



DIABETIC RISK LEVEL OF A PATIENT WITH A BETTER ACCURACY BY USING PAC AND ANN

¹ E.Aruna , ² Rudrangi Rohan , ³ Chanda Arun Kumar, ⁴ Venuturumilli V S Sai Vamsi Krishna

¹ Assistant Professor, Department of Information Technology, Teegala Krishna Reddy Engineering College
Hyderabad, Telangana, India.

¹ aruna.etikalas@gmail.com

^{2,3,4} UG Scholars Department of Information Technology, Teegala Krishna Reddy Engineering College ,
Hyderabad, Telangana, India.

² rudrangirohan112@gmail.com , ³ chandaarun1234@gmail.com , ⁴ venturumilli2001@gmail.com

Abstract

The diabetes is one of lethal diseases in the world. It is additional an inventor of various varieties of disorders foe example: coronary failure, blindness, urinary organ diseases etc. In such case the patient is required to visit a diagnostic center, to get their reports after consultation. Due to this they have to invest their time and currency. But with the growth of Machine Learning methods, we have got the flexibility to search out an answer to the current issue, we have got advanced system mistreatment information processing that has the ability to forecast whether the patient has polygenic illness or not. Furthermore, forecasting the sickness initially ends up in providing the patients before it begins vital. Information withdrawal has the flexibility to remove unseen data from a large quantity of diabetes associated information. The aim of this analysis is to develop a system which might predict the diabetic risk level of a patient with a better accuracy. Model development is based on categorization methods as Passive Aggressive Classifier algorithm (), Artificial Neural Network (ANN), and Support Vector Machine algorithm (SVM). For PAC, the models give precisions of 78%, 87% for Support Vector Machine and for ANN 89.5% which is highest among three algorithms. Outcomes show a significant accuracy of the methods.

1.INTRODUCTION

Diabetes is a situation which causes deficiency due to less amount of insulin in the blood. Warning sign of high blood sugar results in frequent urination, feeling thirsty, increased hunger. If it is not medicated, it will lead to many difficulties. When there is a rise within the sugar level within the blood, it is referred to as prior diabetes. Various information mining algorithms presents different decision support systems for assisting health specialists. The effectiveness of the decision support system is recognized by its accuracy. Therefore, the

aim is to build a decision support system to predict and diagnose a certain disease with extreme amount of precision. The AI consists of ML which is its subfield that resolves the real-world difficulties by "providing learning capability to workstation without supplementary program writing.

Diabetes is a most common disease caused by a group of metabolic disorders. It is also known as Diabetic mellitus. It affects the organs of the human body. It can be controlled by predicting this disease earlier.



If diabetics patient is untreated for a long time, it may lead to increase blood sugar.

Currently various methods are being used to predict diabetes and diabetic inflicted diseases. In the proposed work, we have used the Machine Learning algorithms Support Vector Machine & Random Forest that would help to identify the potential chances of getting affected by Diabetes Related Diseases. After pre-processing the data, features which influences the prediction are selected by implementing step forward and backward feature selection. The Principal Component Analysis dimensionality reduction method is analysed after the selection of specific features and the accuracy of the prediction is 83% implementing Random Forest (RF) which is significant in comparison with Support Vector Machine (SVM) with accuracy of 81.4%.

2. LITERATURE SURVEY

□ Veena Vijayan V. And Anjali C has discussed, the diabetes disease produced by rise of sugar level in the plasma. Various computerized information systems were outlined utilizing classifiers for anticipating and diagnosing diabetes using decision tree, SVM, Naive Bayes and ANN algorithms.

□ P. Suresh Kumar and V. Umatejaswi has presented the algorithms like Decision Tree, SVM, Naive Bayes for identifying diabetes using data mining techniques.

□ Ridam Pal, Dr. Jayanta Poray and Mainak Sen has presented the Diabetic Retinopathy (DR) which is one of the leading cause of sight inefficiency for diabetic patients. In which they reviewed the performance of a set of machine learning algorithms and

verify their performance for a particular data set.

□ Dr. M. Renuka Devi and J. Maria Shyla has discussed about the analysis of various skills of mining to guess diabetes using Naive Bayes, Random Forest, Decision Tree and J48 algorithm.

□ Rahul Joshi and Minyechil Alehegn has discussed the ML techniques which are used to guess the datasets at an initial phase to save the life. Using KNN and Naive Bayes algorithm.

□ Zhilbert Tafa and Nerxhivane Pervetica has discussed the result of algorithms that are implemented in order to progress the diagnosis reliability.

□ Prof. Dhomse Kanchan B. and Mr. Mahale Kishor M. has discussed the study of Machine Learning Algorithms such as Support Vector Machine, Naïve Bayes, Decision Tree, PCA for Special Disease Prediction using Principal of Component Analysis.

3. PROBLEM STATEMENT

In Previous work Marius et al. have proposed this system that implements rather fast generating nearest neighbour and appropriate algorithm configuration. In this system, this system they have built up a framework that chooses a fitting algorithm in view of the data bolstered which rather creates the fastest nearest neighbour. This algorithm is selected based on dimension of the data. For some PC vision issues, the most tedious segment comprises of nearest neighbor coordinating in high dimensional spaces. There are no known correct algorithms for tackling these high-dimensional issues that are speedier than

straight pursuit. Rough algorithms are known to furnish expansive speedups with just minor misfortune in exactness, however numerous such algorithms have been distributed with just negligible direction on choosing an algorithm and its parameters for any given issue.

4. PROPOSED SYSTEM

The aim of our system is to work on a larger dataset to increase the efficiency of the overall system. The number of medical tests also affects the performance of the system; thus, our aim is to reduce the number of medical tests to increase the efficiency of the system. The admin of the system will then choose one of the two appropriate algorithms available. Thus, after using the system, the prediction will be done whether the patient is diagnosed with diabetes or not. If the patient is found diabetic expert recommendations would be provided to the patient so that he/she can recover from diabetes. Whole report will be provided to the patient in the printed form, as usually provided in the hospitals like a report. This system would be very much useful in the field of healthcare. The proposed system focuses using algorithms combinations shown above in the block diagram. The base classification algorithms are: Decision tree, Support Vector Machine, Naive Bayes and ANN for accuracy authentication.

5. DESIGN

System design is transition from a user-oriented document to programmers or data base Personnel. The design is a solution, how to approach to the creation of a new system. Designing goes through logical and

physical stages of development, logical design reviews the present physical system, prepare input and output specification, details of implementation plan and prepare a logical design walkthrough.

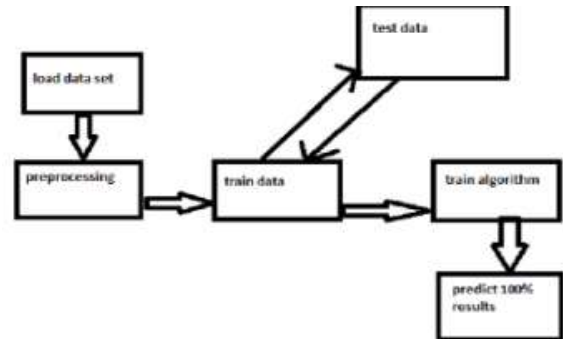
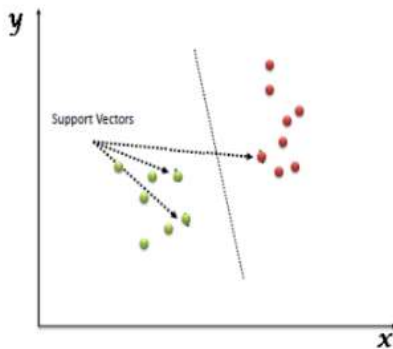


Fig 5.0.1 System Architecture

6. IMPLEMENTATION

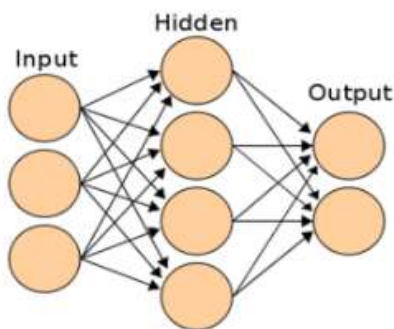
6.1 SVM Algorithm

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes.



6.2 ANN Algorithm

It is truly said that the working of ANN takes its roots from the neural network residing in human brain. ANN operates on something referred to as Hidden State. These hidden states are similar to neurons. Each of these hidden states is a transient form which has a probabilistic behavior. A grid of such hidden state act as a bridge between the input and the output.



6.3 Passive Aggressive Classifier Algorithm

The Passive Aggressive algorithms are a family of Machine learning algorithms that are not very well known by beginners and even intermediate Machine Learning enthusiasts. However, they can be very useful and efficient for certain applications. This is a high-level overview of the algorithm explaining how it works and when to use it. It does not go deep into the mathematics of how it works Passive

Aggressive algorithms are generally used for large-scale learning. It is one of the few “online-learning algorithms”. In online machine learning algorithms, the input data comes in sequential order and the machine learning model is updated step-by-step, as opposed to batch learning, where the entire training dataset is used at once. Passive: If the prediction is correct, keep the model and do not make any changes. i.e., the data in the example is not enough to cause any changes in the model.

Aggressive: If the prediction is incorrect, make changes to the model. i.e., some change to the model may correct it. If you want to work on big data, this is a very important classifier and I encourage you to go ahead and try to build a project using this classifier and use live data from a social media website like Twitter as input. There will be a huge amount of data coming in every second and this classifier will be able to handle data of this size.

7. INTERNAL MODULES

7.1 NumPy

NumPy is a Python package which stands for ‘Numerical Python’. It is the core library for scientific computing, which contains a powerful n-dimensional array object, provide tools for integrating C, C++ etc. It is also useful in linear algebra, random number capability etc. NumPy array can also be used as an efficient multi-dimensional container for generic data. Now, let me tell you what exactly is a python NumPy array. To install Python NumPy, go to your command prompt and type “pip install NumPy”. Once the installation is completed, go to your IDE

(For example: PyCharm) and simply import it by typing: “import NumPy as np”.

7.2 Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data. In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data. Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.

8. RESULTS



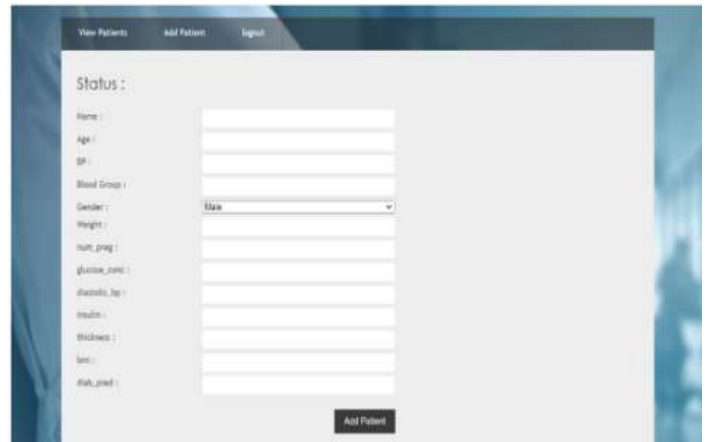
6.6.1 Home Page



6.6.2 Reception Login



6.6.3 Reception Page



6.6.4 Add Patient Page Details



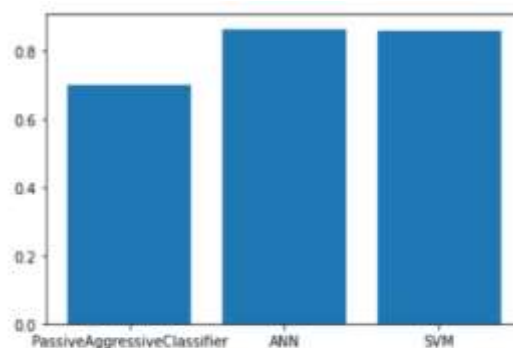
6.6.5 Admin Login



6.6.9 Doctor Page



6.6.6 Admin Page



6.6.10 Bar graph



6.6.7 Add Doctor



6.6.8 Doctor Login

9. CONCLUSION

Diabetes is a serious and chronic condition. Diabetes can be detected early enough which can result in more effective treatment. This study also compares various classification models based on machine learning algorithms for predicting a patient's diabetic condition at the earliest feasible stage. After dataset balancing, classifiers accuracy was compared. The professionals in the medical field have approved the features chosen for the early diagnosis of diabetes prediction. It is recommended that using state of art algorithms for the early prediction can help in decreasing the upward trend of diabetes three algorithms including SVM, PAC and ANN algorithms were utilized for this purpose amongst all we achieved ANN algorithm has the highest accuracy of 89%. To determine the best and



most accurate diabetes prediction algorithm, a variety of various algorithms and combinations of algorithms can be examined. In this application we will upload some test data and will be consider as passive aggressive classifier and SVM and ANN model on test data to predict patient condition and send resultant data to this application. As we don't have sensors to sense data so we consider uploaded test data as sense data.

10. FUTURE SCOPE

We can make this application as an android app so that it becomes easier for the users. We can add more algorithms to find outputs and algorithms can be compared to find efficient algorithm. We can add visitor query module. Where visitor can post queries to administrator and admin can send reply to those queries. We can add treatment modules.

11. REFERENCES

- 1) Veena Vijayan V. And Anjali C, Prediction and Diagnosis of Diabetes Mellitus, "A Machine Learning Approach", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10- 12 December 2015 | Trivandrum.
- 2) 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.
- 3) Ridam Pal ,Dr. Jayanta Poray, and Mainak Sen, , "Application of Machine Learning Algorithms on Diabetic Retinopathy", 2017 2nd IEEE International Conference On Recent Trends In Electronics

Information & Communication Technology, May 19-20, 2017, India.

- 4) Berina Alic, Lejla Gurbeta and Almir Badnjevic, "Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases", 2017 6th Mediterranean Conference On Embedded Computing (MECO), 11-15 JUNE 2017, BAR, MONTENEGRO.

- 5) Dr. M. Renuka Devi and J. Maria Shyla, "Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 1 (2016) pp 727-730 © Research India Publications. <http://www.ripublication.com>

- 6) Rahul Joshi and Minyechil Alehgn, "Analysis and prediction of diabetes diseases using machine learning algorithm": Ensemble approach, International Research Journal of Engineering and Technology Volume: 04 Issue: 10 | Oct -2017

- 7) Zhilbert Tafa and Nerxhivan Pervetica, "An Intelligent System for Diabetes Prediction", 4th Mediterranean Conference on Embedded Computing MECO – 2015 Budva, Montenegro.

- 8) Sumi Alice Saji and Balachandran K, "Performance Analysis of Training Algorithms in Diabetes Prediction", International Conference on Advances in Computer Engineering and Applications (ICACEA) IMS Engineering College, Ghaziabad, India 2015.

- 9) Aakashsha Rathore and Simran Chauhan, "Detecting and Predicting Diabetes Using Supervised Learning". International Journal



of Advanced Research in Computer Science,
Volume: 08, May/June 2017.

10) April Morton, Eman Marzban and Ayush Patel, “Comparison of Supervised Machine Learning Techniques for Predicting Short-Term In-Hospital Length of Stay Among Diabetic Patients, 13th International Conference on Machine Learning and Applications”, 2014.

11) Prof. Dhomse Kanchan B. and Mr. Mahale Kishor M. “Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis”. International Conference on Global Trends in Signal Processing, Information Computing and Communication 2016.

12) Deeraj Shetty, Kishor Rit, Sohail Shaikh and Nikita Patil “Diabetes Disease Prediction Using Data Mining”. International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) 2016.