Journal of Cardiovascular Disease Research

ISSN: 0975-3583,0976-2833 VOL12, ISSUE7, 2021

DIABETES AND INSULIN DOSAGE PREDICTION

[1]**Dr B Ravi Prasad,** Professor, CSE Department, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, rprasad.boddu@gmail.com

[2]**Ch.V.Krishna Mohan** Associate Professor, CSE DEPARTMENT, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, kmphd2014@gmail.com

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 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 paper we have proposed Gradient boosting, decision tree and logistic regression models to predict the proper amount of insulin needed for the diabetic patient. The proposed models was trained and tested using several patients data containing many factors such as weight, fast blood sugar and gender. The proposed models showed good results in predicting the appropriate amount of insulin dosage.

1. Introduction

Diabetes mellitus, commonly referred to as diabetes, is a group of metabolic diseases characterized by high blood glucose concentrations resulting from defects in insulin secretion, insulin action or both. Diabetes has been classified into two major categories, namely, type 1 and type 2. Type 1 diabetes, which accounts for only 5-10% of those with diabetes, is caused by the cell-mediated autoimmune destruction of the insulin producing βcells in the pancreas leading to absolute insulin deficiency. On the other hand, type 2 diabetes is a more prevalent category (i.e. accounts for ~90-95% of those with diabetes) and is a combination of resistance to insulin action and an inadequate compensatory insulin secretion. In addition to the general guidelines that the patient follows during his daily life, several diabetes management systems have been proposed to further assist the patient in the self-management of the disease. One of the essential components of a diabetes management system concerns the predictive modeling of the glucose metabolism. It is evident that the prediction of glucose concentrations could facilitate the appropriate patient reaction in crucial situations such as hypoglycemia. Thus, several recent studies have considered advanced datadriven techniques for developing accurate predictive models of glucose metabolism. The fact that the relationship between input variables (i.e. medication, diet, physical activity, stress etc.) and glucose levels is nonlinear, dynamic, interactive and patient-specific, necessitates the application of non-linear regression models such as artificial neural networks, support vector regression and Gaussian processes.

2. literature survey

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

ISSN: 0975-3583,0976-2833 VOL12, ISSUE7, 2021

ELMB and ANNB models provided satisfactory 1-day lead-time forecasts of similar accuracy, the 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. 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.

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 diabetes using data mining techniques. The dataset has taken 768 instances from PIMA Indian Diabetes 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 logistic regression model, Naïve Bayes, ANN and the SVM gave the lowest accuracy IndexTerms—Data mining, Prediction, Naïve Bayes, Logistic Regression, C5.0 Decision Tree, Artificial Neural Networks (ANN) and Support Vector Machine (SVM).

3.Existed models

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. 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.

ISSN: 0975-3583,0976-2833 VOL12, ISSUE7, 2021

4. Proposed model

In this study, we work with de-individualized prediction models that do not require knowledge of physiological parameters or any manual work other than the collection of EHR data, thus they are more generalizable and feasible in practice.

we are using Gradient boosting, decision tree and logistic regression algorithms to predict diabetes and then applying Random Forest Regression Algorithm to predict insulin dosage. To train both algorithms we have PIMA and UCI dataset. 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.

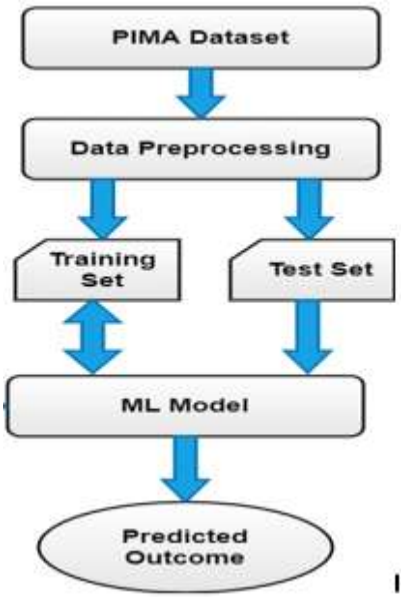

Figure :1 proposed model diagram

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

Modules:

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing
- Splitting data into train & test: using this module data will be divided into train & test

ISSN: 0975-3583,0976-2833 VOL12, ISSUE7, 2021

- Model generation: Build Gradient boosting, Decision tree and logistic regression Classifiers. Algorithms accuracy calculated.
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

Algorithms

Gradient boosting: Gradient boosting is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error.

Decision tree: A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes. Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves.

logistic regression: Logistic regression is commonly used for prediction and classification problems. Some of these use cases include: Fraud detection: Logistic regression models can help teams identify data anomalies, which are predictive of fraud.

Python is currently the most widely used multi-purpose, high-level programming language.

Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java.

Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.

Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber… etc.

The biggest strength of Python is huge collection of standard library which can be used for the following –

- Machine Learning
- GUI Applications (like Kivy, Tkinter, PyQt etc.)
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like Opencv, Pillow)
- Web scraping (like Scrapy, BeautifulSoup, Selenium)
- Test frameworks
- Multimedia

ISSN: 0975-3583,0976-2833 VOL12, ISSUE7, 2021

Python downloads with an extensive library and it *contain code for various purposes like regular expressions, documentation-generation, unit-testing, web browsers, threading, databases, CGI, email, image manipulation, and more.* So, we don't have to write the complete code for that manually.

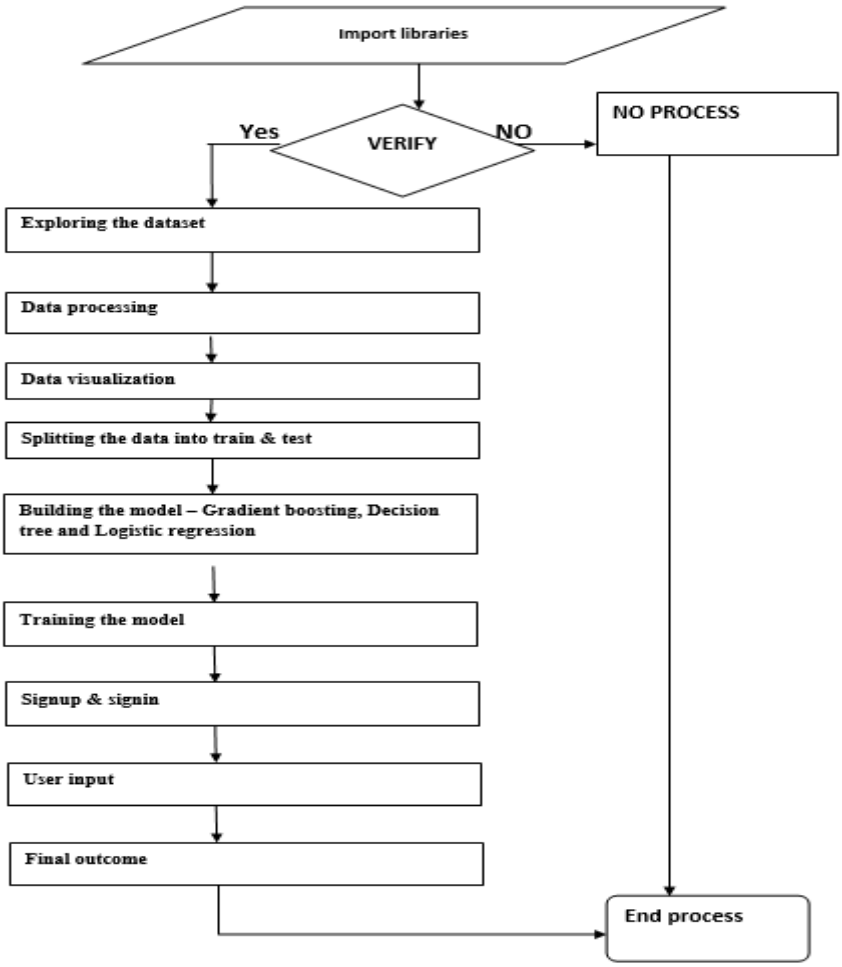

Figure :2 Flow of work

Machine Learning is the most rapidly growing technology and according to researchers we are in the golden year of AI and ML. It is used to solve many real-world complex problems which cannot be solved with traditional approach.

5. Results and discussion

Figure :3 screen

ISSN: 0975-3583,0976-2833 VOL12, ISSUE7, 2021

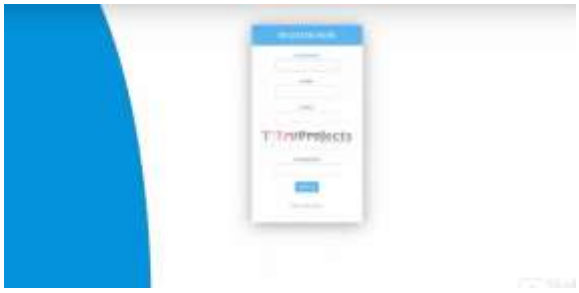

Figure :4 registration

Figure :5 Diabetes prediction

CONCLUSION

This paper was aimed at modeling neural network for the prediction of amount of insulin dosage suitable for diabetic patients. A model based on ANN trained with BP was used. The model uses four input information about each patient its length, weight blood sugar, and gender. Many experiments were conducted on 180 patient's data. we are using gradient boosting, random forest and logistic regression models to predict diabetes and then predict insulin dosage if diabetes detected. The models converged fast and gave results with high performance.

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.

ISSN: 0975-3583,0976-2833 VOL12, ISSUE7, 2021

[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.