iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4651

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Diabetes Disease Prediction using Neural Network

Jatin N Bagrecha[1], Chaithra G S[1], Jeevitha S[1]
[1, 2, 3]*Department of Computer Science Engineering, Dayananda Sagar College of Engineering, Karnataka, India*

Abstract: ***This project aims in assisting doctors to predict diabetes disease in prior based on present health parameters like blood plasma, age ,insulin level, pregnancy, body mass, skin thickness, paediatric conditions and blood pressure. Machine learning tools have huge impact in medical field day by day. It is quite difficult to make correct decisions in future prediction of disease. But this project uses convolution neural network system to make efficient classification. Here classification happens as Diabetic or Non-Diabetic based on the health parameters. The results obtained has an accuracy level of 84% and further the accuracy can be enhanced by more interesting deep neural networks, which is a further improvement step for this project. Multilayer perceptron neural network is the algorithm used for binary classification of diabetes. It involves feature analysis of all those 8 parameters and their reflection on being diabetic or not. This is a computer aided system, which doesn't require frequent blood tests of patients in order to make predictions. Henceforth, saves both time and money making the hospital system efficient. The GUI is developed to fetch the data to send it for backend analysis. The dataset used here is Pima Indian Diabetes Dataset which is a collection of 768 patients' health records.***

Keywords: Health Parameters, Convolution Neural Network, Pima Indian Dataset, Feature Analysis, GUI, Machine Learning

I. INTRODUCTION

These days, diabetes has become a serious issue and has fourth most high mortality rate. The daily diet of modern people has huge amount of sugar and fat. Unhealthy eating conditions may lead to diabetes disease in future. Early diagnosis of this disease plays a vital role. Diabetes when not in control may lead to major damage in kidney, heart and liver etc.

Neural networks are prime machine learning algorithms for accurate prognosis of disease in medical field. It uses health parameters as input for neurons, here 8 input parameters are given as input for the neural network. Initially, weights are assigned randomly for neuron and later on during the training phase weights are updated accordingly. Iteration continues, until the output of neural network matches with target output.

The algorithm used here is multilayer perceptron neural network. It has 8 inputs in the visible layer and 32 and 16 neurons in the hidden layers respectively with Relu activation function. Output layer has 1 neuron with Sigmoid activation function. Neural network is trained for 700 epochs. The batch size of these epochs is 10. ADAM optimizer and binary cross entropy loss function enhances the performance of the network.

II. LITERATURE SURVEY

Thiyagarajan C, Dr. K. Anandha Kumar, Dr. A. Bharathi proposed "A Survey on Diabetes Mellitus Prediction Using Machine Learning Techniques" [1]. Data mining and machine learning algorithms are used in medical field for diagnosis and treatment of diseases. .Machine learning algorithm finds the hyper-plane to make a efficient and accurate classification.

Kumari Deepika and Dr. S. Seema put forwarded "Predictive Analysis to prevent and control Chronic Disease" [2]. The algorithms used to perform predictive analysis are Decision tree, Naïve Bayes and SVM. Experimental results validates that Naïve Bayes classifier gives highest accuracy of 73.558% for diabetes disease classification.

Panigrahi Srikanthad and Dharmaiah Deverapalli propounded "A Critical Study of Classification Algorithms Using Diabetes Diagnosis" [3]. In this paper, a comparative study of Decision tree algorithm, Rule based algorithm and Bayes algorithm is done. On analysis it's proven that decision tree algorithms like J48, AD tree and VF tree outperforms than other two.

Nirmala Devi. M, Appavu alias Balamurugan. S, Swathi U.V proposed "An amalgam KNN model to predict Diabetes mellitus" [4].An amalgam obtained as result of combining K-means model with K-nearest neighbour model yields a greater accuracy than native other algorithms. It uses Weka software tool along with multi-step preprocessing techniques to make a classification as Diabetic or not.

Fikirte Girma Woldemichael, Sumitra Menaria propounded "Prediction of Diabetes using Data Mining Techniques" [5]. Backpropogation algorithm is used to train the model with 8 input parameters, one hidden layer with 6 neurons and 1 output layer. In order to evaluate the model fivefold-cross validation is used. The performance was found to be better than existing old techniques like J48, Naïve Bayes and SVM.

Roxana Mirshahvalad and Nastaran Asadi Zanjani put forwarded "Diabetes Prediction Using Ensembled Perceptron Algorithm" [6].Perceptron algorithm alone does not have good performance hence ensembling of it with other algorithms is crucial. This paper depicts pocket algorithm with perceptron algorithm to make an efficient ensembled algorithm. The AUC performance on ensembling got improved from 0.72 to 0.75.

Miss Sneha Joshi and Prof. Megha Borse proposed "Detection and Prediction of Diabetes Mellitus Using Back-Propagation Neural Network" [7]. This paper depicts the implementation and development of software tool built in MATLAB. The algorithm used is Back-Propagation Neural Network (BPNN) and performance ranges around 81 percent.

T. Jayalakshmi and Dr. A. SanthaKumaran propounded "Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks" [8].This paper summarizes the importance of pre-processing and missing values. Combining pre-processing methodologies along with agorithm can yield a greater accuracy.

Sushruta Mishra, Pamela Chaudhury, Brojo Kishore Mishra and Hrudaya Kumar Tripathy proposed "An Implentation of
Feature Ranking Using Machine Learning Techniques for Diabetes disease prediction" [9]. A comparative analysis is done among four filter based feature selection methods (Chi-Square method, Information gain method, Cluster variation method and Correlation method).The study revealed that the performance got accelerated with filter based feature selection.

Rajeeb Dey, Vaibhav Bajpai, Gagan Gandhi and Barnali Dey put forwarded "Application of Artificial Neural Network (ANN) technique for Diagnosis of Diabetes Mellitus" [10]. In this paper, artificial neural network is trained with low running rate and more number of epochs to obtain higher accuracy. The study concluded that with increasing number of hidden layers with more neurons and high learning can give better performance.

III. PROPOSED SYSTEM

The proposed system is a user friendly website whose frontend is designed using PHP and backend has python programming. The website as a frontend tool accepts user data and sends it for backend processing.

A. Building Model

In order to build the model, import required libraries and load the dataset in .csv format. The algorithm contains two variables, namely X and Y. X holds the parameters to predict result and Y has predicted result.

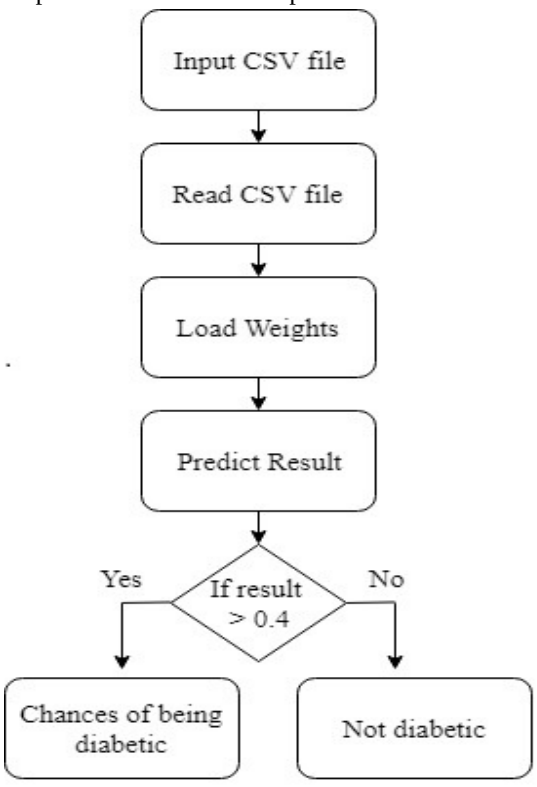


Fig. 1 Block Diagram of Proposed Work

The data is splitted in the ratio of 80:20 as training and testing respectively. Create neural network model with 8 neurons in input layer and 1 neuron in output layer with 2 hidden layers consisting of 32 and 16 neurons respectively. Fit the model with 10 batch size containing 700 iterations in each batch size. Save the model by converting to JSON file. Then save the weights obtained for future prediction.

The figure 2, represents the architecture of the multilayer perceptron neural network algorithm. This algorithm consists of 8-32-16-1 neurons. The hidden layer has ReLu activation function. Rectified Linear Unit (ReLu) activation function is used in convolution neural networks and is half rectified.

When z is less than zero, the function f(z) is zero else f(z) is equal to z for zero and greater than zero. This indicates that all negative inputs turn out to be zero in ReLu activation function which results in unmapping of negative values.

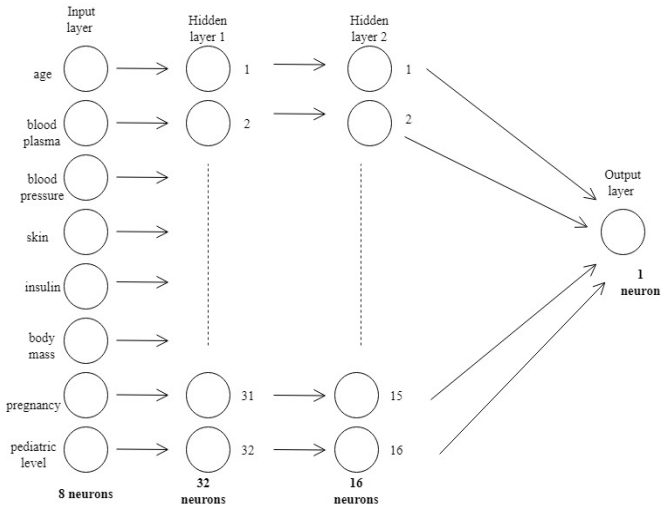


Fig. 2 Neural Network Architecture

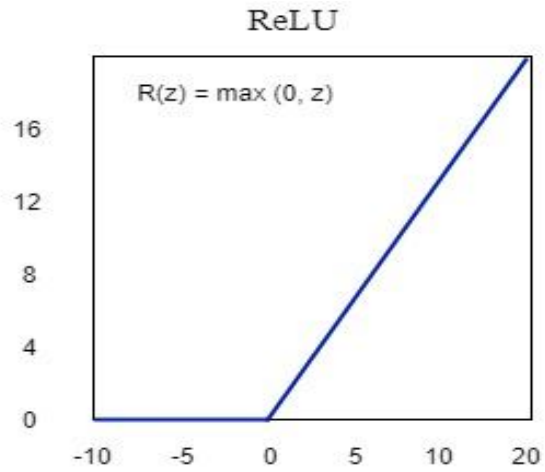


Fig. 3 ReLu Activation Function

The output layer has Sigmoid activation function. It is a 'S' shaped graph and non-linear in nature. Mathematically, it is expressed as:

$f(x) = 1/(1+ e^{-x})$ (1)

The output of this function ranges between 0 and 1.

B. Prediction

In the prediction phase, import required libraries and open the JSON file for reading and storing the content in them. The dataset is loaded in csv format and the values read are stored in an array. The result obtained as an output from neural network undergoes comparison. If result is greater than 0.4, the outcome is 1 indicating to be diabetic in near future else 0.

TABLE I
DATASET DESCRIPTION

Sl No	Attribute	Description
1	Preg	Number of times pregnant
2	Plas	Blood plasma level
3	Pres	Blood pressure
4	Skin	Skin fold thickness
5	Insu	Insulin level
6	Mass	Body mass index
7	Pedi	Paediatric level
8	Age	Age of the individual

IV. WORKING METHODOLOGY

A. Diabetes Disease Prediction Website Appears As

Fig. 4 Home page of the Website

1) Step 1: Load the Dataset For testing Single patient data: Through input form, all 8 parameter values is taken from user and on submit the .csv file gets generated for further processing of the model. Figure 5, depicts this.When multiple patients' records need to be checked and tested at the same time, then their overall data can be written in single .csv file and it can be uploaded for further processing. The figure 6 shows multiple data acceptance through file input.

2) Step 2: Parsing the input data in this step, the excel or .csv file generated are read one by one and sent for machine learning model for testing.

3) Step 3: Feature Extraction: During training phase, the model learns the importance of each feature in resulting patient as diabetic or not. Henceforth, while testing the most affecting features are given more priority to yield any result. In the feature extraction process set of derived values (features) are obtained from initial set of values. The figure 7 depicts the feature importance graph and from the graph it is clear that, blood plasma has highest importance in causing diabetes.

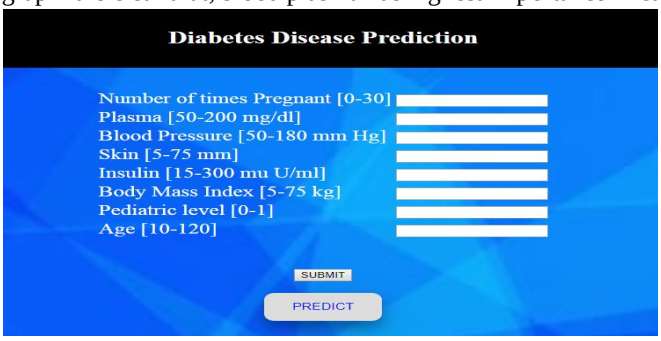


Fig. 5 Input Form to accept Single patient Data

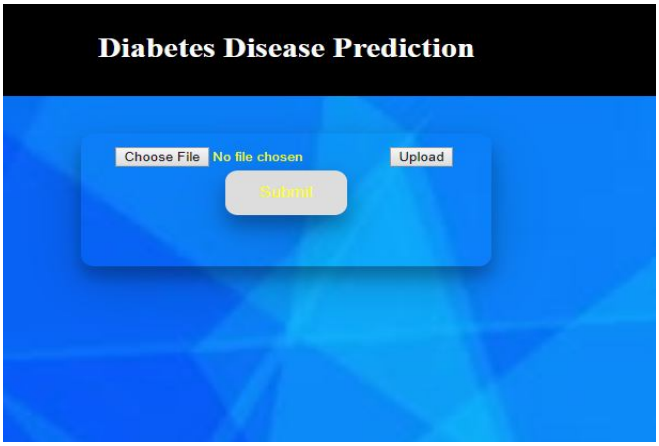


Fig. 6 Multiple Patients' Records can be uploaded in single CSV form

4) *Step 4:* Result analysis on submitting, the data in input form or .csv file is fetched by the neural network. The algorithm runs over those data and appropriate result is shown in output screen. Figure 8, depicts the prediction result of single patient. Figure 9, illustrates the prediction results of multiple patients.

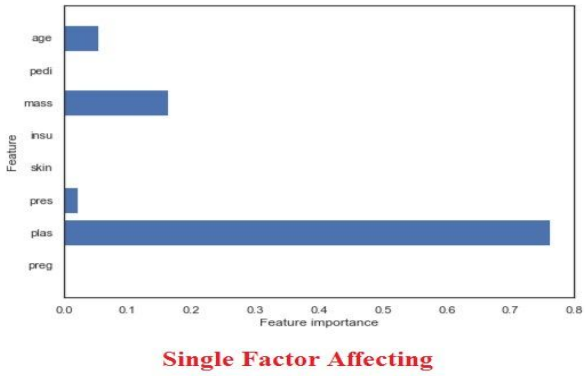


Fig. 7 Feature Importance Graph

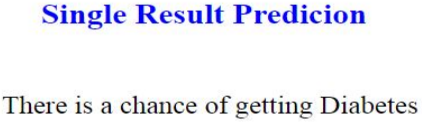


Fig. 8 Report of Single Patient

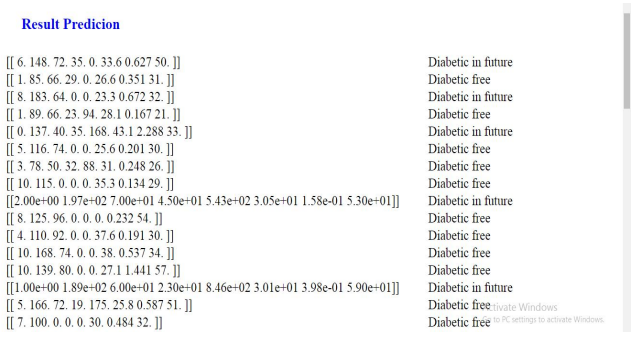


Fig. 9 Report of Multiple Patients

V. RESULTS

The experimental result obtained has an accuracy of 84.4%. The results here fall into two categories Diabetes free or Chances of getting Diabetes in near future. The feature extraction plays crucial role in result prediction. Reports of single patient and multiple patients can be generated in a single run of the application through form and file input respectively.

VI. CONCLUSION

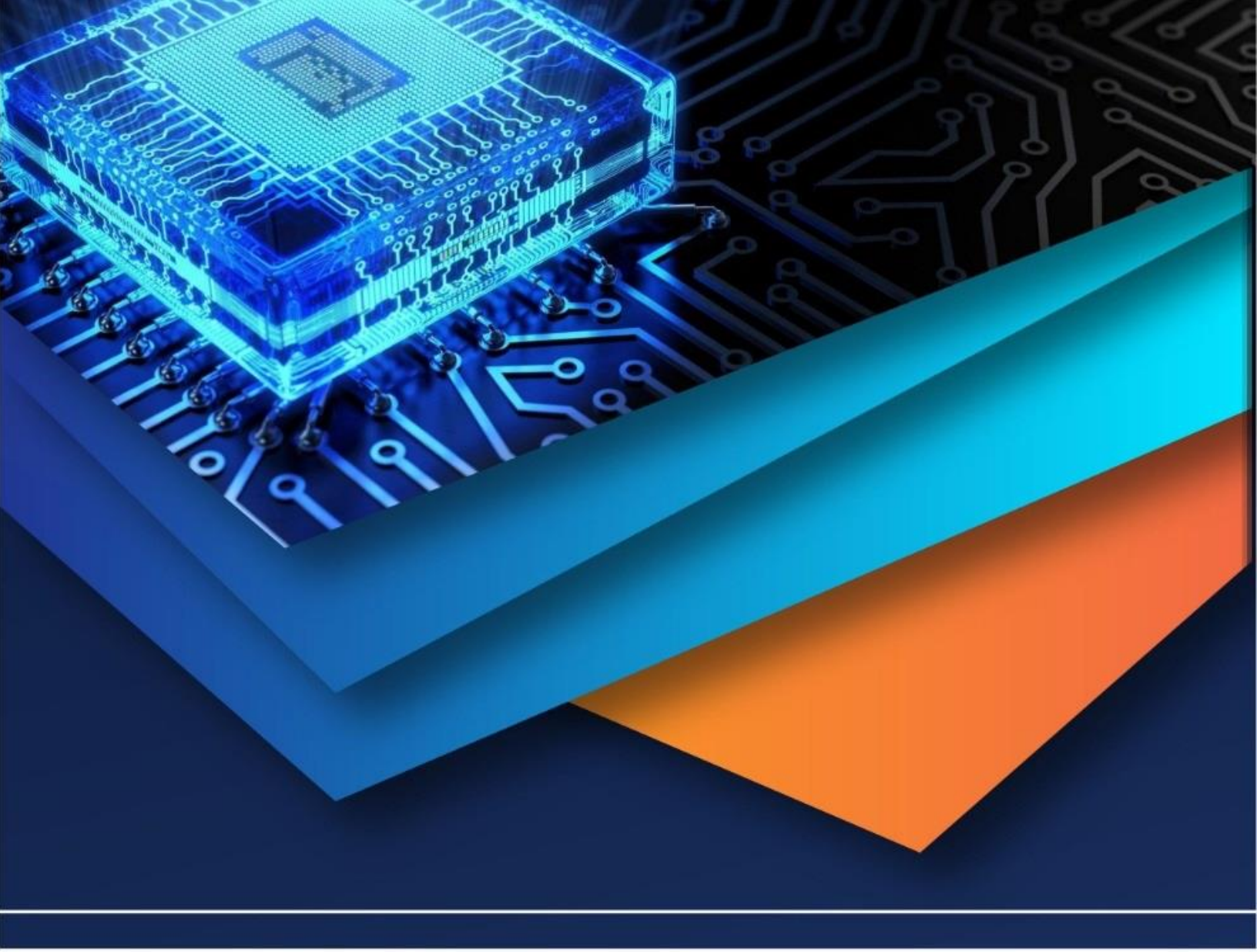
Convolution Neural Network (CNN) is a technique used for prediction of diseases. It has a unique approach in regularization and convenes more complex patterns from existing hierarchical patterns. CNN algorithm has accuracy of 84.4% which is comparatively higher than the previously implemented techniques. The UI designed has graphical representation of data which is layman understandable. The activation functions like ReLu and Sigmoid has improved the performance of neural networks.

VII.ACKNOWLEDGMENT

Primarily, we would like to thank almighty for giving us all the courage and patience to accomplish this work. We express our heartful thanks to our guide Prof. Asnika for all the guidance and support.

REFERENCES

[1] Thiyagarajan C, Dr. K. Anandha Kumar, Dr. A. Bharathi. "A Survey on Diabetes Mellitus Prediction Using Machine Learning Techniques." [Online]. Available: https://pdfs.semanticscholar.org/74c6/6e5295c13cac9ae7283c883f336f2f675618.pdf
[2] Kumari Deepika and Dr. S. Seema. "Predictive Analysis to prevent and control Chronic Disease." [Online]. Available:https://ieeexplore.ieee.org/document/7912028/
[3] Panigrahi Srikanthad and Dharmaiah Deverapalli. "A Critical Study of Classification Algorithms Using Diabetes Diagnosis"[Online]. Available: https://ieeexplore.ieee.org/document/7544843/
[4] Nirmala Devi. M, Appavu alias Balamurugan. S, Swathi U.V. "An amalgam KNN model to predict Diabetes mellitus" [Online]. Available: https://ieeexplore.ieee.org/document/6528591/
[5] Fikirte Girma Woldemichael, Sumitra Menaria. "Prediction of Diabetes using Data Mining Techniques" [Online]. Available: https://www.researchgate.net/publication/329393208_Prediction_of_Diabetes_Using_Data_Mining_Techniques
[6] Roxana Mirshahvalad and Nastaran Asadi Zanjani. "Diabetes Prediction Using Ensembled Perceptron Algorithm" [Online]. Available: https://ieeexplore.ieee.org/document/8319383
[7] Miss Sneha Joshi and Prof. Megha Borse. "Detection and Prediction of Diabetes Mellitus Using Back-Propagation Neural Network" [Online]. Available: https://ieeexplore.ieee.org/document/7938895
[8] T. Jayalakshmi and Dr. A. SanthaKumaran. "Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks". [Online]. Available: https://ieeexplore.ieee.org/document/5452614
[9] Sushruta Mishra, Pamela Chaudhury, Brojo Kishore Mishra and Hrudaya Kumar Tripathy . "An Implentation of Feature Ranking Using Machine Learning Techniques for Diabetes disease prediction" [Online]. Available: https://dl.acm.org/citation.cfm?id=2905100
[10] Rajeeb Dey, Vaibhav Bajpai, Gagan Gandhi and Barnali Dey. "Application of Artificial Neural Network (ANN) technique for Diagnosis of Diabetes Mellitus" [Online]. Available:https://ieeexplore.ieee.org/abstract/document/4798367.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY