INTERNATIONAL JOURNAL OF ENVIRONMENTAL SCIENCES
ISSN: 2229-7359
VOL. 11 NO. 23S, 2025
HTTPS://WWW.THEASPD.COM/IJES.PHP

Prophecy Of Diabetic Disease Via Information Scrutiny

Dr.A. Manikandan[1*], P. Sakthivel[2], A. Areef Khan[3], S. Noorjahan[4]

[1*]Professor, Department of CSE, Paavai Engineering College, Namakkal (Dt), Tamil Nadu, India
Manikandanarumugampec@paavai.edu.in
[2]Assistant Professor, Department of CA, Muthayammal College of Arts & Science, Namakkal (Dt), Tamil Nadu, India, 24capsl@muthayammal.in
[3]Assistant Professor, Department of CS & CA, Vivekanandha College of Arts & Science for Women (Autonomous) Namakkal (Dt), TamilNadu, India, areefkhana@vicas.org
[4]Assistant Professor, Department of CS Vysya College, Salem(Dt), Tamil Nadu, India, Snoor2006@gmail.com

Abstract: *Diabetes Mellitus is a most common metabolic disease globally. According to World health organization diabetes is one of the leading causes of deaths in world, and with an estimate of 415 million people suffering from it. According to the studies it is said that by the year 2030 India will have around 1million patients suffering from Type-II diabetes. Diabetes has several complications related to itself .The impact of the diabetes over the various organs such as heart, kidney, eyes, peripheral nerves, blood vessels leads to coronary heart disease, ischemic problems, renal diseases, diabetic foot, and nerve damage. Diabetic complications such as heart and strokes account for 65% of diabetic deaths. More than 30% of the diabetic population suffers from diabetic nephropathy. Prediction of these diabetic complications at an earlier stage would help with the disease management, prevention of prognosis of the disease and help reduce the economic burden. This can be done using the data analysis method. In this paper a novel method to find out the diabetic diseases using the blood parameters of the patients such as HbA1C, serum cholesterol, serum albumin and others is done. The data of 140 patients was used for the predictive analysis and the KNN classifier was used for the analysis. The accuracy was found to be 71.42% using this classifier.*

Keywords-*Diabetic diseases, Disease prediction, Predictive analysis, Blood parameters, Data analysis, KNN classifier*

I. INTRODUCTION

Diabetes Mellitus is a non-communicable and most common metabolic disease, with an estimate of 415 million people suffering from it worldwide. A peptide hormone called insulin produced by the pancreas helps our body in the absorption of glucose from carbohydrates which are taken in the form of food. When the body loses its ability to produce enough insulin it results in a metabolic disease known as Diabetes mellitus popularly known as Diabetes. Diabetes is of three types, Type-1, Type-2, and Gestational diabetes. Type-1diabetesisachroniccondition, where the beta cells that produce insulin is destroyed by the body's immune system and affect insulin secretion. Type-2diabetes is a progressive condition that prevents the usage of insulin in the body, due to this cell cannot utilize insulin efficiently. Gestational diabetes is a condition where a high glucose level in the blood is developed during pregnancy and disappears after giving birth.

For the year 2016, Diabetes was the 7th biggest cause of death, according to the World Health Organization. By 2017, the number of Americans living with diabetes had grown from 108 million in 1980 to 425 million. Type1 diabetes is expected to account for 5 percent to 10 percent of all cases by 2045, while type2 diabetes would account for 80 percent to 90 percent of all cases. There has been a dramatic increase in the percentage of persons over the age of 18 diagnosed with diabetes from 4.7% in 1980 to 8.8% in 2017. According to the study, India is expected to have roughly 1 million Type-II diabetics by 2030. Pre-diabetes affects an estimated 86 million people in the United States. More than 20- 40% of diabetics suffer from diabetes illnesses including diabetic nephropathy. More over 60% of people with type-2 diabetes have diabetic retinopathy, making it one of the most common causes of blindness. Heart disease and stroke claim the lives of diabetics at a rate of 68%. More than half of diabetic patients in India die from heart-related causes. Diabetes has its own complications related to it. The impact of the blood glucose or diabetes over the various organs such as heart, kidney, eyes, peripheral nerves, blood vessels would lead to coronary heart disease, ischemic problems, renal diseases, diabetic foot, and nerve damage.

These complications are difficult to predict at an earlier stage and symptoms are seen at the final stage which increases the risk of a patient's life. Prior to the manifestations of the symptoms at a later stage, many of the bodily parameters change conversely over time. However, due to many factors identification and prediction of diabetic complications at an early stage becomes difficult. In this project our main goal is to predict the diseases

that are caused due to the high sugar level in the blood in diabetic patients by collecting their blood parameters and constructing a computer aided diagnostic model which will help predict the onset of diabetic diseases at an earlier stage. This will help improve the quality of the patient's life and hence reduce the economic burden Therefore; the implementation of a method to predict the complications can be an important step to control the disease. Data analysis is a technique used to inspect, clean, transform and obtained at a model from large set of data to discover the hidden pattern and to draw out useful information. Data analysis is attained in detection and diagnosis of the complications faced by diabetic patients in future. Increased accuracy, cost reduction, and reduction of human resources are the benefits of data analysis in medical environments. Clinical data mining could provide the practitioners with a new method for finding patterns in health data. Diabetics and related complications are diagnosed based on the attributes such as age, fasting blood sugar (fbs), postprandial blood sugar (ppbs), HbA1c, average glucose, serum cholesterol, monocytes, albumin and creatinine. This study is mainly focused on the prediction of diabetic heart and kidney risks. The current system on diabetes disease prediction works on a small dataset.

II. LITERATURE REVIEW

Saeed Farzi et.al [1], in the year 2017 proposed a method in the paper titled "Predicting serious diabetic complications using hidden pattern detection". In this paper they have predicted the complications of type-2 diabetes such as heart diseases, diabetic retinopathy, and diabetic kidney diseases. They have used a data set which contains the data of type-2 diabetic patients. They have used several algorithms and trained and tested them for detecting diabetic complications. They have done this over a dataset of 1000's of patients. They found out that Random Forest algorithm is the best to predict the complications and is useful in health care as a decision making S.Ananthi and V. Bhuvaneswari [2], presented a paper in the year 2017 titled "Prediction of heart and kidney risks in diabetic prone population using fuzzy classification," in this paper they have predicted the risk factors of diabetes. Data mining techniques were applied on patient data. They aimed to develop a fuzzy classifier model to predict heart and kidney complications. The results showed that the risk complication of heart and kidney is high for the diabetic population.

In the year 2017 Dheraj. Shetty, K. Rit, S. Shaikh and N. Patil [3], proposed the study titled "Diabetes illness prediction using data mining" offered an approach. Using Bayesian and KNN (K-Nearest Neighbor) algorithms, they've created a network that can assist doctors forecast diabetes disorders. Specifically, they used the diabetic patient's database and evaluated it using various aspects of diabetes to predict the condition.

B. Shivakumar and S. Alby [4] in the year 2014 presented a paper titled "A survey on data- mining technologies for prediction and diagnosis of diabetes". In this paper they have carried out a survey on different data mining techniques that have been used for prediction of diabetes disease and diabetes data analysis.

T. Jayalakshmi and A. Santhakumaran [5] in the year 2010 proposed a method in the paper titled "A novel classification method for diagnosis of diabetes mellitus using artificial neural networks". This paper aims at various techniques to find missing values to improve the accuracy of classification. It also talks about the preprocessing during the classification. They used Pima Indian Diabetes Dataset and applied the classifier. They found that the results were very much improved when using certain combination of preprocessing techniques. The classification accuracy was found to be 99%.

In the year 2008 J. Han, J. C. Rodriguez and M. Beheshti [6] proposed "Diabetes Data Analysis and Prediction Model Discovery using Rapid Miner," in this research, they collected data on people with and without diabetes. The data mining was carried out with the aid of Rapid Minier. Pre-processing of data, such as identifying and removing outliers, standardising data and numerically discretizing the data sets, visual data analysis and discovering hidden linkages, is carried out in this section finally, a model for predicting the onset of diabetes has been developed. Predicting diabetic patients was made easier by this model.

Chien-Lung Chan et.al [7], in the year 2008 proposed a method in the paper titled "Investigation of diabetic micro vascular complications using data mining techniques". In this they use two data mining techniques such as C5.0 and neural networks to analyse and explore the relationship diabetic complication such as diabetic retinopathy, diabetic nephropathy and diabetic foot problem. They made the analyses on 8736 patients. They found out that creatinine is the most important predictor for diabetic retinopathy. If creatinine is more and other parameters are normal still the patient will suffer from diabetic retinopathy. Diabetic foot problem will occur when the medical parameters of the patient are worse than target values. For diabetic nephropathy for females with high BMI, HbA1C, and AC sugar and males with most uncontrolled blood pressure causes the

diabetic kidney diseases.

III. METHODS

The blood parameters of the diabetic patients are collected. A data bank is created where the collected parameters are stored. Data analysis is done of the acquired data. Pattern recognition is obtained by using the KNN classifier algorithm. Based on the output of the algorithm a decision rule is made.

IV. METHODOLOGY

This study was proposed to predict the diabetic diseases using data analysis by making use of the blood parameters such as age, fasting blood sugar (FBS), postprandial blood sugar (PPBS), HbA1c, average glucose, serum cholesterol, monocytes, albumin and creatinine which were collected from diabetic patients. A repository of information obtained from the subjects was organized and a data bank was created to facilitate information retrieval and to process many continual queries over a long period of time.

The process of Data analysis is done using python programming which is focused more on detecting diabetic diseases such as heart diseases and diabetic kidney related complications.

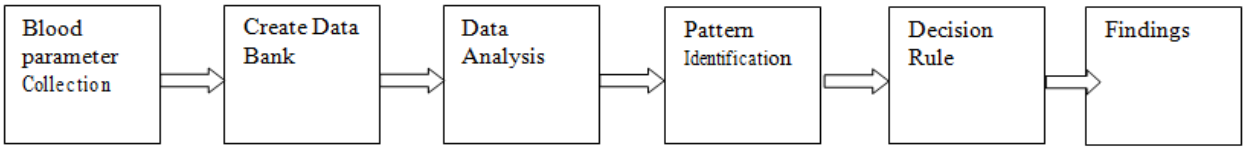

Figure1. Flow chart of the methodology

Fig [1] shows the flow chart of the process in predicting heart and kidney risks in diabetic patients. Primarily, a data bank was created by collecting the blood parameters of around 140 patients from the laboratory with the permission of legal authorities. There are two kinds of data set called continuous and categorical data set, here the data we collected is of continuous type. Data analysis of the obtained set of data was classified into two types, training and testing data set. 70% of the data was used to train the software, whereas 30% of the data was subjected for testing. Pattern recognition was procured by building an algorithm to predict heart and kidney risks, K Neighbor classifier was used to predict the diabetic kidney diseases. The below figure [2] shows the flow chart of the steps in the algorithm used to predict diabetic complications using KNN classifier. Firstly required blood parameters from the patients are collected and stored in a data base, and then KNN classifying algorithm is applied to the stored data set. Later a threshold value is set for the parameters which were taken by consulting health experts.

Table in figure [3] shows the normal values and abnormal values of each blood parameter chosen for this project and their respective complications. These threshold values are necessary for the algorithm to give an accurate prediction of the complications.

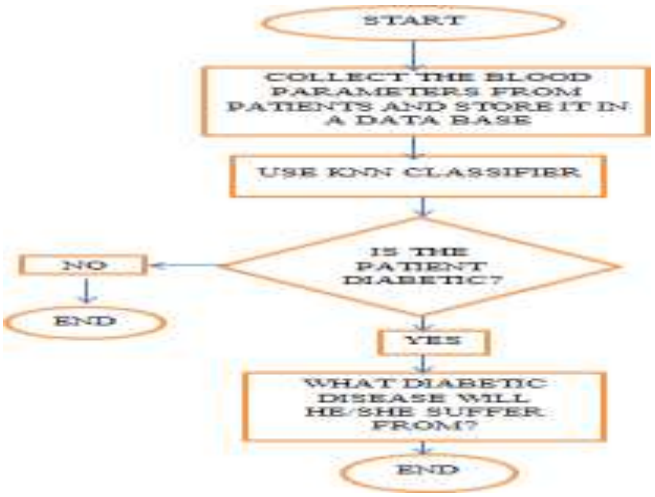

Figure 2. Flow diagram of the algorithm

If the patient is found to be diabetic, then the algorithm detects it as abnormal and later the complications are found out. The other parameters of heart and kidney diseases are checked by the algorithm, if the values of those parameters are above the threshold level or normal level the algorithm classifies the patient into respective diseases they will be suffering from.

Figure 3. Table showing the normal and abnormal values for each parameter and their complications

Tests	Normal values	Mid-range	High values	Complications
HbA1C	4-6	7-9	10-13	Heart diseases, stroke
Serumcreatinine	0.7-1.3 mg/dL	-	Greaterthan1.3	Diabetic kidney diseases
Serum albumin	3.4-5.4g/dL	-	Lowerthan3.4	Mild kidney disease
Serumcholestrol	Lessthan 200mg/dL	-	Higherthan 200mg/dL	Heart diseases or stroke
Monocytes	2-8%	-	Greaterthan8%	Arterogenis is (heart diseases)

V. DATA ANALYSIS

Data analysis is a method used to find out hidden patterns from the set of data that can provide useful information. This paper deals with predictive analysis, where blood parameters of diabetic patients are collected to predict the art and kidney related complications. The two prominent methods in data analysis are quantitative research and qualitative research with their own technique. Data bank is formed from the collected data which is organized and then processed for data cleaning. Once the data is cleaned, it ready to be analyzed by applying a technique referred as exploratory data analysis. We have coded for the KNN classifier algorithm to make sure the accuracy of prediction, and thus obtain the desired results. Decision is made based on the output.

VI. RESULTS AND DISCUSSIONS

After the data was collected and stored in a data base, as shown in figure (4) the predictive analysis on the collected data was done using the KNN classifier for 140 instances. KNN algorithms perform classification using the new data points based on similarity measurements. The algorithm grouped the data set into 20 set neighbour with 5 subjects each. Classification is done by a majority vote to its neighbors. The data is assigned to the nearest neighbours, if we increase the number of neighbours the accuracy of the model increases.
By using the KNN classifier we got an accuracy of 71.42%. The data set was grouped into 20 set neighbour with 5 subjects' data each. The accuracy was found at the 5th neighbour as shown in the graph in figure (5).
A threshold value for each parameter was set by consulting health experts, the algorithm made use of the threshold values and prediction of the diabetic complications was made.

Age	FBS(mgs/dl)	PPBS(mgs/dl)	HBAC(%)	Avg.Glucose	Monocytes(%)	Serum Cholestral	Serum Creatine	Album
70	115	191	7.18	171	3	239	1	2.4
67	103	159	5.12	126	4	183	0.7	4
38	125	170	6.37	97	2	171	1.1	3.7
64	110	187	6.9	149	4	190	1.4	3.9
44	179	240	7.78	200	8.7	407	0.7	1.9
65	195	272	7.29	218	3	234	0.9	3.2

Figure 4. Pre-processed dataset used for analysis

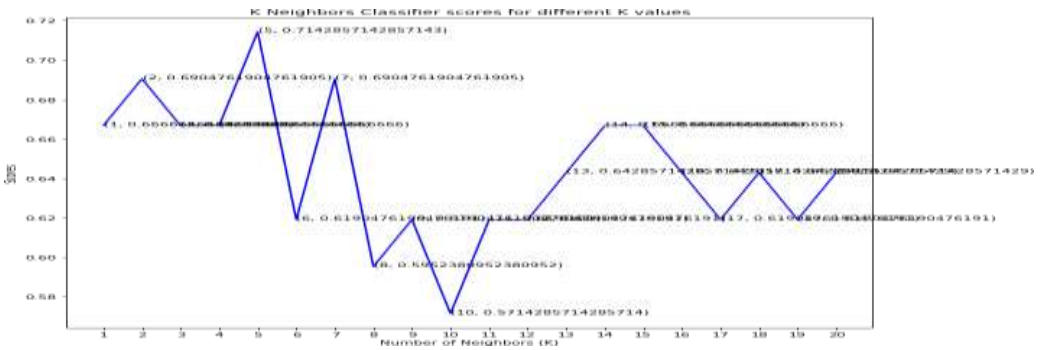

Figure 5. Graph showing KNN classifier scores for different values of K

VII. CONCLUSION

In this paper, KNN classifier is used to analyze and predict the diabetic complications such as heart and kidney diseases. Generally this model uses the medical parameters such as blood glucose levels, HbA1C, serum albumin, serum creatinine and so on. In the selected population of patients that was used to predict the diabetic diseases the model's performance accuracy was 71.42%. The experimental results showed that around 40% of the patients suffer from heart diseases, about 50% patients suffer from both heart and diabetic kidney diseases and only 5% suffer only from diabetic nephropathy and others are non-diabetic and didn't suffer from any complications. This model can be helpful in the field of health care as a decision support system for health worker and physicians. The accuracy of the prediction of risk complications can be increased by using a large set of data and using different more attributes.

Limitations and Future Studies

Our study has some limitations. This study reports data from single source, therefore, our findings was established with small set of data. Furthermore, this technique can be used with larger set of data to obtain more accuracy.

REFERENCES

1. S.Farzi, S.Kianian and I.Rastkhadive, "Predicting serious diabetic complications using hidden pattern detection," 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI),Tehran, 2017, pp. 0063-0068.
2. S.AnanthiandV.Bhuvaneswari," Prediction of heart and kidney risks in diabetic prone population using fuzzy classification," 2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, 2017
3. D.Shetty, K.Rit, S.Shaikh and N.Patil, "Diabetes disease prediction using data mining," 2017International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore,2017.
4. B.Shivakumar and S.Alby," A survey on data-mining technologies for prediction and diagnosis of diabetes," in Intelligent Computing Applications (ICICA),2014 International Conference on, 2014, pp.167-173.
5. T.Jayalakshmi and A.Santhakumaran, "A novel classification method for diagnosis of diabetes mellitus using artificial neural networks," in Data Storage and Data Engineering (DSDE), 2010 International Conference on, 2010, pp. 159-163.
6. J.Han, J.C.Rodriguez and M.Beheshti, " Diabetes Data Analysis and Prediction Model Discovery Using Rapid Miner," 2008 Second International Conference on Future Generation Communication and Networking,HainanIsland,2008
7. Chien-Lung Chan, Yu-Chen Liu and Shih-Hui Luo, "Investigation of diabetic micro vascular complications using data mining techniques," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp.830-834.
8. Bengio, Y., and Grandvalet, Y. (2005). Bias in Estimating the Variance of K -Fold Cross-Validation. New York, NY: Springer, 75–95. https://doi:10.1007/0-387-24555-3
9. Habibi, S., Ahmadi, M., and Alizadeh, S. (2015). Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining. Glob. J. Health Sci. 7, 304–310. https://doi:10.5539/gjhs.v7n5p304
10. Han, L.,Luo, S., Yu, J., Pan, L., and Chen, S. (2015). Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes. IEEE J. Biomed. Health Inform. 19, 728–734. https://doi:10.1109/JBHI.2014.2325615
11. Liao, Z. J., Wan, S., He, Y., and Zou, Q. (2018). Classification of small GTPases with hybrid protein features and advanced machine learning techniques. Curr. Bioinform. 13, 492–500.https://doi:10.2174 /1574893612666171121162552
12. Liao, Z., Ju, Y., and Zou, Q. (2016). Prediction of G protein-coupled receptors with SVM-Prot features and random forest. Scientifica 2016:8309253. https://doi:10.1155/2016/8309253
13. Lin, C., Chen, W., Qiu, C., Wu, Y., Krishnan, S., and Zou, Q. (2014). LibD3C: ensemble classifiers with a clustering and dynamic selection strategy. Neuro computing 123, 424–435.https://doi:10.1016 /j. neucom.2013.08.004
14. .Su, Z. D., Huang, Y., Zhang, Z. Y., Zhao, Y. W., Wang, D., Chen, W., et al. (2018). iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. Bioinformatics https://doi:10.1093/bioinformatics/bty508 [Epub ahead of print].
15. Tang, W., Wan, S., Yang, Z., Teschendorff, A. E., and Zou, Q. (2018). Tumor origin detection with tissue-specific miRNA and DNA methylation markers. Bioinformatics 34, 398–406.https://doi:10.1093 /bioinformatics/btx622