PREDICTION OF DIABETIC INSULIN DOSAGE USING MACHINE LEARNING

1 Y.Ashok Assistant Professor, Department of CSE, Chalapathi Institute of Technology, Guntur.
2 Appala Maruthi Mohan, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
3 Ch.Likhitha Padmavathi,B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
4 Kosuru.Srinivasarao, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
5 Kari Alekhya,B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.

ABSTRACT: Diabetes Mellitus is a chronic metabolic disorder. Normally, with a proper adjusting of blood glucose levels (BGLs), diabetic patients could live a normal life without the risk of having serious complications that normally developed in the long run. However, blood glucose levels of most diabetic patients are not well controlled for many reasons. Although the traditional prevention techniques such as eating healthy food and conducting physical exercise are important for the diabetic patients to control their BGLs, however taking the proper amount of insulin dosage has the crucial rule in the treatment process. In this project we are using Gradient Boosting Classifier to predict diabetes and then using Logistic Regression algorithm to predict insulin dosage in diabetic detected patients. To implement this project we are using PIMA diabetes dataset and UCI insulin dosage dataset. We are training both algorithms with above mention dataset and once after training we will upload test dataset with no class label and then Gradient Boosting will predict presence of diabetes and Logistic Regression will predict insulin dosage if diabetes detected by Gradient Boosting.

1. INTRODUCTION

Diabetes Mellitus (DM) is a chronic disease caused by a disorder in the glucose metabolism leading to abnormal Blood Glucose Level (BGL) that can be higher (Hyperglycemia) or lower (Hypoglycemia) than normal range: 70 mg/dl to 140 mg/dl (3.9 mmol/L to 7.8 mmol/L) [1]. This disorder can be either due to the deficiency of the pancreas which does not produce enough insulin (Type 1 Diabetes Mellitus, T1DM), or to an inability of the insulin to be effectively used by the body cells (Type 2 Diabetes Mellitus, T2DM) or during pregnancy when placenta produces high levels of hormones impairing the action of the insulin (Gestational Diabetes Mellitus, GDM) [2]. According to the International Diabetes Federation (IDF Diabetes Atlas) [3], diabetes prevalence in 2019 has been estimated at 9.3% of the global adult population aged between 20 and 79 years old, which represents 463 million people. This number is expected to reach 578 million people by 2030 (10.2%) and 700 million people by 2045 (10.9%). Diabetes Mellitus therapy aims to maintain blood glucose levels within a normal range [4]. Otherwise, diabetic patients could face higher risks of complications including damage to blood vessels, cardiovascular diseases, blindness, kidney damage, coma, or even death [2]. Fortunately, complications are highly less common in patients who effectively manage their BGL [5]. To maintain a good glycaemia control, diabetic patients need to measure regularly their blood glucose level. Measurements could be performed using Continuous Glucose Monitoring (CGM) sensors or manually using sticks several times a day [6]. Furthermore, diabetic patients need to predict their blood glucose level in the future depending on the ingested meals, the injected insulin bolus, the performed physical activity, and other factors impacting the glucose dynamics. Predicting BGL helps patients not only regulate their glycaemia but also avoid hyperglycemic and hypoglycemic episodes [7]. Machine Learning (ML) techniques have become more popular among data mining methods [8]. They have been successfully used in many eHealth fields including cardiology [9], oncology [[10], [11], [12]] and endocrinology [[13], [14], [15]]. The availability and growing use by diabetic patients of sensors to capture their BGL, insulin bolus, energy expenditure and other physiological measurements contributed to the build-up of large historical data repositories [[16], [17], [18], [19]]. Hence, the availability of these data allows the use of ML techniques to build regression models for blood glucose level prediction. Consequently, several ML techniques have been developed and evaluated to predict diabetic patients' BGL including Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Genetic Programming (GP) [15]. Moreover, many reviews have been published to analyze and summarize the findings of primary studies dealing with the use of ML techniques in diabetes self-management [20] and BGL prediction in particular. Woldaregay et al. [15] conducted a literature review of blood glucose prediction using ML strategies in type 1 diabetes where 55 papers published between 2000 and February 2018 have been assessed and analyzed. El Idrissi et al. [20] performed a systematic mapping and review study on the use of predictive techniques in diabetes self-management where 38 papers published between 2000 and April 2017 have been reviewed. Kavakiotis et al. [21] conducted a systematic review of the application of ML and Data Mining Methods in diabetes research where an assessment has been performed on 103 articles published between July 2011 and July 2016. Oviedo et al. [22] presented a methodological review of models for predicting blood glucose by analyzing 140 papers published between 2010 and April 2016. Nonetheless, using a single ML technique to predict BGL values depends on the evaluation context and therefore might provide inaccurate values when applied to different patients and even to the same patient under different circumstances [15]. In contrast, ensemble methods are learning techniques aggregating multiple single ML models into one predictive model using specific combination schemes. This might allow to reach higher prediction accuracy and a better variance/bias trade-off by consolidating the advantages and mitigating the limitations of each single learner [23]. To the best of our knowledge, no systematic literature review (SLR) has yet been performed on ensemble techniques in BGL prediction. However, many reviews have been conducted on the use of ensembles in medical subfields such as oncology [12,24], cardiology [25], and neurology [26]. Moreover, Fernandes-Aleman et al. [14] conducted an SLR of the usage of classification ensemble techniques in diabetes; they found that the diagnosis is the most investigated clinical task and pointed out the need of carrying out more research on the other medical tasks including monitoring, management and treatment of diabetes. The aim of this paper is, therefore, to carry out an SLR to synthetize and analyze the state of the art of the current studies that explore ensemble techniques in BGL prediction.

2. LITERATURE SURVEY

1. Water demand forecasting using extreme learning machines

AUTHORS: Tiwari, Mukesh, Jan Adamowski, and Kazimierz Adamowski

ABSTRACT: The capacity of recently-developed extreme learning machine (ELM) modelling approaches in forecasting daily urban water demand from limited data, alone or in concert with wavelet analysis (W) or bootstrap (B) methods (i.e., ELM, ELMW, ELMB), was assessed, and compared to that of equivalent traditional artificial neural network-based models (i.e., ANN, ANNW, ANNB). The urban water demand forecasting models were developed using 3-year water demand and climate datasets for the city of Calgary, Alberta, Canada. While the hybrid ELMB and ANNB models provided satisfactory 1-day lead-time forecasts of similar accuracy, the. diabetes using data mining techniques. The dataset has taken 768 instances from PIMA IndianDiabetes Dataset to determine the accuracy of the data mining techniques in prediction. Then we developed five predictive models using 9 input variables and one output variable from the Dataset information; we evaluated the five models in terms of their accuracy, precision, sensitivity, specificity and F1 Score measures. The purpose of this study is to compare the performance analysis of Naïve Bayes, Logistic Regression; Artificial neural networks (ANNs), C5.0 Decision Tree and Support Vector Machine (SVM) models for predicting diabetes using common risk factors. The decision tree model (C5.0) had given the best classification accuracy, followed by the logisticregression model, Naïve Bayes, ANN and the SVM gave the lowest accuracy Index Terms— Data mining, Prediction, Naïve Bayes, Logistic Regression, C5.0 Decision Tree, Artificial Neural Networks (ANN) and Support Vector Machine (SVM).

2. Comparison Data Mining Techniques To Prediction Diabetes Mellitus AUTHORS: Aswan Supriyadi Sunge

ABSTRACT: Diabetes is one of the chronic diseases caused by excess sugar in the blood. Various methods of automated algorithms in various to anticipate and diagnose diabetes. One approach to data mining method can help diagnose the patient's disease. In the presence of predictions can save human life and begin prevention before the disease attacks the patient. Choosing a legitimate classification clearly expands the truth and accuracy of the system as levels continue to increase. Most diabetics know little about the risk factors they face before the diagnosis. This method uses developing five predictive models using 9 input variables and one output variable from the dataset information. The purpose of this study was to compare performance analysis of Naive Bayes, Decision Tree, SVM, K-NN and ANN models to predict diabetes mellitus. ANNW and ELMW models provided greater accuracy, with the ELMW model outperforming the ANNW model. Significant improvement in peak urban water demand prediction was only achieved with the ELMW model. The superiority of the ELMW model over both the ANNW or ANNB models demonstrated the significant role of wavelet transformation in improving the overall performance of the urban water demand model.

3. Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus. AUTHORS: Devi, M. Renuka, and J. Maria Shyla

ABSTRACT: Data mining approach helps to diagnose patient's diseases. Diabetes Mellitus is a chronic disease to affect various organs of the human body. Early prediction can save human life and can take control over the diseases. This paper explores the early prediction of diabetes using various data mining techniques. The dataset has taken 768 instances from PIMA Indian Dataset to determine the accuracy of the data mining techniques in prediction. The analysis proves that Modified J48 Classifier provide the highest accuracy than other techniques.

4.Predicting Diabetes Mellitus using Data Mining Techniques AUTHORS: J. Steffi, Dr. R. Balasubramanian,Mr. K. Aravind Kumar

ABSTRACT: Diabetes is a chronic disease caused due to the expanded level of sugar addiction in the blood. Various automated information systems were outlined utilizing various classifiers for anticipate and diagnose the diabetes. Data mining approach helps to diagnose patient's diseases. Diabetes Mellitus is a chronic disease to affect various organs of the human body. Early prediction can save human life and can take control over the diseases. Selecting legitimate classifiers clearly expands the correctness and adeptness of the system. Due to its continuously increasing rate, more and more families are unfair by diabetes mellitus. Most diabetics know little about their risk factor they face prior to diagnosis.This paper explores the early prediction of

3. EXISTING SYSTEM

Insulin enables blood glucose to get into the cells and this glucose is used for energy. So blood glucose is kept in a narrow range. Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. However, early prediction of diabetes is quite challenging task for medical practitioners due to complex interdependence on various factors. Diabetes affects human organs such as kidney, eye, heart, nerves, foot etc.

DISADVANTAGES

- Low Accuracy to predict diabetis

4. PROPOSED SYSTEM

In this project we are using Gradient Boosting Classifier to predict diabetes and then using Logistic Regression algorithm to predict insulin dosage in diabetic detected patients. To implement this project we are using PIMA diabetes dataset and UCI insulin dosage dataset. We are training both algorithms with above mention dataset and once after training we will upload test dataset with no class label and then Gradient Boosting will predict presence of diabetes and Logistic Regression will predict insulin dosage if diabetes detected by Gradient Boosting.

ADVANTAGES

Individual blood glucose level and insulin dosing are highly erratic and precise prediction of them is likely impractical. Average blood glucose level over 24 hours can be more reliably predicted and determining whether the patient's glucose level is going to be high is a more feasible task. And by using these algorithms we can get accurate results and fast execution.

5. UML DIAGRAMS

1. CLASS DIAGRAM

The cornerstone of event-driven data exploration is the class outline. Both broad practical verification of the application's precision and fine-grained demonstration of the model translation into software code rely on its availability. Class graphs are another data visualisation option.

The core components, application involvement, and class changes are all represented by comparable classes in the class diagram. Classes with three-participant boxes are referred to be "incorporated into the framework," and each class has three different locations:

• The techniques or actions that the class may use or reject are depicted at the bottom.

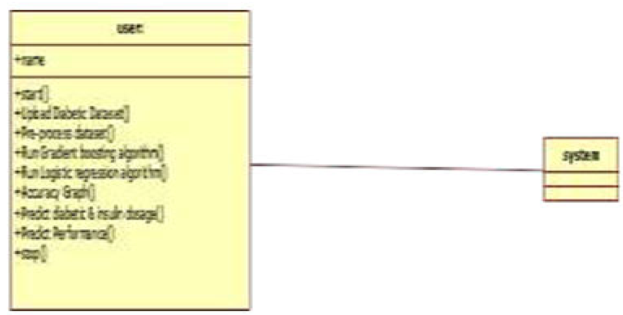

Fig 5.1 shows the class diagram of the project

1. USECASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

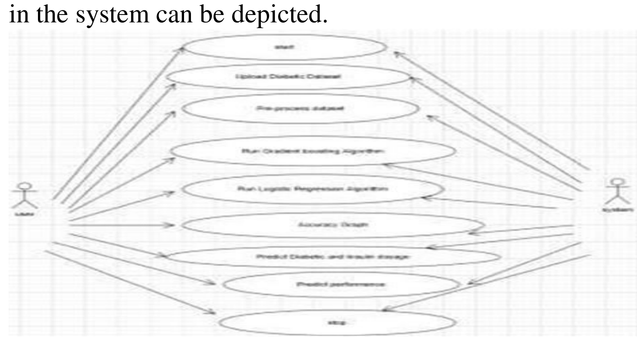

Fig 5.2 Shows the Use case Diagram

3. SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

Fig 5.3 Shows the Sequence Diagram

6. RESULTS

Both dataset available inside Dataset folder andbelow screen is showing dataset details

6.1 Output Screens

Fig 6.1 Diabetes dataset

To run project double click on 'run.bat' file to getbelow screen

Fig 6.2 Upload the dataset

In above screen click on 'Upload Diabetic Dataset' button to upload dataset

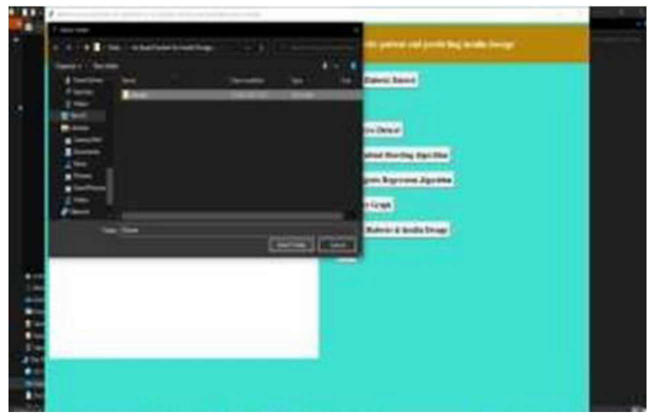

Fig 6.3 Upload the dataset

In In above screen selecting and uploading entire 'Dataset' folder to load both diabetes and insulin dataset and then click on 'Select Folder' button to get below screen

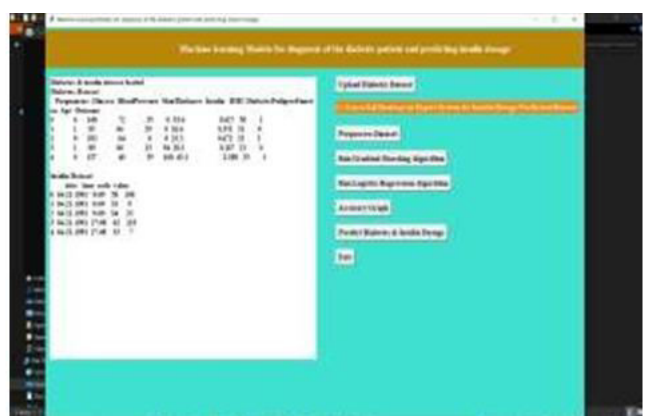

Fig 6.4 Uploaded Data

In above screen we can see both datasets loaded and we can see some records from each dataset and will getbelow graph also

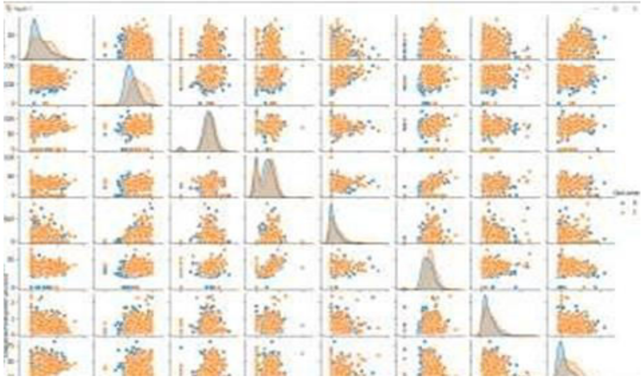

Fig 6.5 Diabetic Data Graph

In above graph we can see diabetes for each column where red colour dots indicate presence of diabetes and blue represents no diabetes detected. We are plotting graph for each column value to show with which value diabetes is present and with which value diabetes is not present for example in above graph in first column we are plotting graph for 'number of Pregnancies' with 'presence or no presence of diabetes' and now close above graph and then click on 'Preprocess Dataset' button to remove missing values and to split dataset into train and test

Fig 6.6 Run Gradient Boosting Algorithm

In above screen we can see dataset contains total 768 records and application using 80% records for training and 20% records to test ML accuracy and now dataset is ready and now click on 'Run Gradient Boosting Algorithm' button to train gradient boosting with above dataset

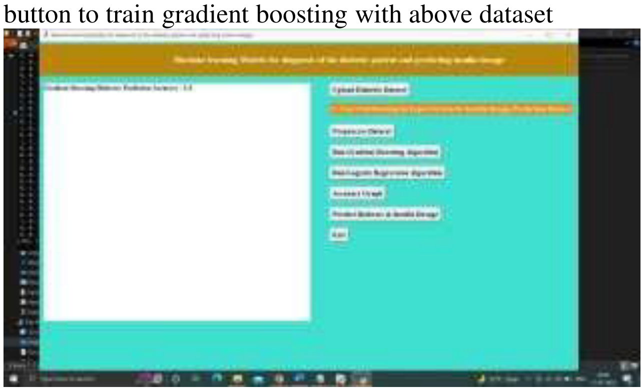

Fig 6.7 Run Gradient Boosting Algorithm Accuracy

In above screen we got gradient boosting accuracyas 100% and now click on 'Run Logistic Regression Algorithm' button to build logistic regression algorithm

Fig 6.8 Run Logistic Regression Algorithm

In above screen with logistic regression we got 78.7% accuracy and now click on 'Accuracy Graph' button to get below graph

Fig 6.8 Accuracy Comparison Graph

In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and now close above graph and then click on 'Predict Diabetes & Insulin Dosage' button to upload test values and then will get predictionresult

7. CONCLUSION

This paper was aimed at modeling for the prediction of amount of insulin dosage suitable for diabetic patients. A model based on gradient boosting trained with BP was used. The model uses eight input information each patient. Pregnancies, glucose, blood pressure, skin Thickness, Insulin, BMI, diabetes pedigree Function, age. We are using Gradient boosting to predict diabetes and then applying Logistic Regression Algorithm to predict insulin dosage if diabetes detected by Gradient boosting algorithm. The Gradient boosting model converged fast and gave results with high performance.

Future Enhancement

The results of the research conducted in this thesis are promising for a wide range of applications in T1D therapy. Nevertheless, the performance and safety of the predictions can be improved further by generating a set of interchangeable models that predict useful Blood Glucose values for control and therapy purposes based on the determination of individual specific dynamics, lifestyle, and other factors. An extension of this work will include testing personalized Blood Glucose prediction models in a more challenging situation involving real subjects.

8. REFERENCES

[1] American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care, Vol. 31, No. 1, 2008, 55-60, 1935- 5548.

[2] I. Eleni Georga, C. Vasilios Protopappas and I. Dimitrios Fotiadi,. "Glucose Prediction in Type 1 and Type 2 Diabetic Patients Using Data Driven Techniques," Knowledge-Oriented Applications in Data Mining, InTech pp 277-296, 2011.

[3] V. Tresp,, T. Briegel, and J. Moody, "Neural- network models for the blood glucose metabolism of a diabetic," IEEE Transactions on Neural Networks, Vol. 10, No. 5, 1999, 1204-1213, 1045-9227.

[4] C. M. Bishop, . Pattern Recognition and Machine Learning, Springer, 2006, New York.

[5] S. Haykin, Neural networks and learning machines. Pearson 2008.

[6] W.D. Patterson, Artificial Neural Networks- Theory and Applications, Prentice Hall , Singapore. 1996.

[7] M. Pradhan and R. Sahu, "Predict the onset of diabetes disease using Artificial Neural Network (ANN)," International Journal of Computer Science & Emerging Technologies, 303 Volume 2, Issue 2, April 2011.

[8] W, Sandham, D,Nikoletou, D.Hamilton, K, Paterson, A. Japp and C. MacGregor, "BLOOD Glucose Prediction for Diabetes THERAPY USING A RECURRENT Artificial Neural Networks ", EUSIPCO, Rhodes, 1998, PP. 673-676

[9] M, Divya, R. Chhabra, S. Kaur and S. , "Diabetes Detection Using Artificial Neural Networks & Back-Propagation Algorithm", International Journal of Scientific & Technology Research". V 2, ISSUE 1, JANUARY 2013.

[10] G, Robertson, E. Lehmann, W. Sandham, and D. Hamilton Blood, "Glucose Prediction Using Artificial Neural Networks Trained with the AIDA Diabetes Simulator: A Proof-of-Concept Pilot Study". Journal of Electrical and Computer Engineering , V 2011 (2011), Article ID 681786, 11 pages.