Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed) Engineering and Science Journal Impact Factor:7.984(SJIF) Volume-4, Special Issue-3; ISSN: :2582-5887

An Expert System for Insulin DosagePrediction

Bolem Hamitha, Reddi Jhansi Rani,Mandru.Amulya,Samunuru.Hruthika.
Computer Science And Engineering
RK College of Engineering
Vijayawada, India.
hamithaboleam@gmail.com
DOI:10.53414/UIJES:2024.43.507

Abstract – **Death rates are increasing tremendously due to diabetes, which could be predicted and controlled in the early stage by prescribing insulin. Diabetes is a chronic disorder where the pancreas does not produce insulin or does not use it efficiently. Which can lead to life risk complications like heart stroke, eye damage, kidney failure Etc. Diabetes is manageable by giving a proper dosage of insulin to a patient.**

The project automates the diagnosis of diabetes for the patient. Based on PIMA Indian Diabetes dataset used to predict diabetes by using Extreme Gradient Boosting model. PIMA dataset contains nine attributes in total such as insulin, age, BMI, pregnancies, Etc. After the diagnosis of diabetes, a proper amount of insulin dosage is given to the diabetes patient.The dosage is given to patients based on the following factors as diet, level of physical activity, and severity of diabetes. This model automatically Predict Insulin dosage for the diabetes patient based on the linear regression (LR) algorithm. For the experimental analysis, the UCI Diabetes Dataset uses for dosage prediction. The diabetes dataset contains 20 features such as Regular insulin dose, NPH insulin dose, Ultra Lente insulin dose, Hypoglycaemic symptoms, Typical mean ingestion, Etc. The diabetic dataset is used to predict the insulin dosage for the patient according, to their symptoms and lifestyle characteristics as well as automating the entire process to reduce the burden on the healthcare profession andat the same time improving the health of the patient in an effective manner.

Keywords – **Diabetis Mellitus, Blood Glucose Levels (BGLs), Insulin Dosage, Gradient Boosting Classifier, Linear Regression.**

I. INTRODUCTION

Type 1 diabetes is a common chronic disease that has no treatment yet [1]. This disease is well known in children and is known as completely insulin dependent diabetes where the body fail to make any insulin [2]. This type of diabetes is one of the diseases that accompanies the patient and increases risk with time, and it is unable to be solved until today [3]. This is what drives patients to follow diets, continuously monitor the glucose level and take the appropriate doses of insulin [2].

The fumigation process in the diabetics treatment is mainly referred to the insulin injection that is necessary to manage the blood glucose level. People with type 1 diabetes must inject insulin up to four or five times per day to balance the blood glucose level. The insulin is delivered using a pump and the patient inserts a new cannula under the skin every two to three days [1-3]. Unfortunately, many patients fail to calculate the appropriate insulin doses [4]. The errors in dealing with insulin/ glucose levels lead to serious complications that may lead to the emergence of disorders caused by advanced diabetes such as loss of feeling in the limbs, loss of vision, loss of ability to heal wounds, and may cause an amputation of some extremities [5-6].

The methods used to monitor and maintain the blood glucose levels are still suffering from complications and have some effects on the patient's health [3]. A diabetic patient has difficulty in stopping bleeding on any wound [3-4]. Also, children have difficulty in tolerating and using acupuncture [3, 5]. Unstable blood glucose level is a main concern for many diabetes mellitus (DM) patients and the insulin therapy with self-monitored blood glucose (SMBG) can cause unstable blood glucose level.

The sliding scale method can be used to determine the insulin requirement in proportion to fasting blood glucose (FBG) [7-9]. The main challenge that we still face is due to the difficulty in determining the appropriate amount of insulin dose in a smart and faster way [8-10]. A smart system is necessary to measure the cumulative diabetes of the patient and reduce the negative consequences of the wrong prediction of the insulin dose.

Some recent research confirms the real-time prediction of insulin [6,11], but it still depends on an immediate attachment to the patient and face many challenges with negative complications [12-14]. Research work is working to develop the best ways to predict the amount of insulin necessary for maintaining the blood glucose level [15-19].

Moreover, medical and technical attempts focus on limiting negative complications of the disease. Several systems and methods used to deal with diabetes and regulate insulin. Iokibe et al. [9] employed a local fuzzy reconstruction method based on chaos theory for predicting fasting blood glucose (FBG) at peak time, and the amount of insulin was adjusted based on the predicted FBG level. The success rate of the prediction of the FBG at peak time was 70-90%. Otoom et al. [15] reported a real-time insulin injection system based on continuous updating of glucose level and continuous pumping of the required insulin to reach the desired glucose level. The same system was used with gestational diabetes [16]. This system suffers from several challenges due to the pain caused by the need of acupuncture several times a day [9, 13-15, 20]. However, some statistics indicate that the quality of life for the patients using this system has been improved [14,16]

Previous work was done to use an insulin pump to build a system for giving the patient the required insulin through inhalation as a more comfortable method than insulin injection [21-27]. Inhaled insulin is an insulin powder that is delivered to the body through the lungs [21]. Thus, this system can reduce the negative effects caused using insulin pump and might improve the efficiency of the treatment. Inhaled mono-inhaled insulin was developed by Mankind approved by the FDA in 2014 [26]. However, chronic lung disease cases caused the European Medicines Agency to withdraw the drug in 2018. This method might lead to a decreased efficiency compared to insulin injection under the skin, and some effects on the respiratory system, such as coughing and danger to asthma patients [27-28].

II. LITERATURE SURVEY

Diabetes in developing countries:

There has been a rapid escalation of type 2 diabetes (T2D) in developing countries, with varied prevalence according to rural vs urban habitat and degree of urbanization. Some ethnic groups (eg, South Asians, other Asians, and Africans), develop diabetes a decade earlier and at a lower body mass index than Whites, have prominent abdominal obesity, and accelerated the conversion from prediabetes to diabetes. The burden of complications, both macro- and microvascular, is substantial, but also varies according to populations. The syndemics of diabetes with HIV or tuberculosis are prevalent in many developing countries and predispose to each other. Screening for diabetes in large populations living in diverse habitats may not be cost-effective, but targeted high-risk screening may have a place. The cost of diagnostic tests and scarcity of health manpower pose substantial hurdles in the diagnosis and monitoring of patients. Efforts for prevention remain rudimentary in most developing countries. The quality of care is largely poor; hence, a substantial number of patients do not achieve treatment goals. This is further amplified by a delay in seeking treatment, "fatalistic attitudes", high cost and non-availability of drugs and insulins. To counter these numerous challenges, a renewed political commitment and mandate for health promotion and disease prevention are urgently needed. Several low-cost innovative approaches have been trialed with encouraging outcomes, including training and deployment of non-medical allied health professionals and the use of mobile phones and telemedicine to deliver simple health messages for the prevention and management of T2D.

Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset:

Diabetes Mellitus is a dreadful disease characterized by increased levels of glucose in the blood, termed as the condition of hyperglycemia. As this disease is prominent among the tropical countries like India, an intense research is being carried out to deliver a machine learning model that could learn from previous patient records in order to deliver smart diagnosis. This research work aims to improve the accuracy of existing diagnostic methods for the prediction of Type 2 Diabetes with machine learning algorithms. The proposed algorithm selects the essential features from the Pima Indians Diabetes Dataset with Goldberg's Genetic algorithm in the pre-processing stage and a Multi Objective Evolutionary Fuzzy Classifier is applied on the dataset. This algorithm works on the principle of maximum classifier rate and minimum rules. As a result of feature selection with GA the number of features is reduced to 4 from 8 and the classifier rate is improved to 83.0435 % with NSGA II in training rate of 70% and 30% testing.

Using the ADAP learning algorithm to forecast the onset of diabetes mellitus:

Neural networks or connectionist models for parallel processing are not new. However, a resurgence of interest in the past half decade has occurred. In part, this is related to a better understanding of what are now referred to as hidden nodes. These algorithms are considered to be of marked value in pattern recognition problems. Because of that, we tested the ability of an early neural network model, ADAP, to forecast the onset of diabetes mellitus in a high risk population of Pima Indians. The algorithm's performance was analyzed using standard measures for clinical tests: sensitivity, specificity, and a receiver operating characteristic curve. The crossover point for sensitivity and specificity is 0.76. We are currently further examining these methods by comparing the ADAP results with those obtained from logistic regression and linear perceptron models using precisely the same training and forecasting sets. A description of the algorithm is included.

III. PROPOSED SYSTEM

Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed) Engineering and Science Journal Impact Factor:7.984(SJIF) Volume-4, Special Issue-3; ISSN: :2582-5887

Methodology

An expert system for insulin dosage prediction utilizes a combination of Gradient Boosting and Logistic Regression to enhance accuracy and reliability in managing diabetes. The system integrates these two machine learning techniques to provide a more robust and efficient predictive model.

Gradient Boosting is a powerful ensemble learning technique that combines the predictions of multiple weak models to create a strong predictive model. It works by sequentially training models to correct errors made by the previous ones, thereby improving overall accuracy. In the context of insulin dosage prediction, Gradient Boosting can learn complex relationships between various patient features and insulin requirements.

Logistic Regression, on the other hand, is particularly useful for binary classification problems. In the context of diabetes management, it can be employed to predict the likelihood of a patient needing a certain insulin dosage level. Logistic Regression provides a probabilistic interpretation of its predictions, making it suitable for estimating the probability of an event, such as the need for a specific insulin dosage.

The expert system leverages the strengths of both methods by combining their predictions, creating a more robust model for insulin dosage prediction. The system is trained on a diverse dataset that includes patient demographics, lifestyle factors, and historical insulin usage. It adapts to individual patient characteristics, allowing for personalized predictions.

The utilization of Gradient Boosting and Logistic Regression in tandem ensures that the expert system captures both linear and non-linear relationships within the data, providing a comprehensive understanding of the factors influencing insulin requirements. Regular model updates and refinements based on new patient data contribute to the system's adaptability and continuous improvement.

In summary, the expert system for insulin dosage prediction integrates Gradient Boosting and Logistic Regression to create a powerful and adaptive model. This combination enhances accuracy, making it a valuable tool for personalized diabetes management by predicting insulin requirements based on individual patient characteristics and historical data.

In the healthcare industry, patient safety is of utmost importance, and the advancement in connection is crucial in augmenting this facet. By turning on real-time monitoring, healthcare providers can proactively react and reduce potential hazards because the system can quickly identify irregularities or failures. [1] By doing this, the possibility of un favorable outcomes is reduced and a more dependable and safe drug delivery system is established.

The assimilation of remote monitoring technology is in harmony with the wider tele health movement, providing a patient-centered method of healthcare provision. Nowadays, patients can take their meds in the comfort of their own homes, with medical professionals monitoring and adjusting the drug delivery parameters remotely as needed. This improves patient convenience while also making healthcare management more effective,[3] especially for those with long-term illnesses.

Although the development in connectivity has great potential, issues like interoperability and data security need to be resolved before it can be widely used. Continuous research and development endeavors are imperative in order to enhance and broaden the functionalities of this inventive solution, guaranteeing its smooth assimilation into the more extensive healthcare system

DATASET

The PIMA Indian Diabetes Dataset, originally from the National Institute of Diabetes and Digestive and Kidney Diseases, contains information of 768 women from a population near Phoenix, Arizona, USA. The outcome tested was Diabetes, 258 tested positive and 500 tested negative. Therefore, there is one target (dependent) variable and the 8 attributes (TYNECKI, 2018): pregnancies, OGTT(Oral Glucose Tolerance Test), blood pressure, skin thickness, insulin, BMI(Body Mass Index), age, pedigree diabetes function. The Pima population has been under study by the National Institute of Diabetes and Digestive and Kidney Diseases at intervals of 2 years since 1965. As epidemiological evidence indicates that T2DM results from interaction of genetic and environmental factors, the Pima Indians Diabetes Dataset includes information about attributes that could and should be related to the onset of diabetes and its future

Companies.

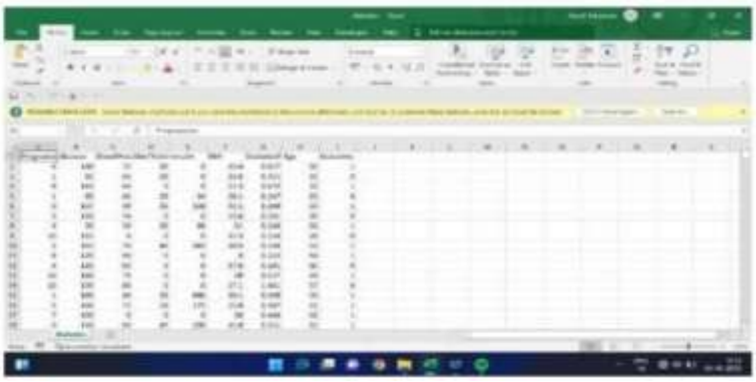


Fig 1 Diabetes Dataset

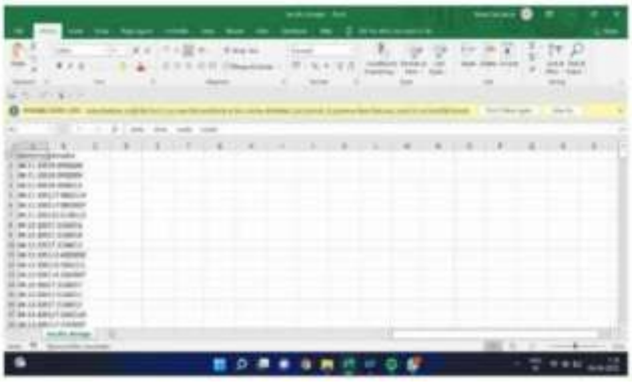


Fig2:Insulin Dataset

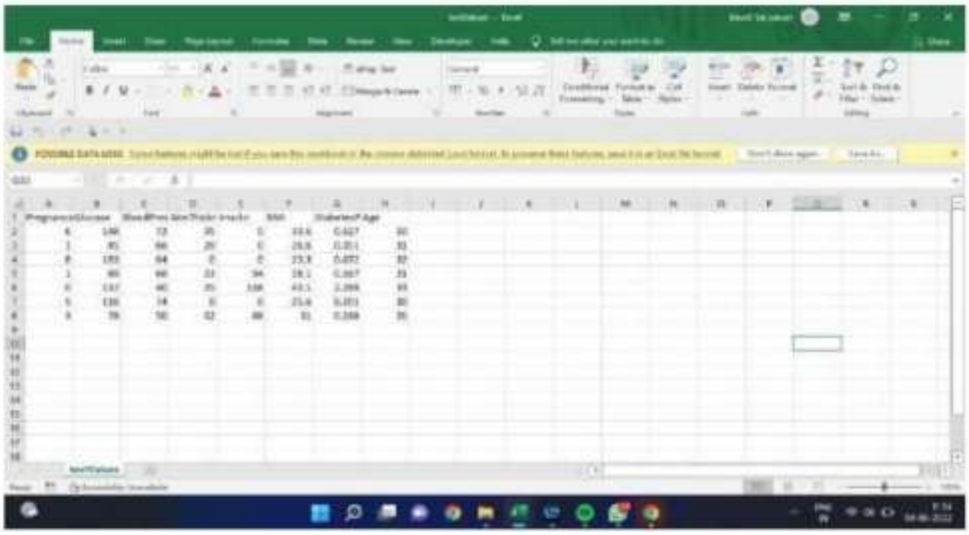
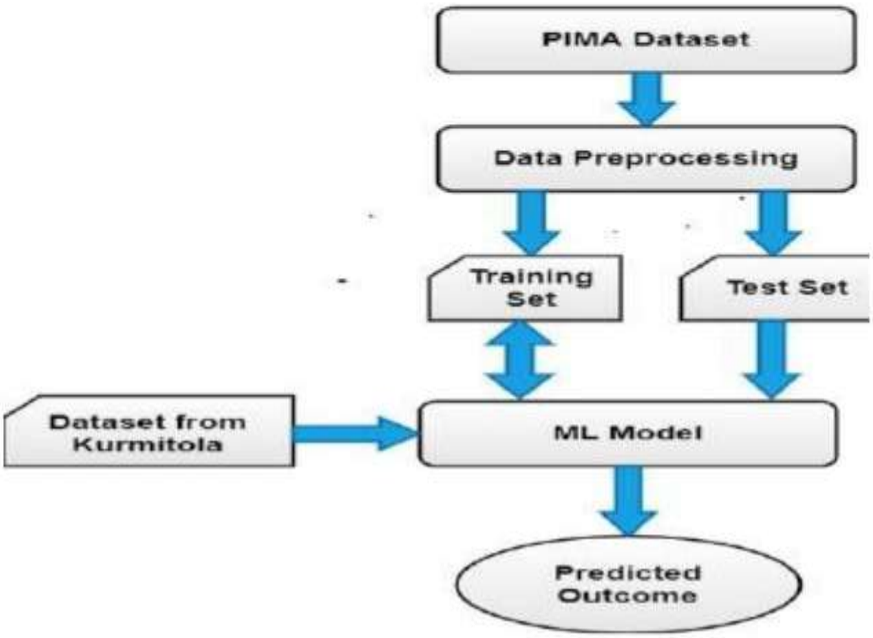


Fig3:TestValue Dataset

ARCHITETURE



Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed) Engineering and Science Journal Impact Factor:7.984(SJIF) Volume-4, Special Issue-3; ISSN: :2582-5887

ALGORITHMS AND FLOWCHART

Algorithm1 (Gradient Boosting)

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensembleof weak prediction models, which are typically decision trees. When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. Gradient boosting involves three elements:

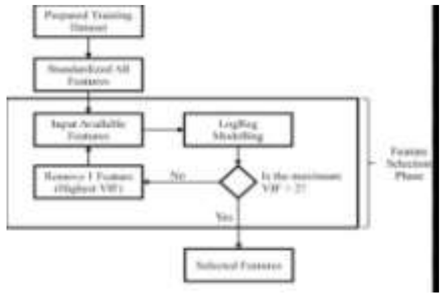
Loss function :A loss function to be optimized. The loss function used depends on the type of problem being solved.

Weak learner :A weak learner to make predictions. Decision trees are used as the weak learner in gradient boosting.

Addictive model :An additive model to add weak learners to minimize the loss function. Trees are added one at a time,and existing trees in the model are not changed.

A gradient descent procedure is used to minimize the loss when adding trees.

Step 1: Creating classification dataset with make classification



Step 2: Building Gradient Boosting Classifier

Step 3: Performing prediction with a classification model

Algorithm2 (Linear Regression*)*

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variableyou want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. It is a regression model that uses a straight line to describe the relationship between variables. It finds the line of best fit through your data by searching for the value of the regression coefficient(s) that minimizes the total error of the model.

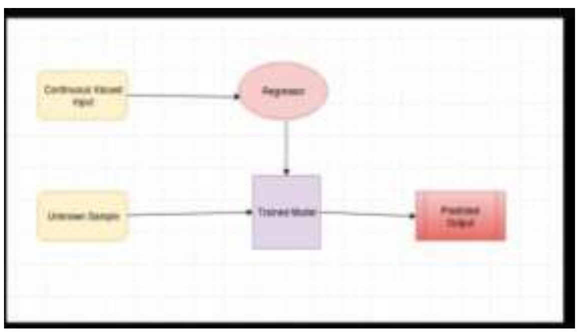
Step 1: Import the packages and classes that you need.

Step 2: Provide data to work with, and eventually do appropriate transformations.

Step 3: Create a regression model and fit it with existing data.

Step 4: Check the results of model fitting to know whether the model is satisfactory.

Step 5: Apply the model for predictions



IV. RESULTS AND DISCUSSION

The outcomes of the proposed system is the level of severity identified from the diabetic patients after performingvarious processing techniques using machine learning algorithms.

Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed) Engineering and Science Journal Impact Factor:7.984(SJIF) Volume-4, Special Issue-3; ISSN: :2582-5887

Processing Screens

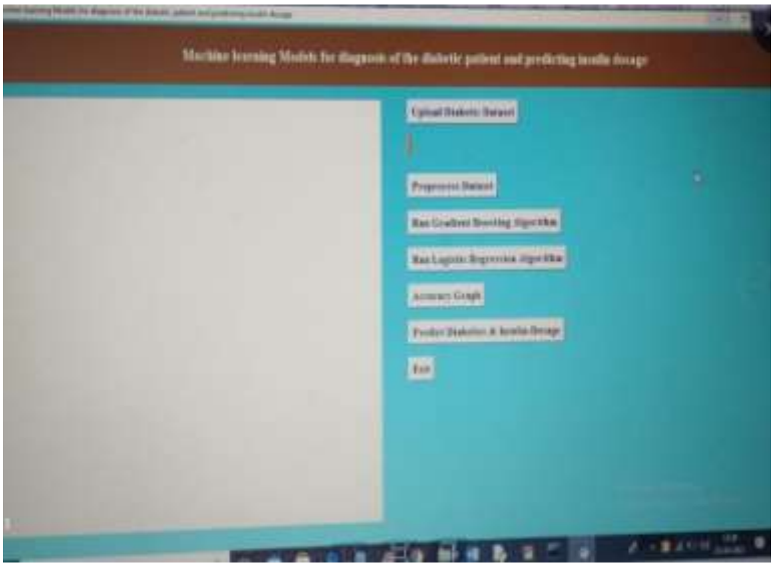


Fig 1:Upload Diabetic Dataset

Fig 2:Uploading Entire Dataset Folder To Load Both Diabetes 0

Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed) Engineering and Science Journal Impact Factor:7.984(SJIF) Volume-4, Special Issue-3; ISSN: :2582-5887

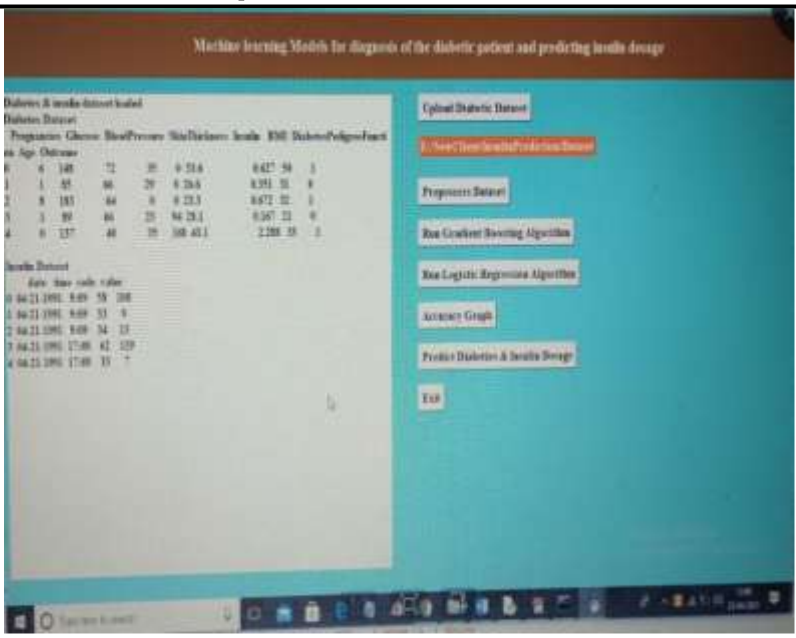


Fig3:Both Diabetics And Insulin Dataset Loaded

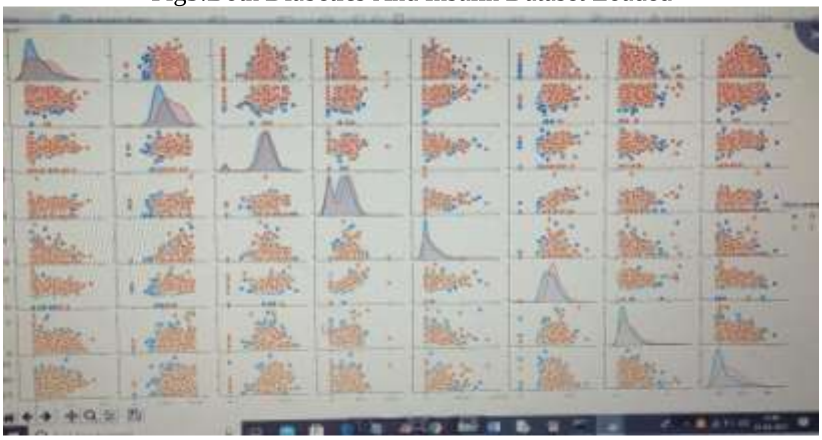


Fig 4: Red Dots Indicates Presences Of Diabetics And Blue Dots Indicates No Diabetics Detected

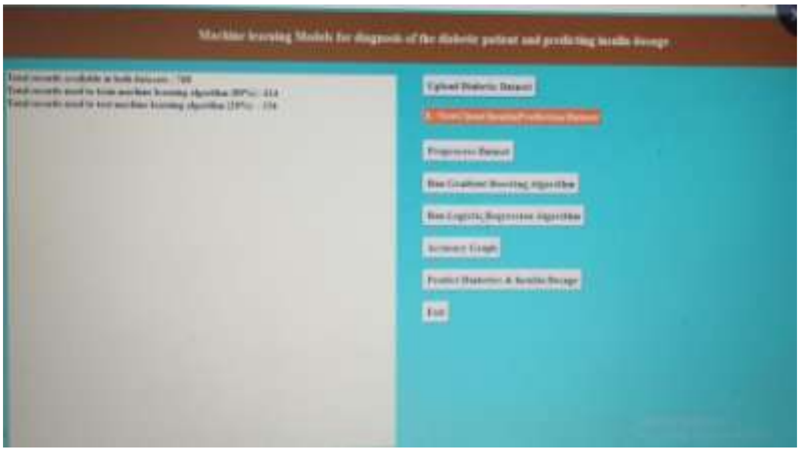


Fig5:Processdataset To Remove Missing Values And To Split Dataset Into Train And Test

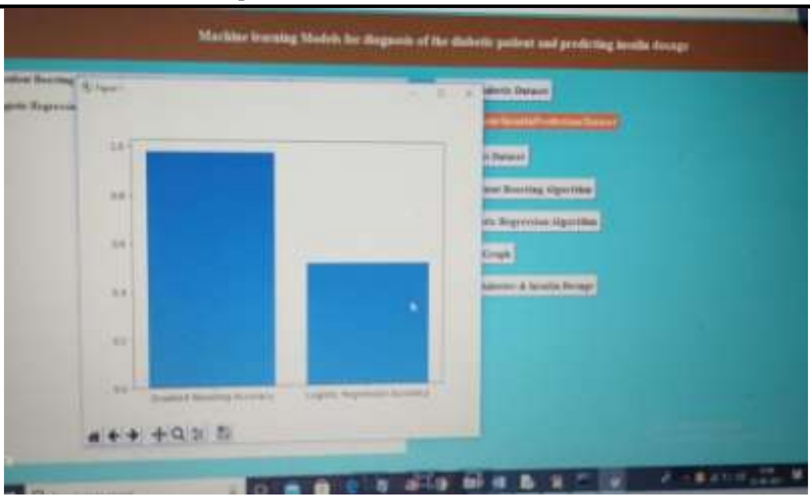


Fig6:Representing Accuracy Of Algorithms

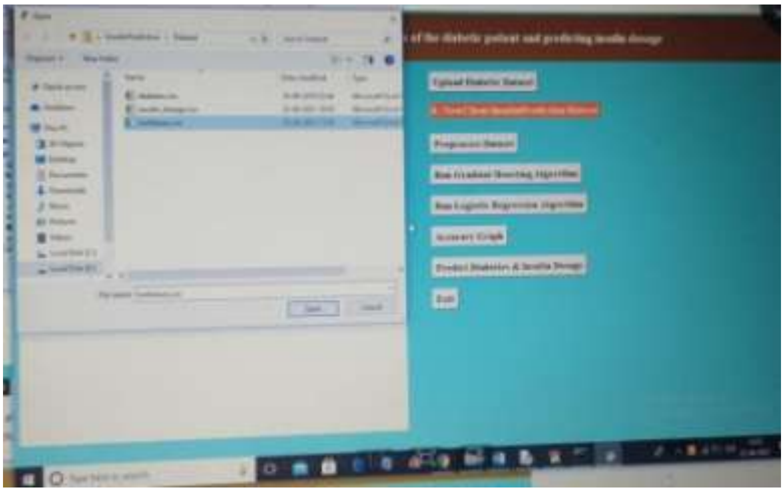


Fig 7:Uploading 'Testvalues.Csv' File

Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed) Engineering and Science Journal Impact Factor:7.984(SJIF) Volume-4, Special Issue-3; ISSN: :2582-5887

Output Screens:

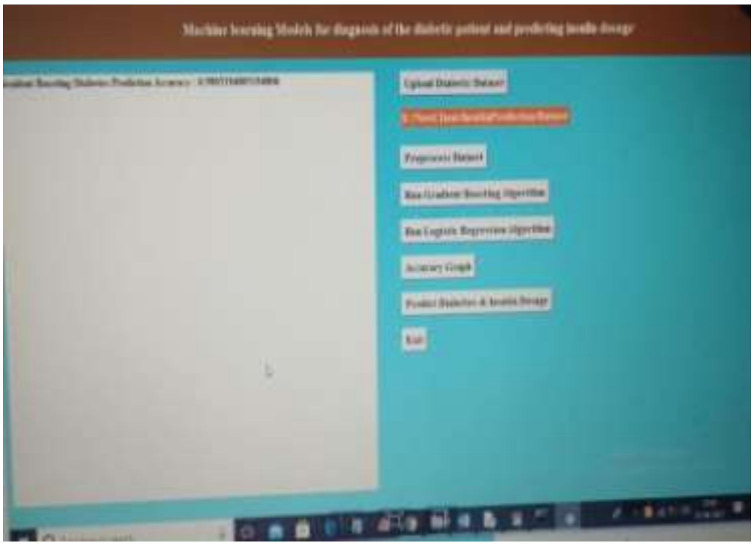


Fig 8:Diadetics Is Predicted With 100% Accuracy

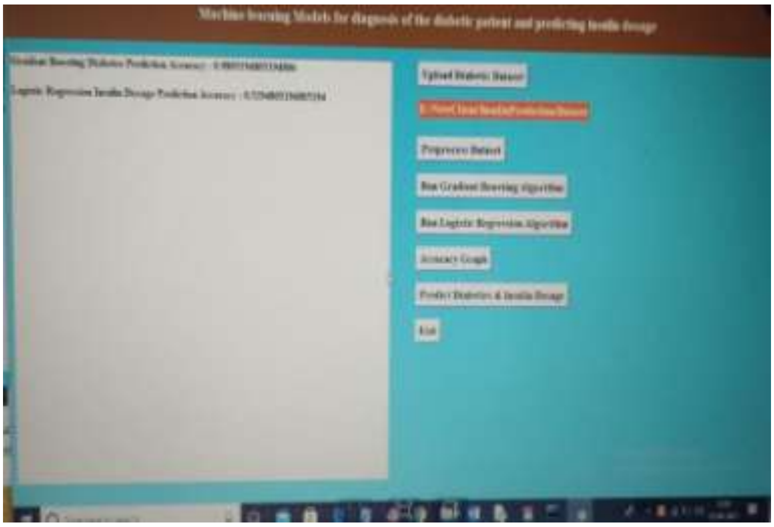


Fig 9:Insulin Dosage Is Predicted With 78% Accuracy

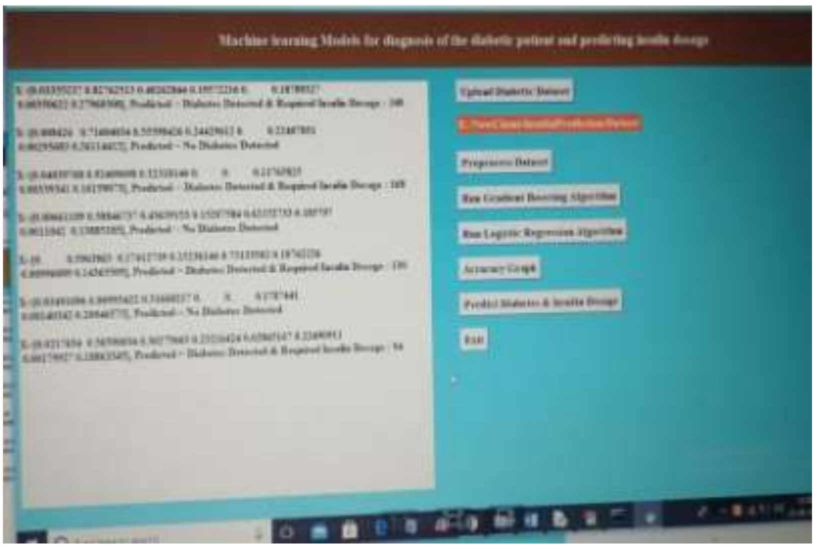


Fig 10 Predicted Result As 'No Diabetes Detected' Or 'Diabetes Detected' And If DiabetesDetected Then Insulin Dosage Predicted

IV. CONSLUSION

In this project we are using Extreme Gradient Boosting Classifier to predict diabetes and then using linear 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 Extreme Gradient Boosting will predict presence of diabetes and Linear Regression will predict insulin dosage if diabetes detected by Extreme Gradient Boosting.

REFERENCES

[1] Misra, H. Gopalan, R. Jayawardena, A. P. Hills, M. Soares, A. A. RezaAlbarrán, and K. L. Ramaiya, Diabetes in developing countries, Journal of Diabetes, vol. 11, no. 7, pp. 522-539, Mar. 2019
[2] R. Vaishali, R. Sasikala, S. Ramasubbareddy, S. Remya, and S. Nalluri, Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset, in Proc. International Conference on Computing Networking and Informatics, Oct. 2017, pp. 1-5.
[3] Emerging Risk Factors Collaboration and other, Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies, The Lancet, vol. 375, no. 9733, pp. 2215- 2222, Jul. 2010.
[4] N. H. Choac, J. E. Shaw, S. Karuranga, Y. Huang, J. D. R. Fernandes, A. W. Ohlrogge, and B. Malandaa, IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045, Diabetes Research and Clinical Practice, vol. 138, pp. 271-281, Apr. 2018.
[5] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karurangaa, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala,
[6] K. Ogurtsova, J. E. Shaw, D. Bright, R. Williams, and IDF Diabetes Atlas Committee, Global and regional diabetes
[7] prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation, Diabetes Research and Clinical Practice, vol. 157, pp. 107843, Nov. 2019.
[8] W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, in Proc. Annual Symposium on Computer Application in Medical Care, Nov. 1988, pp. 261-265.
[9] M. Maniruzzaman, M. J. Rahman, M. A. M. Hasan, H. S. Suri, M. M. Abedin, A. El-Baz, and J. S. Suri, Accurate diabetes risk stratification using machine learning: role of missing value and outliers, Journal of Medical Systems, vol.42, no. 5, pp. 92, May 2018.
[10] G. J. McLachlan, Discriminant analysis and statistical pattern recognition, Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 168, no. 3, pp. 635-636, Jun. 2005.
[11] T. M. Cover, Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, IEEE Transactions on Electronic Computers, vol. 14, no. 3, pp. 326-334, Jun. 1965.
[12] G. I. Webb, J. R. Boughton, and Zhihai Wang, Not So Naive Bayes: Aggregating one-dependence estimators, Machine learning, vol. 58, no. 1, pp. 5-24, Jan. 2005
[13] American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care, Vol. 31, No. 1, 2008, 55-60, 1935- 5548.
[14] 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.
[15] 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.
[16] M. Bishop, . Pattern Recognition and Machine Learning, Springer, 2006, New York
[17] S. Haykin, Neural networks and learning machines. Pearson 2008.
[18] W.D. Patterson, Artificial Neural Networks- Theory and Applications, Prentice Hall , Singapore. 1996.
[19] 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.
[20] 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
[21] 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.
[22] G, Robertson, E. Lehmann, W. Sandham, and D. Hamilton Blood, "Glucose Prediction Using Artificial NeuralNetworks Trained with the AIDA Diabetes Simulator: A Proof-of-Concept Pilot Study". Journal of Electrical andComputer Engineering , V 2011 (2011), Article ID 681786, 11 pages.