

PREDICTION OF DIABETES DISEASE WITH COMPARITIVE STUDY IN DATA MINING ALGORITHMS

Sarikonda Vishali¹
Dr. Ramakrishna Seemakurti²

Abstract:

Diabetes analysis is done by detection of knowledge based on different patterns views from previous research areas and medical data reporter's analysis. There are several algorithms which are being exercised in different areas with several a examples in data mining concepts such as frequent times non linear relationships between independent and dependent variable classification, analysis, regression models are general good examples for supervised data mining algorithms using algorithms such as SVM,C4.5,PNN,K-NN,PLS-DA,PLS-LDA.the data evolved in from all these comparative study analysis report is time precision, data evaluation, bootstrap validation, error checking, positive, negative results. A very typical confusion are being raised by many further researches all over the world in clinical studies. This paper focuses on quick study for algorithmic decision taken to detect disease from UCI with control data set recovery models using mining tool possibility detection in early stage.

Key Words: SVM (support vector machine), C4.5,PNN,K-NN,PLS-DA,PLS-LDA.

Introduction:

The diabetes is causing blindness, heart strokes, renal failures many more diseases and causing to early stage deaths of human beings all over the world due to several factors such as:

1. **Food Habits:** To bring awareness in high consumption of salt in spicy food, pickles and other recipes among Indians. 'Health is Wealth' is possible only through awareness& action. Indian Govt is spending more to improve health care facilities (AROGYA SRI)
2. **Couch Potato's:** In new generation, Indians are addicted to TV-serials, CRICKET matches and movies on TV Screens for more hours in a day. Private Schools are without play grounds and children are resorted to Video games. On line shopping services made things difficult for no physical marketing. Hectic schedules made bicycles missing on roads. All these activities give a scope of symptoms that lead to rise of diabetes in India.

¹ Research scholar, computer science, JJTU, Rajasthan, India, 333001

² Principal, Ph.D(SVIT, Secundrabad , 500003 India

3.Health Care Costs: The development of a nation is not based on financial health, but of individual health of its citizens. Timely identification and prevention methods reduce costs in health care for individual and society.

4.Alternatives: Adapt alternative solutions to reduce stress by modifying the working environment of Banks, insurance, Software sectors without hampering the services. Allow compulsory break for 10-15 minutes in every 2-3 hrs span to have physical movement & fresh air. Health of human resources is an intangible asset to any service sector organization.

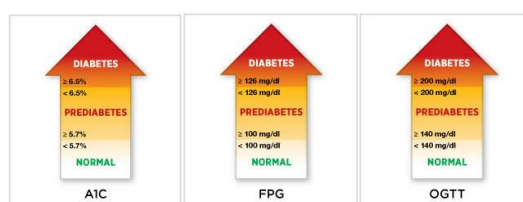
5.Emergency Services: 108 services- Nearest physicians must extend first aid to the patient. Due to traffic & unpleasant roads, Ambulances are finding it difficult to reach to needy ones in time.

6. Research Facilities: By increasing research facilities& awareness, Govt of India could able to eradicate 'Cholera' & 'POLIO' with help of WHO. 'Diabetes' should be next one.

Diagnosing Diabetes and Learning about Pre-diabetes

There are several ways to diagnose diabetes. Each way usually needs to be repeated on a second day to diagnose diabetes.

Testing should be carried out in a health care setting (such as your doctor's office or a lab). If your doctor determines that your blood glucose level is very high, or if you have classic symptoms of high blood glucose in addition to one positive test, your doctor may not require a second test to diagnose diabetes.



A1C

The A1C test measures your average blood glucose for the past 2 to 3 months. The advantages of being diagnosed this way are that you don't have to fast or drink anything.

- Diabetes is diagnosed at an A1C of greater than or equal to 6.5%

Result	A1C
Normal	less than 5.7%
Prediabetes	5.7% to 6.4%

Diabetes	6.5% or higher
-----------------	----------------

Fasting Plasma Glucose (FPG)

This test checks your fasting blood glucose levels. Fasting means after not having anything to eat or drink (except water) for at least 8 hours before the test. This test is usually done first thing in the morning, before breakfast.

- Diabetes is diagnosed at fasting blood glucose of greater than or equal to 126 mg/dl

Result	Fasting Plasma Glucose (FPG)
Normal	less than 100 mg/dl
Prediabetes	100 mg/dl to 125 mg/dl
Diabetes	126 mg/dl or higher

Oral Glucose Tolerance Test (also called the OGTT)

The OGTT is a two-hour test that checks your blood glucose levels before and 2 hours after you drink a special sweet drink. It tells the doctor how your body processes glucose.

- Diabetes is diagnosed at 2 hour blood glucose of greater than or equal to 200 mg/dl

Result	Oral Glucose Tolerance Test (OGTT)
Normal	less than 140 mg/dl
Prediabetes	140 mg/dl to 199 mg/dl
Diabetes	200 mg/dl or higher

Random (also called Casual) Plasma Glucose Test

This test is a blood check at any time of the day when you have severe diabetes symptoms.

- Diabetes is diagnosed at blood glucose of greater than or equal to 200 mg/dl

What is Prediabetes?

Before people develop type 2 diabetes, they almost always have "prediabetes" — blood glucose levels that are higher than normal but not yet high enough to be diagnosed as diabetes.

Doctors sometimes refer to prediabetes as impaired glucose tolerance (IGT) or impaired fasting glucose (IFG), depending on what test was used when it was detected. This condition puts you at a higher risk for developing type2diabetes and cardiovascular disease.

No Clear Symptoms

There are no clear symptoms of prediabetes, so, you may have it and not know it.

Some people with prediabetes may have some of the symptoms of diabetes or even problems from diabetes already. You usually find out that you have prediabetes when being tested for diabetes.

If you have prediabetes, you should be checked for type 2 diabetes every one to two years.

Results indicating prediabetes are:

- An A1C of 5.7% – 6.4%
- Fasting blood glucose of 100 – 125 mg/dl
- An OGTT 2 hour blood glucose of 140 mg/dl – 199 mg/dl

Preventing Type 2 Diabetes

You will not develop type 2 diabetes automatically if you have prediabetes. For some people with prediabetes, early treatment can actually return blood glucose levels to the normal range.

SIGNIFICANCE OF RESEARCH

1. Encourage Individual & Social Health Care through Medical Informatics:
 - (i) Establish diabetes as priority health concern among youth & elder people.
2. To bring awareness of symptoms (diabetes) through patient education (Basic health care)
 - (i) Individual lifestyle requirements and meal planning.
 - (ii) Importance of regular exercise in treatment.
 - (iii) Interaction of food intake, physical activity and oral insulin
 - (iii) Self-monitoring of blood or urine glucose
 - (iv) Life Protection in emergencies like stress
3. To increase Scope of research to solve a problematic of decision support in medical field, a theoretical approach is developed to integrate CBR (Case Based Reasoning) system with various data mining tools. It enables ideas to draw conclusions from pattern

WORK PLAN

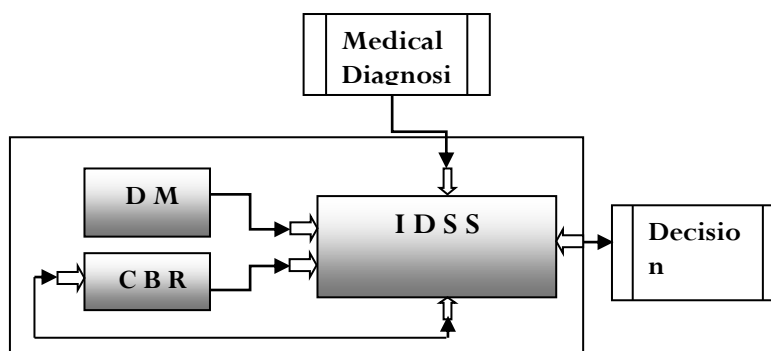


Figure : The adopted Decision Support System using data mining algorithms.

2. BACKGROUND

Diabetes is a lifelong chronic condition that affects the human body by reducing the insulin which carries glucose into the blood cells. This increases the sugar level in the body leading to different complications like stroke, heart disease, blindness, kidney failure and death. Diabetic patients generally have the following symptoms.

- Increased thirst
- Frequent urination
- Weight loss
- Increased hunger
- Slow-healing infections
- Blurred vision
- Nausea and Vomiting
- The following medical tests are used to diagnose the diabetic mellitus [3]
- Urine test
- Fasting blood glucose level
- Random blood glucose level
- Oral glucose tolerance test
- Glycosylated hemoglobin(HbA1c)

Attribute no	Attributes to be tested	Symbols
1	Diastolic blood pressure measured in mmHg	pres
2	Glucose tolerance test to find the plasma glucose level concentration in saliva.	plas
3	Body Mass Index (BMI)	mass
4	Diabetes pedigree function	pedi
5	Skin rashes and thickness fold in mm(Triceps)	skin
6	Age	years
7	2- hour Serum Insulin in mu U/ml(INSULIN)	insu
8	Diabetes Class Variable	binvar

4. PERFORMACE COMPARISONS OF ALGORITHMS

Classification accuracy (CU) is the most common method used for evaluation of performance. Calculation of accuracy is performed by taking ratio of truly classified samples (true negative, true positive) to the total number of samples.

Accuracy = truly classified samples/total samples (3)

Another evaluation methods used for measuring performance are Sensitivity and Specificity. Sensitivity is calculated by dividing the true positive (TP) samples to the sum of true positive (TP) and false negative (FN) samples.

Sensitivity = $TP / (TP + FN)$ (4)

Specificity is calculated by dividing the true negative (TN) samples to the sum of true negative and false positive (FP) samples.

Specificity = $TN / (TN + FP)$ (5)

KNN have the sensitivity value 75% and specificity value 58%. Amalgam-KNN has the highest sensitivity and specificity. The Table 2 shows the performance accuracy of different algorithms. EM algorithm is having the least accuracy rate and it exhibit inaccuracies when high dimensional data sets are given as input. By analyzing the table it is clear that amalgam KNN and ANFIS algorithm with adaptive KNN shows the maximum accuracy compared to other algorithms. To achieve the better accuracy, results were analyzed for different k values.

Table 2. Performance analysis

Method	Accuracy
EM algorithm	<70%
KNN algorithm	73.17%
K means algorithm	66-77%
Amalgam KNN algorithm	>80%
ANFIS algorithm with adaptive KNN	80%

5. RECOMMENDATIONS

ANFIS and Adaptive based KNN algorithms perform the classification with a higher efficiency and reduced complexity. Amalgam KNN extracts both the feature of KNN and K means algorithms. Since the current accuracy of both ANFIS with adaptive based KNN and amalgam KNN is greater than 80 %, they can be combined to produce a better accuracy algorithm than the exiting one. Finally Co- active ANFIS which combines both the features of adaptive neural networks and fuzzy systems which is termed as the successor of ANFIS can also be used to improve the performance of the present algorithm. Their conjoint dependence provides astonishing abilities for learning. CANFIS provides non liner rules for classification, prediction and diagnosis among the input output pairs. Results should be compared and tested with the increased values of k. Greater the k value, more will be accuracy rate.

6. CONCLUSION

Data mining and machine learning algorithms in the medical field extracts different hidden patterns from the medical data. They can be used for the analysis of important clinical parameters, prediction of various diseases, forecasting tasks in medicine, extraction of medical knowledge, therapy planning support and patient management. A number of algorithms were proposed for the prediction and diagnosis of diabetes. These algorithms provide more accuracy than the available traditional systems. This paper includes algorithms like Expectation Maximization Algorithm, K Nearest Neighbor algorithm, K-means algorithm, Amalgam KNN algorithm and Adaptive Neuron Fuzzy Inference System algorithm. From the observation EM possess the least classification accuracy and amalgam KNN and ANFIS provide the better classification accuracy results. Amalgam KNN comprises both the feature of KNN and K means. ANFIS in cooperates both the features of adaptive neural both ANFIS and amalgam KNN is used. Co active ANFIS was extended with some capabilities of its predecessor ANFIS to provide better classification and prediction accuracy. Classification shows better accuracy when the k value is increased to a large value.

References

- Peleg, M., Tu, S., Bury, 2003. Comparing models of decision and action for guideline-based decision support: (JAMIA)(1), 52–68.
- Ramakrishnan, N., Hanauer, D., Keller, B., 2010. Mining electronic health records. IEEE Computer Society 43 (10), 77–81.
- Raza Abidia, Syed Sibte, Manickamb, Selvakumar, 2002. Leveraging XML-based electronic medical records to extract experiential clinical knowledge: an automated approach to generate cases for medical case-based reasoning system.