Diabetes Prediction and Insulin Dosage Prediction Using Machine Learning

Saketh Madabushi, Madhuri Kesari, Shiva Teja Dusa,
Arun Kumar Gongati and Venkata Rao Yanamadni

July 20, 2022

DIABETES PREDICTION AND INSULIN DOSAGE PREDICTION USING MACHINE LEARNING

Madabushi Saketh[1], Kesari Madhuri[1], Dusa ShivaTeja[1], Gongati Arun Kumar Reddy[1] and Mr. Venkata Rao Yanamadni[1*], *Dept. Of CSE,
KG Reddy college of engineering and technology
Corresponding Author: venkataraoyanamadni@gmail.com

Abstract—Chronic metabolic problem caused by the disease known as diabetes mellitus. Blood glucose levels (BGLs) need to be adjusted so that diabetes individuals may continue to lead normal lives without developing the major problems that have historically plagued those with the disease. However, for various reasons, most diabetes individuals do not have their blood glucose levels under good control. Traditional methods of prevention, such as a good diet and regular exercise, are essential for diabetes patients, but the correct insulin dose is the most critical factor in managing their blood glucose levels. In this study, we offer a model for determining the optimal insulin dosage for a diabetic patient that is based on gradient boosting. Patients' weight, fasting glucose, and gender were among the numerous variables used to train and evaluate the suggested model. Good results were obtained when using the suggested approach to estimate the optimal insulin dose.

Index Terms— Diabetes, Prediction, Machine Learning, Data Pre-processing, Regression, Implementation, Visualization, Evaluation, Gradient boosting, ANN.

1. INTRODUCTION

Predicting glucose levels may help patients respond appropriately to life-threatening conditions like hypoglycemia. To this end, various ongoing examinations have analyzed the utilization of state-of-the-art information-driven techniques to make strong expectation models of glucose digestion. The connection between input factors (like drug, diet, active work, stress, and so on) and glucose levels is non-straight, dynamic, intelligent, and patient-explicit, requiring the utilization of non-direct relapse models like counterfeit brain organizations, support vector relapse, and Gaussian cycles.

Predicting diabetes is the primary focus of this research study. Besides the principal rules the patient continues in his routine, numerous diabetes executives' frameworks have been proposed to additional help the patient in self-administration of the sickness. As part of a comprehensive diabetes management strategy, predictive modeling of glucose metabolism should be included.

2. LITERATURE SURVEY

2.1 Water demand forecasting using extreme learning machines

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

ABSTRACT: In this study, we evaluated the performance of newly created extreme learning machine (ELM) modeling techniques to those of similar standard artificial neural network-based models for predicting daily urban water demand from limited data. These include ELM, ELMW, and ELMB (i.e., ANN, ANNW, ANNB). The climate and water demand data for Calgary, Alberta, Canada, were collected over three years to create models for predicting future water use in the city. The hybrid ELMB/ANNB model did well in providing 1-day lead-time predictions, while the ANNW/ELMW model did better. The ELMW model is the only one to significantly increase accuracy in predicting urban water demand peaks. The ELMW model outperformed the ANNW and ANNB models, demonstrating the importance of wavelet transformation to the success of the urban water demand model.

2.2 Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus.

AUTHORS: Devi, M. Renuka, and J. Maria Shyla

ABSTRACT: The use of data mining techniques improves the accuracy of medical diagnosis. As a long-term condition, diabetes mellitus may have far-reaching effects on the human body. Human lives can be saved, and disease outbreaks may be contained with early prediction. Predicting diabetes at an early stage using data mining approaches is investigated in this research. To test the efficacy of data mining approaches for prediction, 768 cases were drawn from the PIMA Indian Dataset. From what we can see, Modified J48 Classifier outperforms all other methods in terms of accuracy.

2.3 Predicting Diabetes Mellitus using Data Mining Techniques

AUTHORS: J. Steffi, Dr. R. Balasubramanian, Mr. K. Aravind Kumar

ABSTRACT: A high blood sugar level is the root cause of diabetes, a chronic condition. Several classifier-based automated information systems for diabetes prediction and diagnosis were presented. The use of data mining techniques improves the accuracy of medical diagnosis. As a long-term condition, diabetes mellitus may have far-reaching effects on the human body. Human lives can be saved and disease outbreaks may be contained with early prediction. Choosing reliable classifiers is a certain way to increase the system's precision and performance. As the prevalence of diabetes mellitus rises, more and more families are impacted negatively. Before diagnosis, most people with diabetes have a limited understanding of the risks they face. In this study, we investigate whether or not data mining methods may be used to detect diabetes in its earliest stages. To measure the accuracy of data mining predictions, 768 cases from the PIMA Indian Diabetes Dataset are used in this study. From that point onward, we used the dataset to fabricate five unmistakable forecast models, each with 9 information factors and 1 result variable, and we thought about them utilizing the measurements of exactness, accuracy, responsiveness, particularity, and F1 Score. The reason for this study was to think about the precision of diabetes risk expectation utilizing the Naive Bayes, Logistic Regression, Artificial Neural Network, C5.0 Decision Tree, and Support Vector Machine (SVM) models. The choice tree model (C5.0) had the most noteworthy precision for characterizing information, trailed by calculated relapse, Naive Bayes, counterfeit brain organization, and backing vector machine. File Some of the expressions utilized here incorporate "information mining," "expectation," "Innocent Bayes," "strategic relapse," "C5.0 choice tree," "brain organizations," and " Support Vector Machine " (SVM).

2.4 Comparison of Data Mining Techniques to Prediction Diabetes Mellitus

AUTHORS: Aswan Supriyadi Sunge

ABSTRACT: Excess glucose in the bloodstream may lead to a variety of chronic disorders, one of which is diabetes. Automated algorithms in several forms have been developed for use in diabetes prediction and diagnosis. The patient's illness may be diagnosed using one data mining strategy. Predictions may save lives and allow for prevention to begin before a disease strikes. Selecting a valid categorization naturally increases the

system's veracity and precision as the tiers progress. Unfortunately, most people with diabetes are uninformed of the dangers they face before they are diagnosed. This strategy involves building five prediction models from the data in the dataset, taking 9 inputs and producing 1 output. The reason for this study was to analyze the precision of expectations made by the Naive Bayes, Decision Tree, SVM, K-NN, and ANN models for diabetes mellitus.

3. Existing System

Glucose from the blood enters cells with the help of insulin and is utilized for cellular metabolism and energy production. A consequence of this is tight control over blood sugar levels. The rising prevalence of type 2 diabetes poses a danger to healthcare systems across the world. However, owing to the complicated interaction of numerous variables, early prediction of diabetes is a tough challenge for medical professionals. The human organs that suffer damage from diabetes include the kidney, eye, heart, nerves, foot, etc.

3.1 Disadvantages:

However, owing to the complicated interaction of numerous variables, early prediction of diabetes is a tough challenge for medical professionals. The human organs that suffer damage from diabetes include the kidney, eye, heart, nerves, foot, etc.

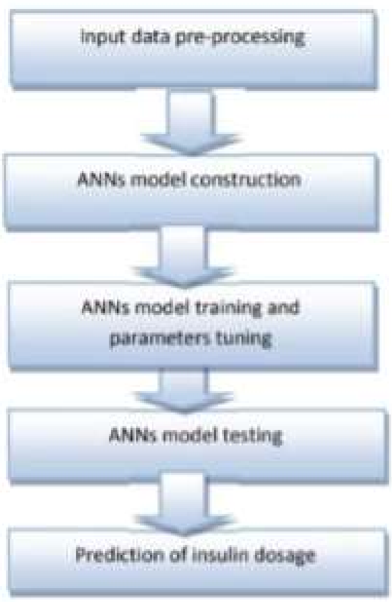
4. Proposed System

In this study, we employ de-individualized prediction models because they are more generalizable and applicable in practice than traditional models, which need knowledge of physiological factors and significant human work in addition to collecting EHR data. we are using Artificial Neural Networks to predict diabetes and then applying Gradient Boosting and Logistics Regression Algorithm to predict insulin dosage if diabetes is detected by the ANN algorithm. To train both algorithms we have PIMA and UCI datasets.

4.1 Advantages:

Due to the inherent unpredictability of both blood glucose and insulin requirements, accurate prediction of either is probably not feasible. Predicting the patient's average blood glucose level over a day and whether the patient's glucose level would be high is now a more manageable challenge.

5. System architecture



6. Implementation

Initially, we have PIMA and UCI datasets.

To test the efficacy of data mining approaches for prediction, 768 cases were drawn from the PIMA Indian Dataset. then data is preprocessed using data mining techniques. After the data preprocessing we use artificial neural networks to predict diabetes. Here We have training data i.e., PIMA and UCI datasets. These data sets are trained using the ANN algorithm.

ANN is a multi-layer algorithm with an input layer, a hidden layer, and an output layer.

The input signal is received by the Artificial Neural Network as a vector representation of a pattern or picture from an external source. For each set of n inputs, a corresponding mathematical notation x(n) is assigned. Each observation's explanatory attribute values will be sent into the system through the input layer.

The number of explanatory variables is often represented by the same number of input nodes in an input layer.

When it comes to the data, the nodes in the input layer are completely mute. One input value is replicated across several outputs. Duplicated values are generated at the input layer and distributed to all the hidden nodes. The input values are transformed internally by the network at the Hidden layers. In this case, the incoming arcs originate from other hidden nodes or input nodes linked to each node. It establishes links with arcs pointing in the direction of output nodes and maybe additional hidden nodes as well. The "connections" in the hidden layer are weighted to do the actual processing. Multiplying the data being fed into a hidden node by weights, which are a collection of integers recorded in the program, yields the output. These weights are often a stand-in for the sturdiness of the connections between individual neurons inside the ANN. The inputs are weighted, and their sum yields a single value. To increase the output to a value greater than zero and so increase the system's reaction, bias is introduced if the weighted sum is zero. The input for bias and the value of weight are both 1. In this case, the sum of the weighted inputs might be any positive number. Here, we utilize the greatest worth as a benchmark, then run the amount of the weighted contributions through the initiation capability to oblige the result to OK ranges. The initiation capability is the assortment of move works that give the ideal outcome. Interfacing the secret layers to a "yield layer" is the following stage. Associations are shipped off the result layer from the information layer and the secret layers. The predicted value of the response variable is provided as the output. Later the diabetes prediction insulin dosage prediction will be carried out. the insulin dosage prediction will be done using Gradient boosting and logistic regression algorithms. We use a gradient boosting approach to train the dataset. For both continuous and categorical targets, to make predictions, the gradient boosting method may be used as a Regressor (as a Classifier). When a regressor is being utilized, the expense capability is the Mean Squared Error (MSE), yet when a classifier is being utilized, the expense capability is the Log misfortune. Further developed preparing proficiency and diminished time spent in preparing: To speed up the training process, Light GBM employs a histogram-based technique, which divides up previously uncategorized continuous feature values into discrete bins.

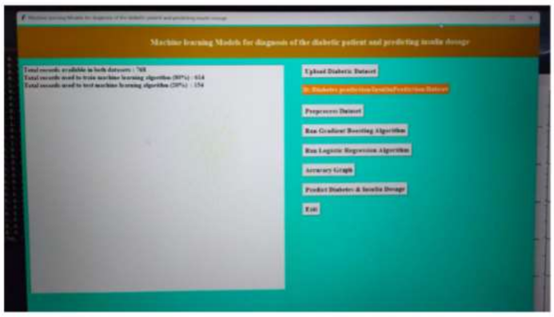
Replacement of continuous values with discrete bins reduces the amount of storage space required for data.

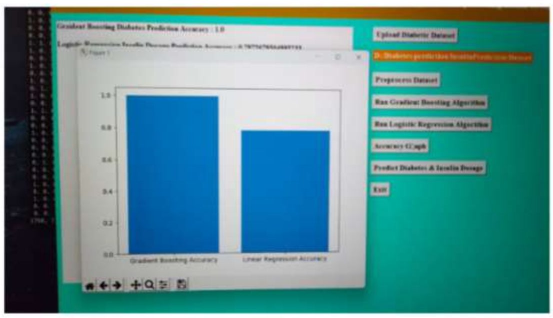
We execute a logistic regression algorithm to train data sets. One of the most used Machine Learning algorithms, logistic regression belongs to the category of Supervised Learning. A categorical dependent variable may be predicted from a collection of independent factors using this method. The goal of logistic regression is to foretell the values of a categorical dependent variable. In politics, for instance, logistic regression may be used to forecast an electoral victory or defeat, while in higher education, it could be used to determine whether a high school senior would be accepted to a certain university.

These simple yes/no choices are made possible by the binary nature of the results.

Later, we use test data to predict diabetes. if a patient has diabetes, it gives results as detected or else not detected. if the result is detected i.e., if the patient has diabetes, it further proceeds to insulin dosage using data that is available and algorithms like gradient boosting and logistic regression. The dosage will be predicted accuracy of gradient boosting is 100% and logistic regression is 78%.

7. output results





8. CONCLUSION

The purpose of this article was to develop a neural network for estimating insulin requirements in diabetic patients. A BP-trained model was employed as the basis for the model. Patient height, weight, blood sugar, and sex are the four data points used to develop the model. Data from 180 patients were used in a variety of investigations. If diabetes is predicted using the gradient boosting technique, the next step is to use the logistic regression approach to determine an appropriate insulin dose. Improvements to Come The research presented in this thesis shows promise for a variety of uses in the treatment of type 1 diabetes. However, by creating a set of interchangeable models that forecast desirable Blood Glucose readings for control and therapeutic purposes based on the assessment of individual dynamics, lifestyle, and other characteristics, the performance, and safety of the predictions may be further enhanced. To take this research further, we want to test our customized Blood Glucose prediction algorithms in a more realistic setting with human participants.

9. REFERENCES

Journals:

[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