IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 6, June 2022

DOI: 10.17148/IJIREEICE.2022.10612

An Expertise System for Insulin Dosage Prediction using Machine Learning Techniques

Dr.K.Vijay Kumar, K.Mounika Sai Sadhvi, M.Bhargavi, K.Sri Varshini, K.Sirisha

Department of Computer Science and Engineering, Vignan‟s Institute of Engineering for Women,

Visakhapatnam, A.P, India

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 Linear Regression algorithm to predict insulin dosage in diabetis 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 Linear Regression will predict insulin dosage if diabetes detected by Gradient Boosting.

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

1. INTRODUCTION:

Diabetes prevalence has been rising more rapidly in middle and low-income countries. The number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014. The global prevalence of diabetes among adults over 18 years of age has risen from 4.7% in 1980 to 8.5% in 2014. A lack of insulin, or an inability to adequately respond to insulin, can each lead to the development of the symptoms of diabetes. For a diabetes patient, insulin dose is necessary to control the level of glucose. The doctor of the patient also has to know the required insulin dose from previous records of doses & from patient’s current calculated blood sugar level. This has inspired us to make a research on how to predict the insulin dose level of a patient before every meal.

Lastly, we want to tell that there are many researches going on to discover newer methods to predict insulin dose.

Despite having all those methods, we believe our research data will help doctors to predict almost accurate insulin dose of diabetes patients.

2. LITERATURE SURVEY

Predicting Diabetes Mellitus using Data Mining Technique: 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, Linear 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 linear regression 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).

DOI: 10.17148/IJIREEICE.2022.10612

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

Comparison Data Mining Techniques to Prediction Diabetes Mellitus: 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.

Diabetes is a disease caused by hyperglycaemia (high blood glucose level). Diabetes affects an estimated 3-4% of the world's population (half of whom are undiagnosed), making it one of the major chronic illnesses prevailing today. It is caused by hyperglycaemia resulting from defects in insulin secretion, insulin action, or both. The chronic hyperglycaemia of diabetes is associated with long-term damage, dysfunction, and failure of various organs, especially the eyes, kidneys, nerves, heart, and blood vessels. This deficiency leads to destruction of the b-cells of the pancreas with consequent insulin deficiency to abnormalities that result in resistance to insulin action and reaction process. Deficiency of insulin results from inadequate insulin secretion. This Improper insulin secretion and defects in insulin action is the primary cause of hyperglycaemia. So, the significance of insulin dose I s clearly visible.

The prediction of glucose concentrations could facilitate the appropriate patient reaction in crucial situations such as hypoglycaemia. Thus, several recent studies have considered advanced data driven 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 patientspecific necessitates the application of non-linear regression models such as artificial neural networks, support vector regression and Gaussian processes.

The aim of project to predict diabetes to predict. 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 modelling of the glucose metabolism.

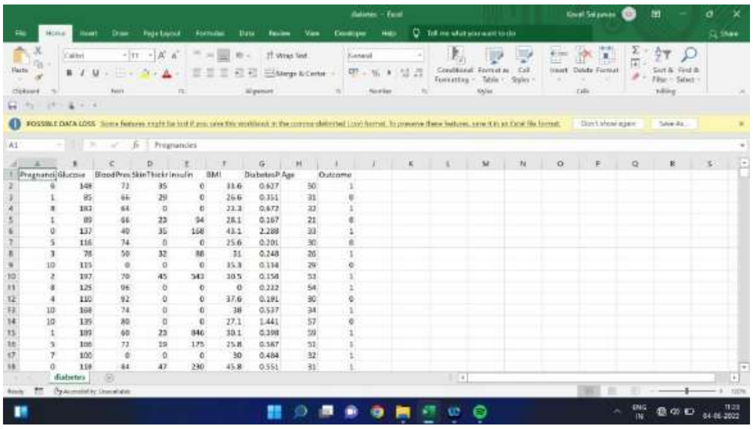
2. PROPOSED SYSTEM

3.1. Methodology

In this project we are using 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.The objective of the PIMA dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. 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 Linear Regression will predict insulin dosage if diabetes detected by Gradient Boosting.

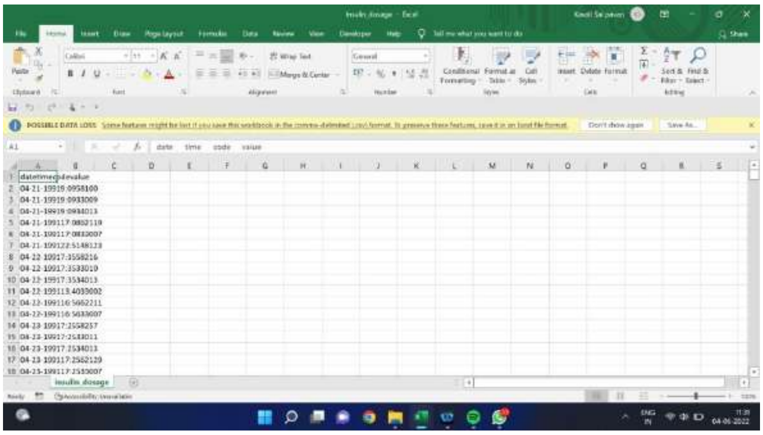
3.2. Dataset

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 complications.



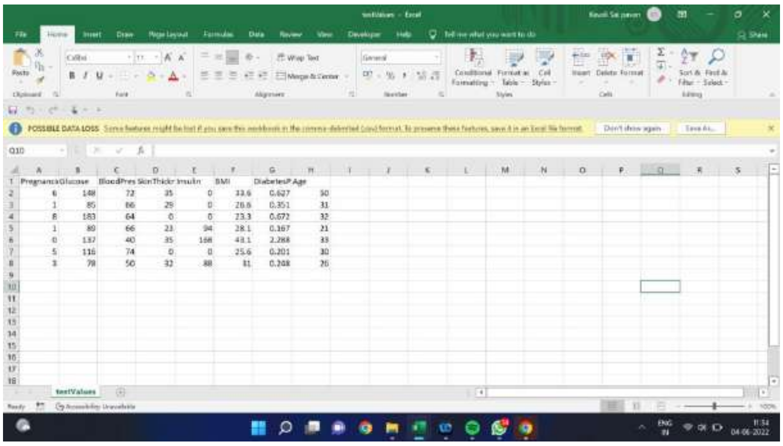
Pregnanci	Glucose	BloodPres	SkinThickr	Insulin	BMI	DiabetesP	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1

Figure 1: Diabetes Dataset



datetimecodevalue
04-21-19919:0958100
04-21-19919:0933009
04-21-19919:0934013
04-21-199117:0862119
04-21-199117:0833007
04-21-199122:5148123
04-22-19917:3558216
04-22-19917:3533010
04-22-19917:3534013
04-22-199113:4033002
04-22-199116:5662211
04-22-199116:5633007
04-23-19917:2558257
04-23-19917:2533011
04-23-19917:2534011
04-23-199117:2562129
04-23-199117:2533007

Figure 2: Insulin Dataset



Pregnanci	Glucose	BloodPres	SkinThickr	Insulin	BMI	DiabetesP	Age
6	148	72	35	0	33.6	0.627	50
1	85	66	29	0	26.6	0.351	31
8	183	64	0	0	23.3	0.672	32
1	89	66	23	94	28.1	0.167	21
0	137	40	35	168	43.1	2.288	33
5	116	74	0	0	25.6	0.201	30
3	78	50	32	88	31	0.248	26

Figure 3: Testvalues Dataset

DOI: 10.17148/IJIREEICE.2022.10612

3.3. Architecture

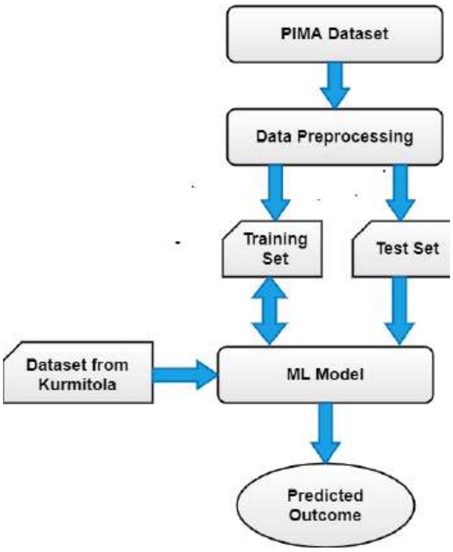


Figure 4: Architecture Diagram

3. ALGORITHMS AND FLOW CHART

4.

4.1. 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

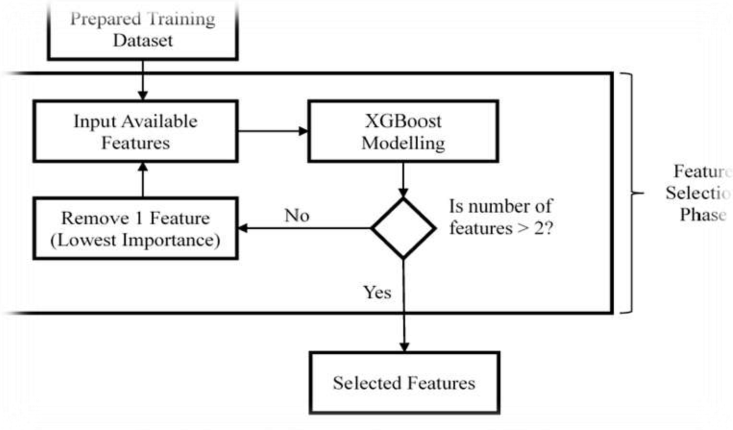

Prepared Training Dataset
Input Available Features
XGBoost Modelling
Remove 1 Feature (Lowest Importance)
No
Is number of features > 2?
Yes
Featur
Selectio
Phase
Selected Features

Figure 5: flowchart of GradientBoosting

4.2. Algorithm2 (Linear Regression)

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you 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

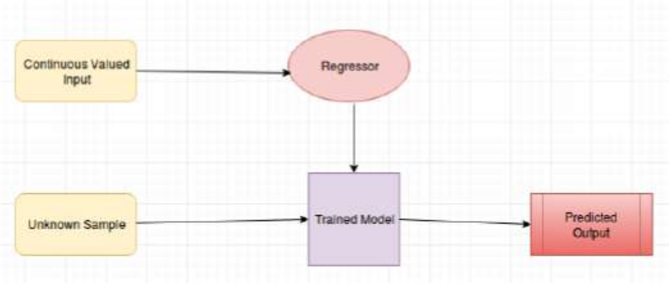
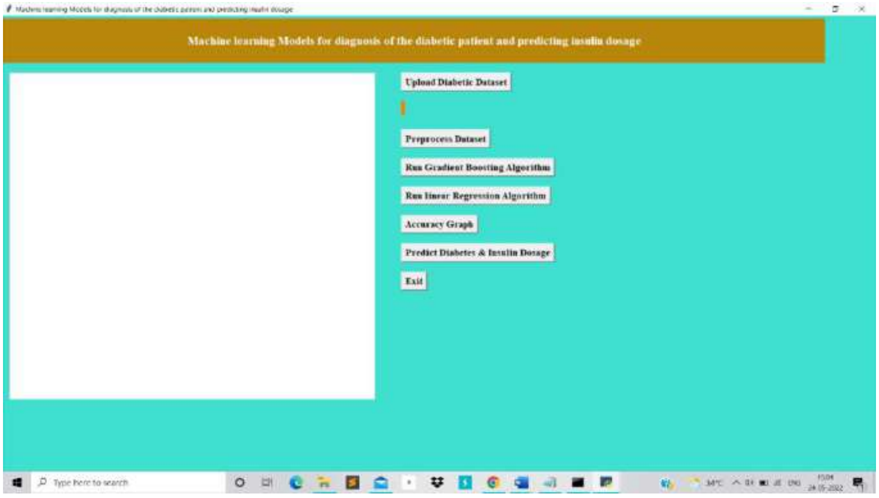


Figure 6: Linear Regression

5. RESULTS AND DISCUSSION

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

5.1. Processing Screens


Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage

Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Upload Diabetic Dataset
Preprocess Dataset
Run Gradient Boosting Algorithm
Run linear Regression Algorithm
Accuracy Graph
Predict Diabetes & Insulin Dosage
Exit
Type here to search

Figure 7:Upload Diabetic Dataset

DOI: 10.17148/IJIREEICE.2022.10612

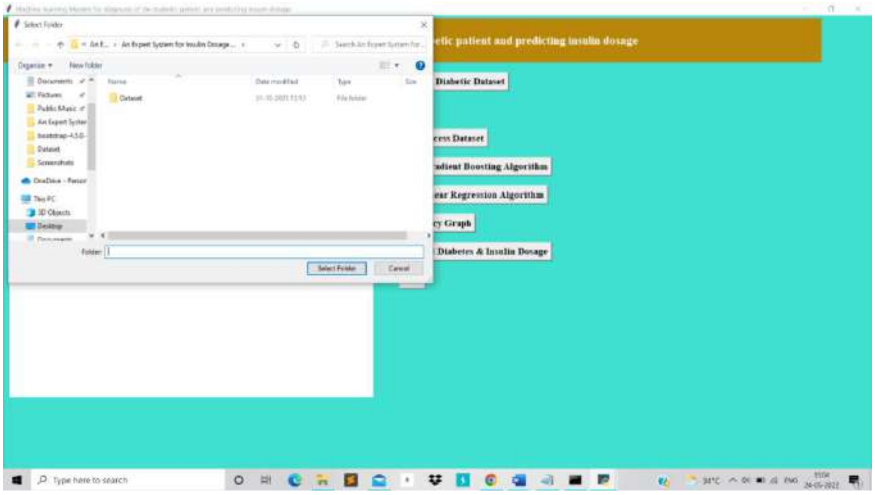
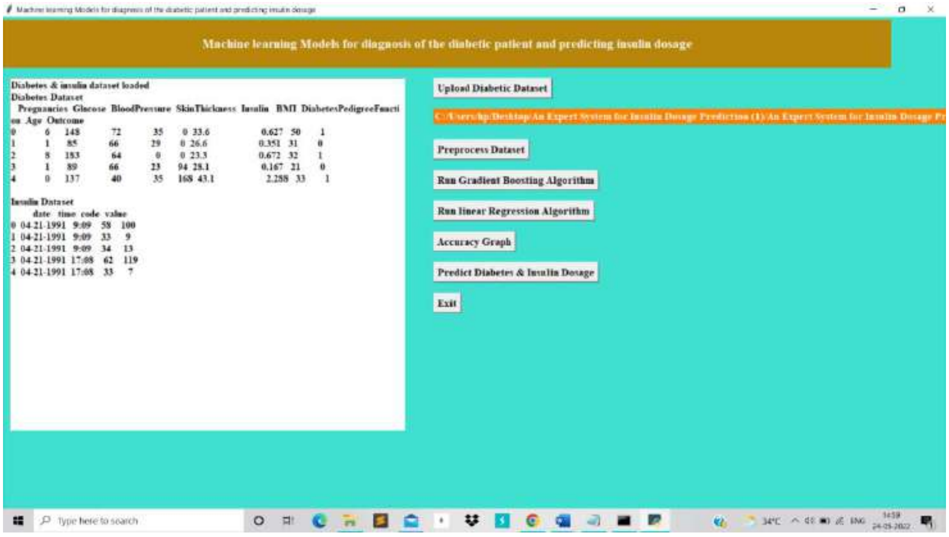


Figure 8: uploading entire Dataset folder to load both diabetes and insulin   datasets


Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Upload Diabetic Dataset
C:/Users/hp/Desktop/An Expert System for Insulin Dosage Prediction (1)/An Expert System for Insulin Dosage Pr
Preprocess Dataset
Run Gradient Boosting Algorithm
Run linear Regression Algorithm
Accuracy Graph
Predict Diabetes & Insulin Dosage
Exit
Type here to search

```
Diabetes & insulin dataset loaded
Diabetes Dataset
  Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  DiabetesPedigreeFuncti
on  Age  Outcome
0         6     148          72            35        0  33.6             0.627   50       1
1         1      85          66            29        0  26.6             0.351   31       0
2         8     183          64             0        0  23.3             0.672   32       1
3         1      89          66            23       94  28.1             0.167   21       0
4         0     137          40            35      168  43.1             2.288   33       1

Insulin Dataset
        date   time  code  value
0  04-21-1991   9:09   58    100
1  04-21-1991   9:09   33      9
2  04-21-1991   9:09   34     13
3  04-21-1991  17:08   62    119
4  04-21-1991  17:08   33      7
```

Figure 9: Both diabetis and insulin dataset loaded

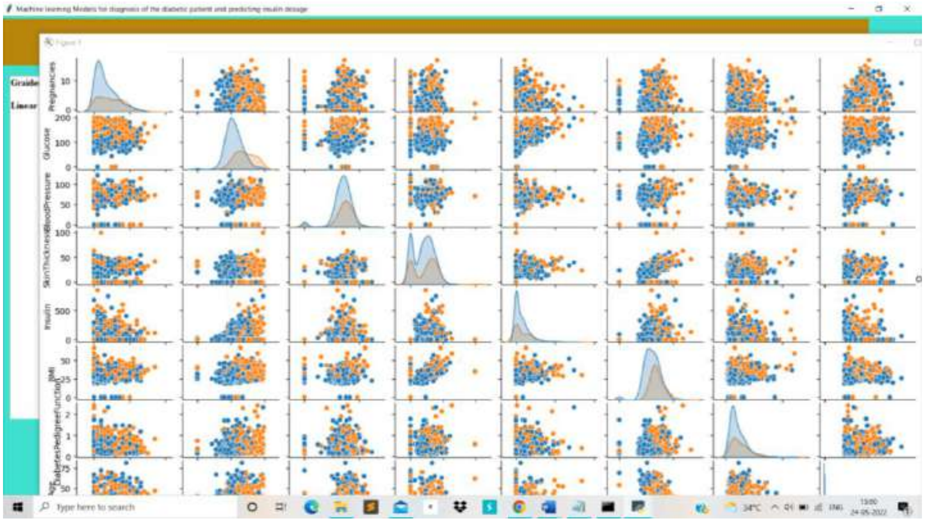
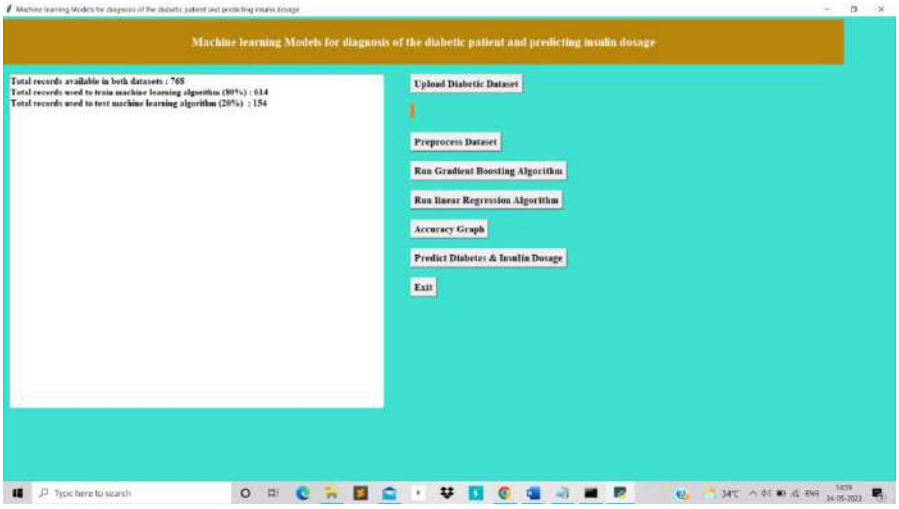


Figure 10: red dots indicate presence of diabetes and blue represents no diabetes detected

DOI: 10.17148/IJIREEICE.2022.10612


Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Upload Diabetic Dataset
Preprocess Dataset
Run Gradient Boosting Algorithm
Run linear Regression Algorithm
Accuracy Graph
Predict Diabetes & Insulin Dosage
Exit
Type here to search

```
Total records available in both datasets : 768
Total records used to train machine learning algorithm (80%) : 614
Total records used to test machine learning algorithm (20%) : 154
```

Figure 11: ‘PreprocessDataset’ to remove missing values and to split dataset into train and test

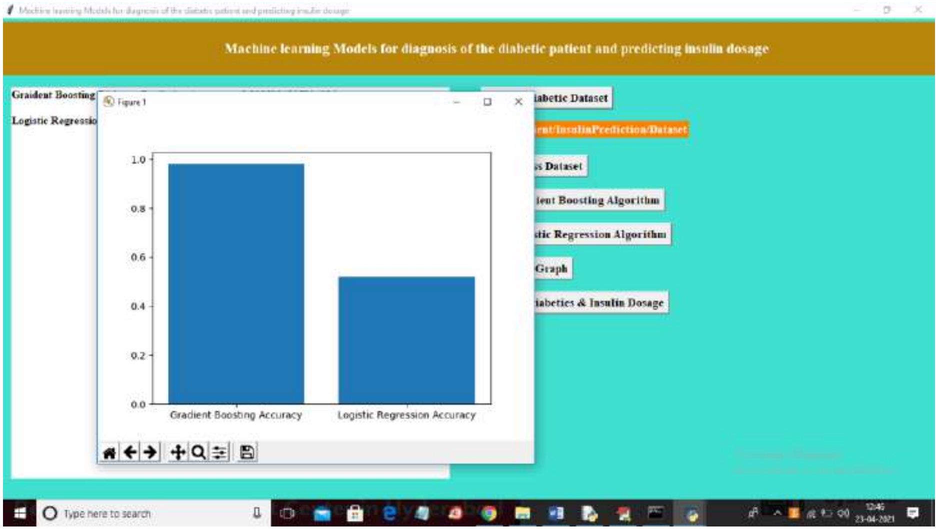
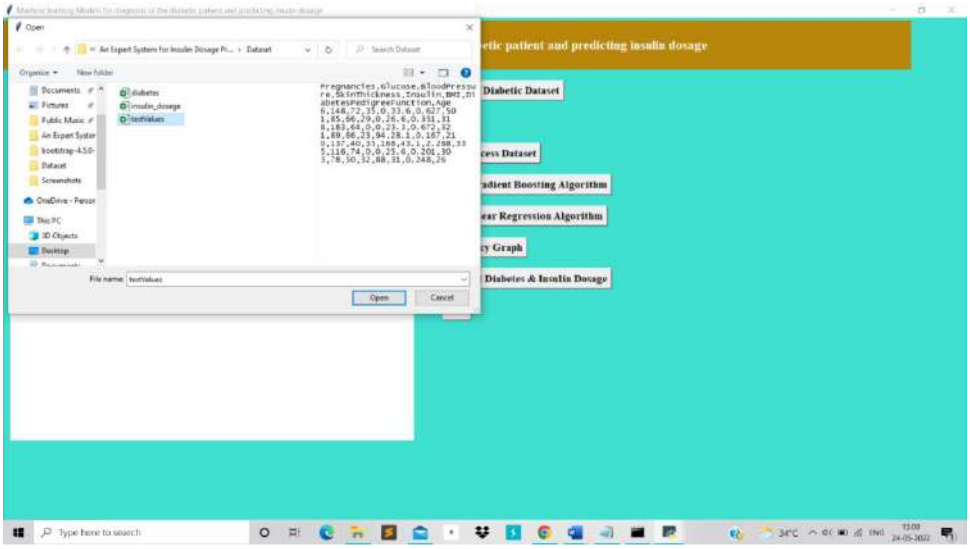


Figure 12: representing accuracy of algorithms

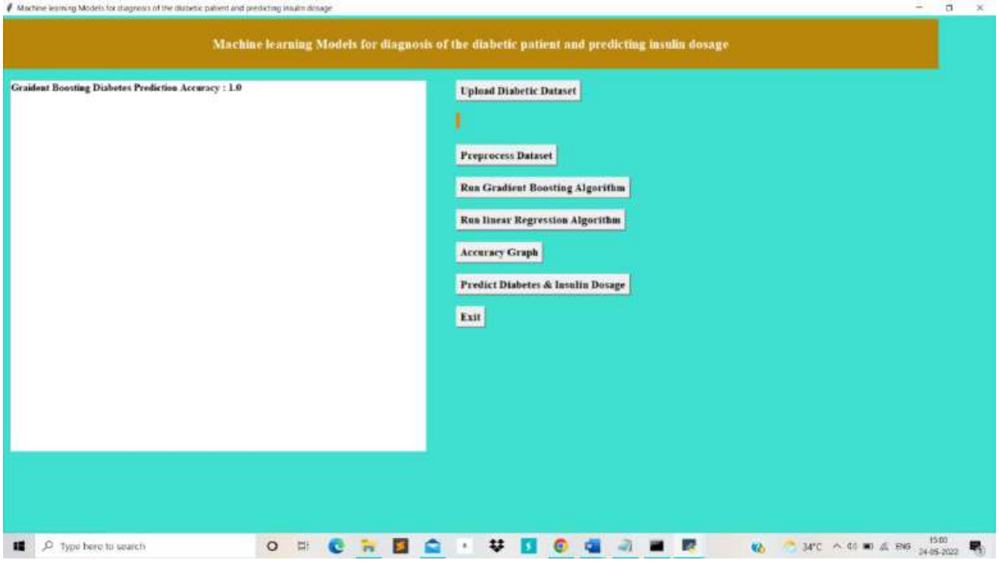


```
Pregnancies,Glucose,BloodPressu
re,SkinThickness,Insulin,BMI,Di
abetesPedigreeFunction,Age
6,148,72,35,0,33.6,0.627,50
1,85,66,29,0,26.6,0.351,31
8,183,64,0,0,23.3,0.672,32
1,89,66,23,94,28.1,0.167,21
0,137,40,35,168,43.1,2.288,33
5,116,74,0,0,25.6,0.201,30
3,78,50,32,88,31,0.248,26
```

Figure 13: Uploading ‘testValues.csv’ file

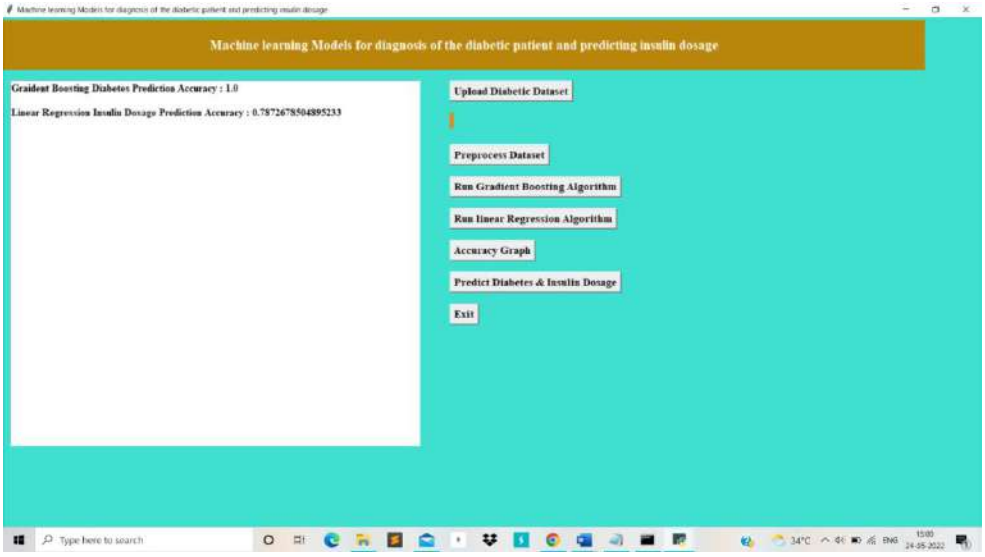
DOI: 10.17148/IJIREEICE.2022.10612

5.2. Output Screens


Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Upload Diabetic Dataset
Preprocess Dataset
Run Gradient Boosting Algorithm
Run linear Regression Algorithm
Accuracy Graph
Predict Diabetes & Insulin Dosage
Exit
Type here to search

```
Graident Boosting Diabetes Prediction Accuracy : 1.0
```

Figure 14: Diabetis is predicted with 100% accuracy

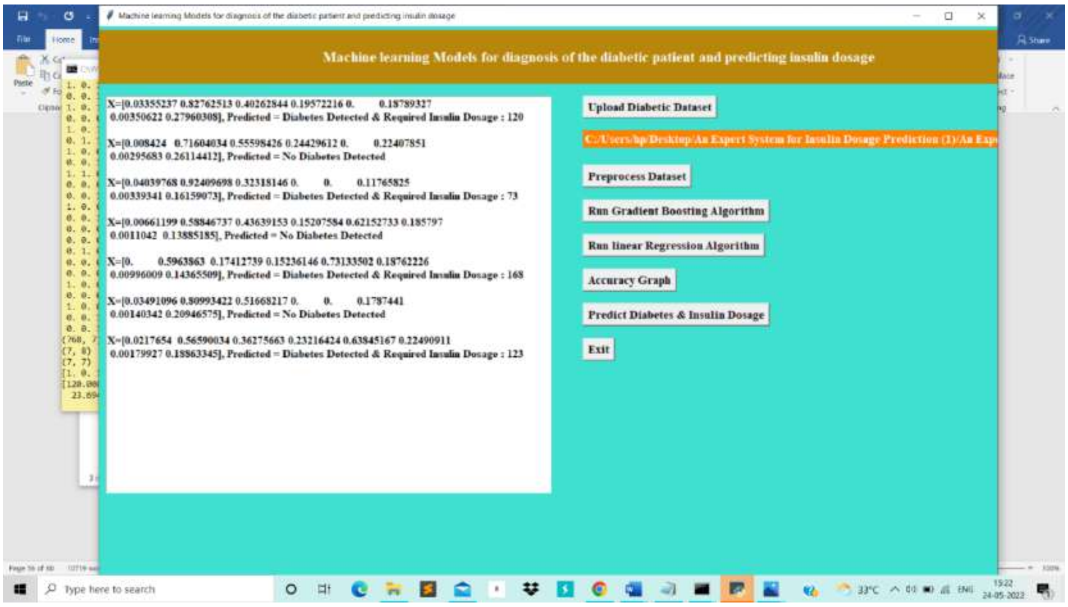

Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Upload Diabetic Dataset
Preprocess Dataset
Run Gradient Boosting Algorithm
Run linear Regression Algorithm
Accuracy Graph
Predict Diabetes & Insulin Dosage
Exit
Type here to search

```
Graident Boosting Diabetes Prediction Accuracy : 1.0

Linear Regression Insulin Dosage Prediction Accuracy : 0.7872678504895233
```

Figure 15: Insulin dosage is predicted with 78% accuracy

DOI: 10.17148/IJIREEICE.2022.10612


Machine learning Models for diagnosis of the diabetic patient and predicting insulin dosage
Upload Diabetic Dataset
C:/Users/hp/Desktop/An Expert System for Insulin Dosage Prediction (1)/An Exp
Preprocess Dataset
Run Gradient Boosting Algorithm
Run linear Regression Algorithm
Accuracy Graph
Predict Diabetes & Insulin Dosage
Exit

```
X=[0.03355237 0.82762513 0.40262844 0.19572216 0.         0.18789327
 0.00350622 0.27960308], Predicted = Diabetes Detected & Required Insulin Dosage : 120

X=[0.008424   0.71604034 0.55598426 0.24429612 0.         0.22407851
 0.00295683 0.26114412], Predicted = No Diabetes Detected

X=[0.04039768 0.92409698 0.32318146 0.         0.         0.11765825
 0.00339341 0.16159073], Predicted = Diabetes Detected & Required Insulin Dosage : 73

X=[0.00661199 0.58846737 0.43639153 0.15207584 0.62152733 0.185797
 0.0011042  0.13885185], Predicted = No Diabetes Detected

X=[0.         0.5963863  0.17412739 0.15236146 0.73133502 0.18762226
 0.00996009 0.14365509], Predicted = Diabetes Detected & Required Insulin Dosage : 168

X=[0.03491096 0.80993422 0.51668217 0.         0.         0.1787441
 0.00140342 0.20946575], Predicted = No Diabetes Detected

X=[0.0217654  0.56590034 0.36275663 0.23216424 0.63845167 0.22490911
 0.00179927 0.18863345], Predicted = Diabetes Detected & Required Insulin Dosage : 123
```

Figure 16: predicted result as 'No Diabetes Detected' or 'Diabetes Detected' and if diabetes detected then insulin dosage predicted

6. CONCLUSION

This paper was aimed at modelling neural network 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 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 to predict diabetes and then applying Linear 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.

REFERENCES

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

[2] I. EleniGeorga, C. Vasilios Protopappas and I. DimitriosFotiadi,. "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.