iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4094

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

General Approach of Prediction of Diabetes Mellitus using Top-K and MOPNAR Algorithm

Yogita S. Wanjari[1], Prof. Gaurav Y. Kawade[2]
[1]Research Scholar(M. Tech), [2] Head of the Department (HOD), Computer Science and Engineering
G. H. Raisoni College of Engineering, Nagpur, India

Abstract: ***Today's world is the digital world all the people are surrounded with the electronic or digital gadgets. Due to this the life style of people is getting change; the individual work is replaced by the electronic gadgets, and from that lot of diseases are caused, in between these diseases some are very usual like sugar, hypertension, blood pressure etc. And some are very dangerous but these usual diseases are not so easy to identify and easy to recover. In this paper we are focusing on the diabetes mellitus diseases which are very common but its effect are very on the human body. The aim of this paper is early detection of diabetes because if such diseases are early detected then it is very easy to diagnosis and it helps patients for the early recovery, for that in this paper the data mining technique are used on the diabetic patients dataset and then applying the mining algorithm for identifying the diabetic patients the algorithms such as TOPK and MOPNAR for early detection of diabetes.***
Keywords: Diabetic Mellitus, Electronic Gadgets, Data Mining, TOPK Algorithm, MOPNAR Algorithm.

I. INTRODUCTION

Diabetes is very common and life threading disease in today's world. According to the US survey near about 28.8 million of people are affecting from this diabetic mellitus disease that means 8 % of its population. It is very common disease and anybody can suffer from this diseases there is no such special symptoms and body mark to identifying it [4] [7]. And in this the main thing is that the many people are not aware about that they suffer from such dangerous disease. This is caused because in the today's century the people are very digitalizes they done all their work with help of the machines that's why the life style of human is totally changed and the people are grabbed with the lots of diseases. These diseases are very effective for the human body and for that it is essential to early identification of such diseases. Because if it early identify then it is easy to save human life, if it early identify then it helps for the recovery and for diagnosis. For early detection of such diseases the data mining techniques are used but before using it, focus on the different keys challenges and approach to improving it. When we deal with all these difficulties then it is possible to handle with it [2]. By the means of US review the diabetic mellitus takes more lives than the Cancer and AIDES, and it is the seventh leading reason of death in the United State [6]. That's why to early identification of such disease is the today's medication need, and this is the main aim of this paper. Now look towards some major factor from that the diabetes can caused such as,

A. What's the Reason to Caused Diabetes

Diabetes can be caused anyone during its any phase of life. The disease called diabetes cannot specially be identified that means it is very common and usual disease that can be caused by anyone in any walk of life [6]. But due to the presence of diabetes some other diseases can be caused which may be very dangerous for the human body such as hit, mental disability, skin complication, retinopathy, heart diseases, stroke, and neuropathy etc. The person who is suffering from diabetes has more chances to have such kind of diseases that's why it is called as co-morbid diseases. But the appropriate management of patients with risk and changed life cycle and the medication can decrease the risk of 30% to 60% [7][8][4]. The body metabolism is depends on the food which is generated by the human body it is also called as energy, the diseases diabetes is the condition in which the body can not generate food properly. The food created by human body gets converted into sugar, glucose etc. As we know the human body is made up of so many cells and tissues, there is organ near to stomach called pancreas which secret hormone called insulin. This hormone insulin play a key role to open a cell and make a way for glucose to enters into the cell for the entire work of the body. The person who is suffer from this disease means the process of hormone secretion is not done properly or the secreted hormone such as insulin is not used in proper manner. Because of it the generated energy is not used totally and the energy is converted into sugar that means the sugar level in the body is gradually increase as compare to the normal requirement and such patient is said as "diabetic" or "high blood sugar" patients [6].

©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

B. Types and Symptoms of Diabetes

There are two types of diabetes first one is the type 1 and the second one is type 2. Type 1 is cause due to the improper secretion of insulin in the body. This type of diabetic can cause in early phase of life that means in the teenage or in adult age or before 45 year. The percentage type 1 diabetic people is approximately 5-10%.

Type 2 diabetic can cause due to the not enough secretion of insulin which is required for the body metabolism. Type 2 diabetic has 90-95 % chances to diagnosis. This type 2 diabetes can caused during the pregnancy it has only 5-6 % chances and it is automatically disappeared after pregnancy over. This type can be cured by doing exercise, diet and home blood sugar test. The risk factor of type 1 is less than type 2.

The diabetes can be very common disease. It does not have expressive or special identified symptoms. It can be identified by using the following symptoms such as,

1) Sudden change in weight
2) Change in eye vision (blurred)
3) Breath- smell in acetone
4) Felling restless
5) Change the BP level in blood[6]

C. Problem Definition, Objective and Scope of Proposed System

The change in life style is the main thing to caused such kind of diseases, but the main problem is that the people who suffering from this disease they do not have any idea that they have such kind of diseases. To solve this problem the expert system is going to design which early identify the disease along with their co-morbid diseases also, if the early detection of diabetes is done then it is very helpful for the medication and the chances to cure is automatically increase.

The traditional system of diabetes detection is depends on the lots of attributes this system was rule base system it generate lots of rule for the detection of diabetes. But the main thing is that these all attributes are not necessary for the detection that's what the created rules are also un- necessary and it is repeatedly occurred which is not needed.

To overcome these entire drawbacks is the main objective of this system, and also increase the accuracy of detection of diabetes as compare to the previous one. For all these the Top-K and MOPNAR algorithm is used here.

The scope of this system is as follows,

1) This system can be used as expert system in hospitals.
2) It can also used in pathology where the test is done for the any kind of diseases detection.

II. METHODOLOGY

A. Related Work

Any disease can be identified earlier then it is very beneficial to the person who is servings from that disease and physician also. Because if disease can be identifying early then the treatment according to the disease will early started and it increase the chances to cure the patients from this disease early. Suppose the disease like cancer if it identify early then there is more chances of patients to live or cure from that disease, but if it is not identify early then very critical condition for that identifying the risk factor of that disease is very necessary. To do this previously some techniques are used which are association rule summarization techniques,

©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

there are number of association rule summarization techniques but these all are not in the used only some of them are used which gives the result in there applicability and strength[9], these techniques are[7],

1) Survival analysis
2) Association rule mining
3) Distributional association rules
4) Distributional association rule for survival outcomes

In the traditional system these four techniques are used for the detection of diseases. But all the rules generate by these techniques are not in used because it may be repeated or co-occurred. And these techniques generate the lots of rules, so in the proposed system apply the rule summarization technique such as Top-k and MOPNAR algorithm which generate minimum rules but the accuracy and the output is very predictive as compare to the previous.

B. Proposed Plan of Work

Diabetes is the life threading disease this disease is the major part for the occurring of the other disease, so there is needs to detect such life threading diseases early. For the detection of such disease which techniques are used previously these are mention in the previous section, now totally discuss about the plan of the proposed system. In the proposed system the Top-K and MOPNAR algorithm are used for the detection of diabetes and the co-morbid diseases which are caused due to the presence of diabetes, if we want to take relief from the co-morbid diseases the first to detect and give proper medication for the diabetes. So for that discuss the flowchart of the proposed system which is as follows.

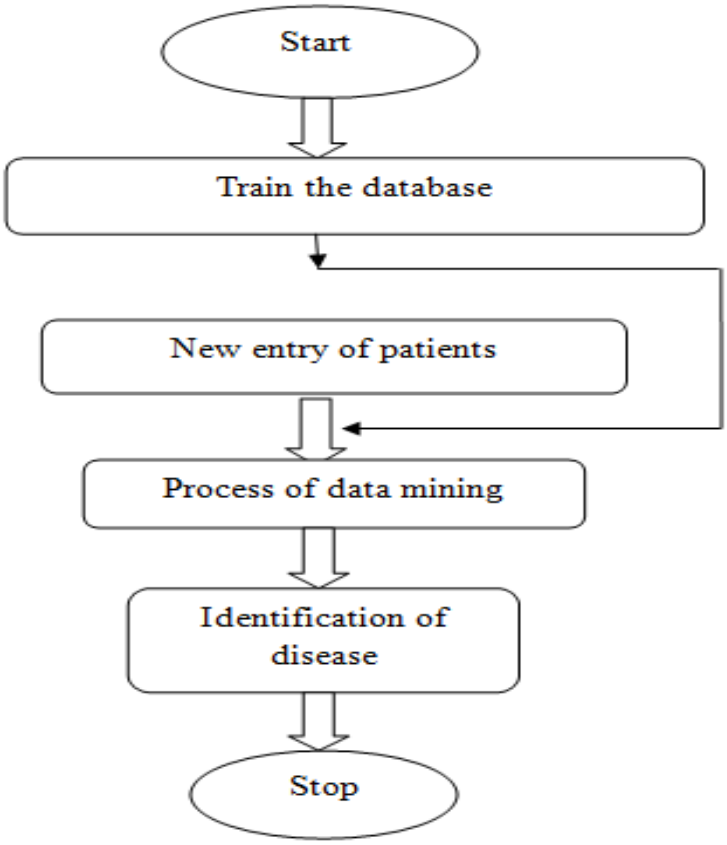


Fig 1: Flowchart of proposed system

According to the flowchart of the proposed system first phase is database training which means to train the database as required and then take the new entry and compare it with the trained database after that apply the mining techniques, these are the rule summarization techniques such as Top-k and MOPNAR algorithms, which gives the required output that the detection or identification of diabetes. In the next section briefly discuss the whole process of identification of diseases.

©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

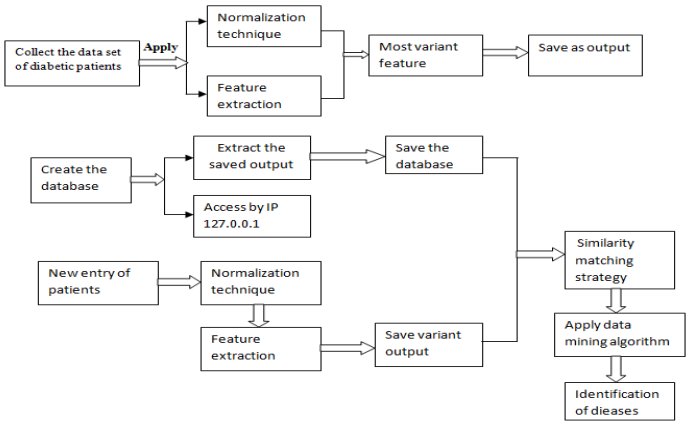


Fig -2: Block Diagram of Proposed System

C. Implementation

In the above section we discuss about the flowchart of the proposed system, now we detail discuss about the whole process of proposed system. For that firstly take the dataset of diabetic patients which is easily available in the UCI repository site, the dataset present here are real and these dataset mostly relevant to the US hospitals, that means these are real patients entries which are in UCI dataset. Or it can also be called as EMR means electronic medical records in this the patients' records are electrically stored in the computer system so that if it required for someone it is easily available [1]. The input dataset of this project is similar to the fig. 3. This is having the 11 attributes such as age, sex, BMI, BP and six sugary components present in the human blood which are as follows.

AGE	SEX	BMI	BP	S1	S2	S3	S4	S5	S6	Y
59	2	32.1	101	157	93.2	38	4	4.8598	87	151
48	1	21.6	87	183	103.2	70	3	3.8918	69	75
72	2	30.5	93	156	93.6	41	4	4.6728	85	141
24	1	25.3	84	198	131.4	40	5	4.8903	89	206

Fig -3: Input Dataset

Then apply the feature extraction technique and normalization techniques on the dataset due to this in the output window the maximum value of each column and the normalized value of all the columns are getting, these normalized value are range in between the 0- 1, then by applying the feature extraction technique we get the most variant features these are the features which are frequently change due to this most variant features the original dataset is compress and stored this compress dataset as output. Now save this dataset into database for saving the dataset we used MySql database i.e. WAMP server. Save all entries of output file of dataset in the database. To access the server first open the browser and type the 127.0.0.1 link to open the wamp server and to show the created database. Take the new entry of patients and apply the same procedure such as feature extraction and normalization technique and then save the output file for these new entries. Then apply the similarity matching strategy with the help of that strategy evaluate the new dataset with the saved database. Finally apply the data mining algorithm i.e. Top-k and MOPNAR to detection of diabetes and it gives the proper output regarding to it. It shows that which patients belong in which class that is in the diabetic or non-diabetic. From this it is clear that the patients are diabetic or not. This system is very helpful for the other also in the dialectologist hospital or in the pathology to early detection of diabetes.

III. COMPERATATIVE ANALYSIS

In this comperatative analysis phase compare the result of or output of (in the error factor manner) two systems first is the previously developed rule base system and the second one is the newly developed system i.e. an expert system.

©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

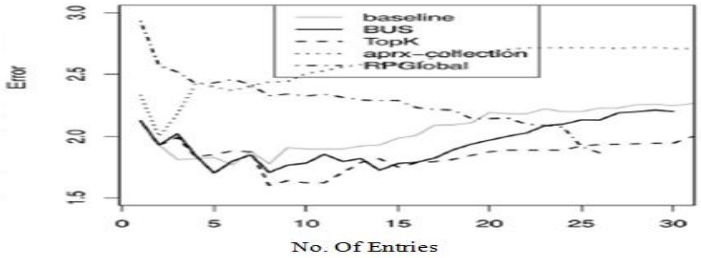


Fig -4: Result of rule base system

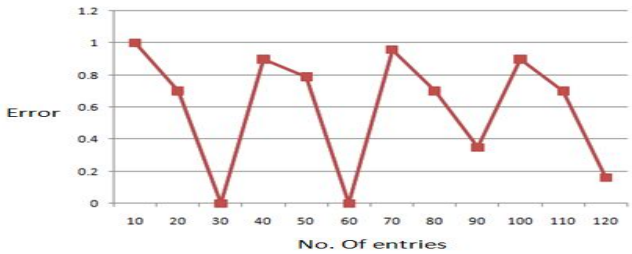


Fig -5: Result of expert system by using Top-K and MOPNAR algorithm

While developing this project one of the main objectives is to increase the accuracy of the system. In the previous system result is based on the error factor and the number of entries means the number of patients, and the graph is regarding to it is shown above. To compare the previous system result with the proposed system the elements on the both axis must be same so it is. As comparing the result of both system then it is clearly visible that in the previous system the error factor is in between 1.5 to 3 and here lots of algorithms are applied, then result is gradually increased for each technique.

As seen the result of proposed system the error factor is vary in between the 0 to 1 which is very less than the previous one. If the error factor is less that means the accuracy is more as compare to the previous system, so it can be said as the proposed system is better than the previous one by using the summarization techniques.

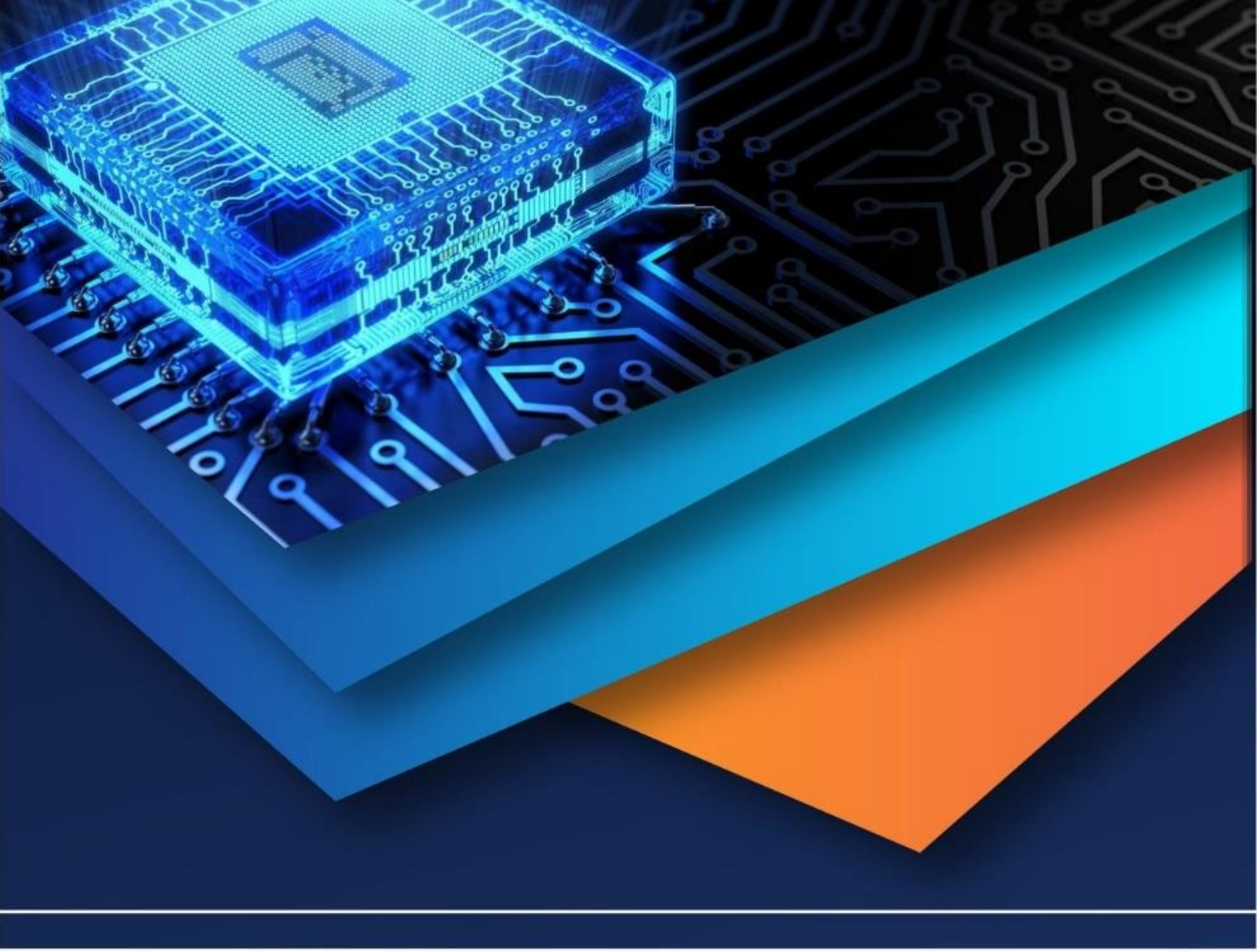
IV. CONCLUSIONS

From the discussion which is done in the section 3 it can be said that the proposed system is better than the previous one. There in the previous system the lots of algorithm is used for the detection of diabetes as compare to the new one and the error regarding to it is also more than the proposed system. So it can be conclude from that the new system is more accurate and better than the previous system.

REFERENCES

[1] G Rahini, P. Dinesh Kumar,"Electronic Medical Record for Diabetes Mellitus Using Association Rule Mining," International Journal of Advance Research in Biology Engineering Science and Technology , vol 2 Issue 15, March 2016, pp. 247-252.
[2] N. Malmbo, "Data Mining: Techniques,Key Challenges and Approaches for Improvement," International Journal of Advance Research in Computer Science ans Software Engineering, vol. 6, Issue 3, March 2016, pp. 59-65.
[3] Swati Gupta, A. M. Karandikar, "A Survey on Methods of Automatic Deyection of Diabetic Reteinopathy," vol 5, Issue 1. January 2015, pp. 1-7.
[4] Priya R., Roshma R., "Pridiction of Co-morbid Condition Association With Diabetes using Split and Merge Algorithm." International Journal of Innovative Research in Computer and Communication Engineering, vol 3, Issue 7, July 2015, pp. 6716-6722.
[5] https://www.cdc.gov/media/presskits/aahd/diabetes.pdf
[6] https://www.diabetesreasearch.org
[7] Gyorgy J. Simom, Pedro J. Caraballo, Terry M. Therneau, Steven S. Cha, M. Regina Castro, and Peter W. Li,"Extending Association Rule Summarization Techniques to Access Risk of Diabetes Mellitus" IEEE Transaction on knowledge and data engineering, Vol. 27, No. 1, January 2015, pp. 130- 141,.
[8] V. Kavitha, R. Mohan, "ARS: Association Rule Summarization Techniques to Detect Risk Of Diabetes Mellitus," International Journal Of Research In Computer Application And Robotics, Vol. 2, Issue 11, , November 2014, pp. 145-149.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)