AUTHORS PROFILE

Aditya Saxena, is honest, curious, confident, and highly organised, can always keep a positive attitude. Hard working doesn't bother me, but I like it. I enjoy facing challenges that help me learn more and faced more challenges to get positive results oriented. I like everything i do but to be well organised. With an excellent Effective communication skills i can even do as much difficult task as soon as possible, can work with all kinds of people always with a positive attitude. I am good at Multi-Tasking, enjoy meeting new people every day. I really enjoy reading about and learning from marketing success stories from social media site. Some people like me, and I like some people. I break some rules but follow most and used to fun with friend with too much free time. No one is perfect in life, and everyone learn from life and i am also following the same path.

Megha Jain, was born in India on October 5, 1988. She received the Engineering degree, B.E. in Computer Science and engineering from TRUBA group of Institute Bhopal in 2010 and post graduate degree MTech. in Computer Science and Engineering from SATI Vidisha Bhopal, Madhya Pradesh, India, in 2012 and pursuing P. HD from VIT Bhopal. She is currently a Assistant Professor in the Department of Computer Science and Engineering, at Lakshmi Narain College of Technology, Excellence, Madhya Pradesh, India, She possesses 8 years of teaching experience. She has published more than 10 scientific papers in International and National reputed Journals and conference proceedings in the field of image processing, machine learning. She had been student project mentors under Smart India Hackathon.

Prof Prashant Kumar Shrivastava, is a faculty in Department of Computer Science & Engineering in LNCT Group of institutions, Bhopal (MP). He has completed his Masters in Information Technology from LNCT Bhopal in 2011. Presently, he is pursuing P H.D from Oriental University in computer science & Engineering. Indore. He has 12 years of experience in teaching and research in Rustamji institute of technology, BSF, Tekanpur, Gwalior. He has various research publications in reputed national and international journals and conferences. His teaching and research area are included VNET, Ad-Hoc Network, Network Security, Web Designing, Web Analysis, and Web Database.