JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A MACHINE LEARNING APPROACH FOR DIABETICS PATIENT DETECTION USING ENHANCED ANN

[1]Anupriya Hamilton , [2]Prof. Sugan Patel,

[1, 2]Department of Computer Science & Engineering,

[1, 2]Lakshmi Narain College of Technology Excellence Bhopal, Madhya Pradesh

Abstract : In this research work present a hybrid model that is combination of linear regression model and cascaded feed forward network with back propagation training. The detection of the diabetic type 2 patient shows better accuracy as compare to other previous methods. In the different previous method analysis got the drawbacks of previous method that is over come in the proposed method that is shown in the result discussion. The outcome proposed method shown that can be detected with an awfully high accuracy, up to 96.54 % in the different diabetic patient datasets. The proposed method also shows lower complexity and better timing due to regression model soft threshold values. In the proposed hybrid model is the combination of CFFNN and Linear Regression method. Cascade-forward networks are similar to feed-forward networks, but include a connection from the input and every previous layer to following layers. As with feed-forward networks, a two-or more layer cascade network can learn any finite input-output relationship arbitrarily well given enough hidden neurons. For greater accuracy on low-dimensional through medium-dimensional data sets, fit a linear regression model. There are major parts of proposed work that is describe in this thesis work. The accuracy of the proposed methods comparison of different methods such as Discrim, MLP, Logdisc, SMART, Bayesnet, Naive bay, Random Forent, J48, SGD, SMO, Backiprop, RBF, LMT nad proposed hybrid model shows better outcome as compare to other previous methods, that is shown in simulation and result.

IndexTerms **-Diabetic Type 2, Cascaded Feed Forward Network, Linear Regression, Machine Learning Processes, Random Forent And Accuracy etc**

I. INTRODUCTION

Diabetes Mellitus ("diabetes" for short) may be a serious illness that happens once your body has issue properly regulation the number of dissolved sugar (glucose) in your blood stream. it's unrelated to a equally named disorder "Diabetes Insipidus" that involves kidney-related fluid retention issues. so as to grasp polygenic disease, it's necessary to initial perceive the role aldohexose plays with reference to the body, and what will happen once regulation of aldohexose fails and glucose levels become perilously low or high. The tissues and cells that structure the anatomy reside things, and need food to remain alive. The food cells eat could be a kind of sugar referred to as aldohexose. mounted in situ as they're, the body's cells ar utterly captivated with the blood stream during which they're bathed to bring aldohexose to them. while not access to adequate aldohexose, the body's cells don't have anything to fuel themselves with and shortly die. citizenry eat food, not aldohexose.

Human foods get regenerate into aldohexose as a vicinity of the conventional digestion method. Once reborn, aldohexose enters the blood stream, inflicting the extent of dissolved aldohexose within the blood to rise. The blood stream then carries the dissolved aldohexose to the assortedtissues and cells of the body. tho' aldohexose could also be out there within the blood, close cells aren't ready to access that aldohexose while not the help of a chemical internal secretion referred to as endocrine. internal secretion acts as a key to open the cells, permitting them to receive and utilize on the market aldohexose. Cells absorb aldohexose from the blood within the presence of internal secretion, and blood glucose levels drop as sugar leaves the blood and enters the cells. internal secretion will be thought of as a bridge for aldohexose between the blood stream and cells. it's vital to grasp once levels of endocrine increase, levels of sugar within the blood decrease (because the sugar goes into the cells to be used for energy).The body is meant to control and buffer the number of aldohexose dissolved within the blood to take care of a gradual provide to satisfy cell wants. The exocrine gland, one in all your body's several organs, produces, stores and releases endocrine into the blood stream to bring aldohexose levels backpedal. The concentration of aldohexose accessible within the blood stream at any given moment depends on the number and sort of foods that individuals eat. Refined carbohydrates, candy and sweets ar straightforward to interrupt down into aldohexose. Correspondingly, glucose levels rise speedily when such foods are consumed. In distinction, blood sugars rises step by step and slowly when consumption a lot of advanced, unrefined carbohydrates (oatmeal, apples, baked potatoes, etc.) that need a lot of biological process steps crop up before aldohexose will be yielded. baby-faced with speedily rising glucose concentrations, the body should react quickly by emotional giant amounts of endocrine all directly or risk a dangerous condition referred to as Hyperglycemias (high blood sugar) which is able to be represented below. The inflow of

endocrine permits cells to utilize aldohexose, and aldohexose concentrations drop. whereas aldohexose levels will rise and fall quickly, endocrine levels modification rather more slowly.

1.1 Machine Learning

ANNs area unit composed of multiple nodes, that imitate biological neurons of human brain. The neurons area unit connected by links and that they act with one another. The nodes will take input file and perform easy operations on the info. The results of these operations is passed to different neurons. The output at every node is termed its activation or node worth. Each link is related to weight. ANNs area unit capable of learning, that takes place by sterilization weight values. the subsequent illustration shows an easy ANN−

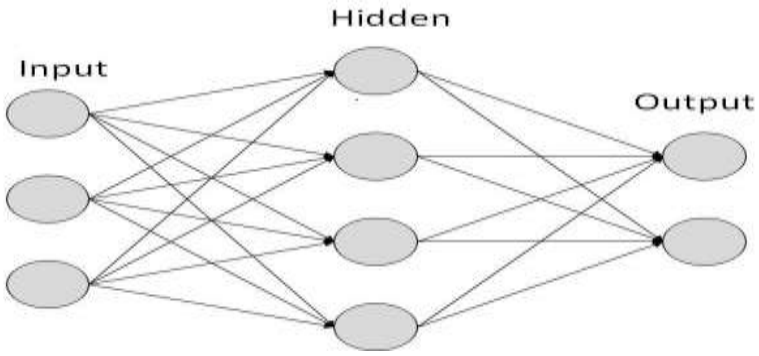


Fig 1. Types of Artificial Neural Networks

Machine Learning in ANNs

ANNs are capable of learning and that they have to be compelled to be trained. There ar many learning ways −

Supervised Learning − It involves a tutor that's scholar than the ANN itself. for instance, the teacher feeds some example knowledge concerning that the teacher already is aware of the answers. For example, pattern recognizing. The ANN comes up with guesses whereas recognizing. Then the teacher provides the ANN with the answers. The network then compares it guesses with the teacher's "correct" answers and makes changes consistent with errors.

- **Unsupervised Learning** − it's needed once there's no example knowledge set with known answers. for instance, finding out a hidden pattern. during this case, cluster i.e. dividing a collection of components into teams in step with some unknown pattern is disbursed supported the prevailing} knowledge sets present.
- **Reinforcement Learning** − This strategy engineered on observation. The ANN makes a call by perceptive its atmosphere. If the observation is negative, the network adjusts its weights to be ready to build a unique needed call consequent time.

1.2 Back Propagation Algorithm

It is the coaching or learning formula. It learns by example. If you meet up with the formula the instance of what you wish the network to try and do, it changes the network's weights in order that it will turn out desired output for a selected input on finishing the coaching. Back Propagation networks are ideal for easy Pattern Recognition and Mapping Tasks.

II. NEURAL NETWORK AND REGRESSION

2.1 Neural Network

Neural network can be a study model, that imitates the animal's neural network behaviors. This model depends on the quality of the system to comprehend the aim of method information by adjusting the affiliation between the inner node. to keep with the connections' vogue, the neural network model are divided into forward network and feedback network. throughout this paper, we tend to tend to used the Neural Pattern Recognition app in MATLAB, that would be a two-layer-feed-back network with sigmoid hidden and soft ax output neurons. The neural network structural is shown in Figure 4. .

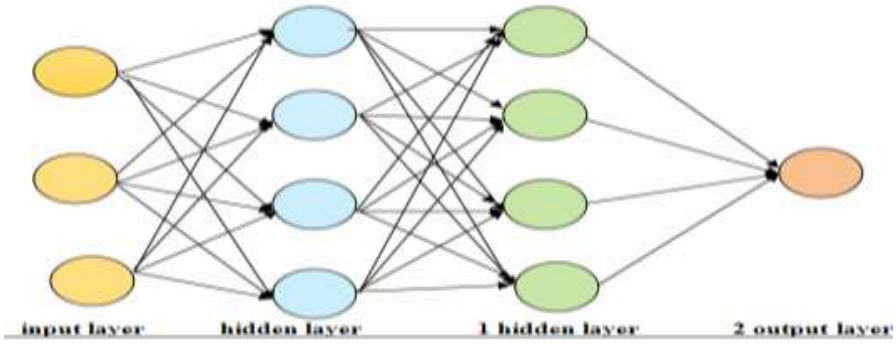


Fig. 2 Neural network

2.2 Types Of Neural Networks

Neural networks are process models that work just like the functioning of a person's system. There are many types of artificial neural networks. These variety of networks are enforced supported the mathematical operations and a group of parameters needed to see the output. Let's look into a number of the neural networks:

1. Feed Forward Neural Network – Artificial Neuron

This neural network is one amongst the only sorts of ANN, wherever the info or the input travels in one direction. The info passes through the input nodes and exit on the output nodes. This neural network could or might not have the hidden layers. In straightforward words, it's a front propagated wave and no back propagation by employing a classifying activation operate sometimes. Below may be a Single layer feed forward network.

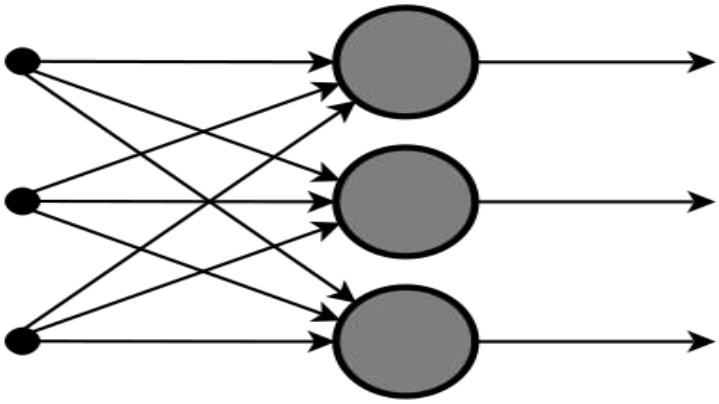


Fig. 3 Single layer feed forward network

2. Radial Basis Function Neural Network:

Radial basic functions contemplate the gap of a degree with relevance the middle. RBF functions have 2 layers, 1st wherever the options are combined with the Radial Basis operate within the inner layer and so the output of those options are taken into thought whereas computing constant output within the next time-step that is largely a memory. Here, the space live utilized in Euclidian, different distance measures can even be used. The model depends on the utmost reach or the radius of the circle in classifying the points into totally different classes. If the purpose is in or round the radius, the probability of the new purpose begin category field into that class is high. There will be a transition whereas dynamic from one region to a different and this will be controlled by the beta perform.

Fig. 4 Power Restoration Systems

III. PROPOSED METHOD AND ALGORITHM

IV.

3.1 Diabetes Patient Prediction Model

In this section, discuss the diabetes patient prediction architecture and define the detection process model. In general diabetes patient prediction first train the system with the help of data set, in the data set there are different body such as Age, BMI, Impaired glucose tolerance (IGT), Family history of DM, Lower high-density lipoprotein cholesterol or (HTG), Hypertension or cardiovascular disease. All these parameters are used to train proposed system. evaluated.

Data Set

Data set is back bone of diabetes patient analysis based projects. There are different type of data set available, in the below table 1 shows the one of the data set that is UCI library, stand stanford university and kaggle.

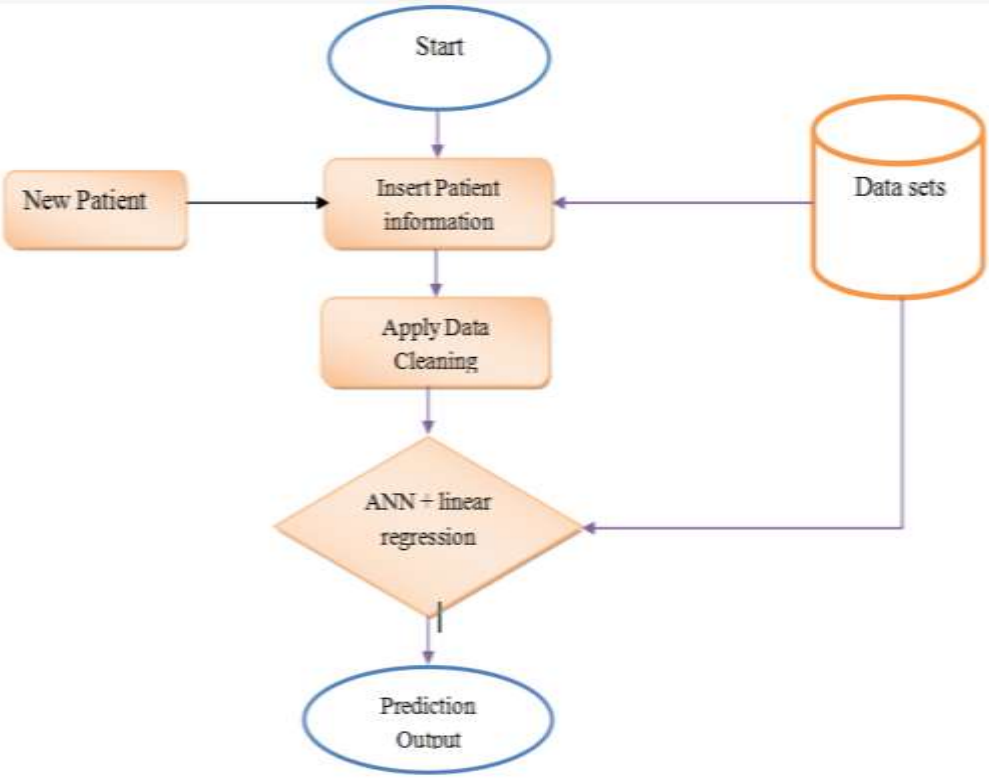


Fig. 5. Shows Type 2 Diabetes Patient Prediction Model

Table 1 Shows the data Set

	preg	plas	pres	skin	insu	mass	pedi	age	class
6	147	73	34	0	34.5	0.626	51		tested_positive
1	86	67	30	0	25.7	0.352	30		tested_negative
8	184	63	0	0	24.4	0.673	33		tested_positive
1	90	67	24	93	27.2	0.166	22		tested_negative
0	138	41	34	169	42.2	2.289	34		tested_positive
5	115	76	0	0	26.7	0.202	31		tested_negative
3	79	51	33	89	32	0.247	27		tested_positive
10	116	0	0	0	34.2	0.133	28		tested_negative

3.2 Cascade-Forward Neural Network

In perception connection that is formed between input and output is a form of direct relation while in FFNN connection formed between input and output is indirect relationship. The connection is nonlinear in shape through an activation function in the hidden layer. If the connection form on perception and multilayer network is combined, then the network with direct connection between the input layer and the output layer is formed, besides the connection indirectly. The network formed from this connection pattern is called Cascade Forward Neural Network (CFNN). The equations are formed from the CFNN model can be written as follows:

$$y = \sum_{i=1}^{n} f^{i}\omega_{t}^{i}\,\chi_i + f^{0}\left(\sum_{j=1}^{n}\omega_j^0\, f_j^h\left(\sum_{i=1}^{n}\omega_{ji}^h\chi_i\right)\right) \quad 4.5$$

Where f^i is the activation function from the input layer to the output layer and ω_{ji}^h is weight from the input layer to the output layer. If a bias is added to the input layer and the activation function of each neuron in the hidden layer is then equation (4.6) becomes

$$y = \sum_{i=1}^{n} f^i\,\omega_i^i\chi_i + f^0\left(\omega^b + \sum_{j=1}^{k}\omega_j^0\, f^h\left(\omega_j^b + \sum_{i=1}^{n}\omega_{ji}^h\,\chi_i\right)\right) \quad 4.6$$

In this research, the CFNN model is applied in time series data. Thereby, the neurons in the input layer are the lags of time series data $Xt-1, Xt-2, \ldots, Xt-p$, whereas the output is the current data Xt . The architecture of CFNN model in predicting time series is shown at Fig.6.

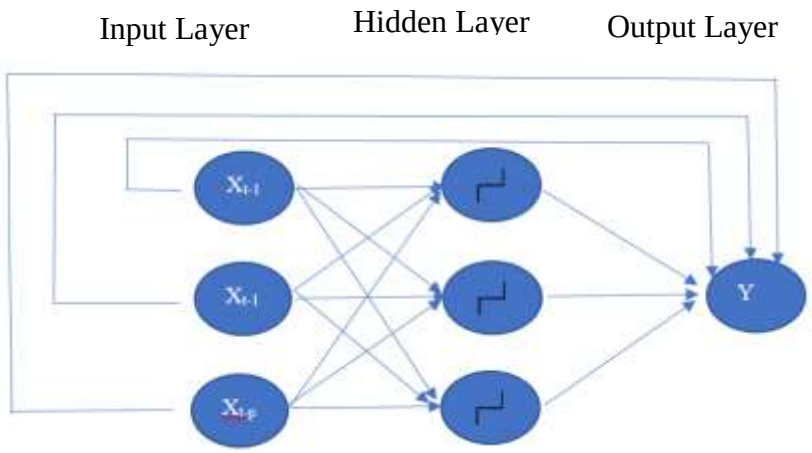


Fig. 6 Shows Neural Cascaded

3.3 Proposed Algorithm

In this proposed hybrid model is the combination of linear regression, Levenberg- back propagation method is use for the training data set, use linear regression to enhance and refine the cleaned input data, use cascaded feed forward neural network machine learning technique for prediction of Type 2 diabetes mellitus prediction. The overall proposed method is divided into three different sections. In the first section, apply data base cleaning of the file. In the next part apply linear regression with different input parameters, for this first complete step 1.

Section -1 Apply Data Cleaning

Section -2 Apply Linear Regression method

Section -3 Apply ANN based feed forwarded network.

Pseudo Codes for file Proposed Algorithm

1. Data Base selection
2. Apply Pre processing task
3. file_ID = fopen(filename,'r');
4. Select = {'AGE','SEX','BMI','BP','S1','S2','S3','S4','S5','S6'}
5. Select file
6. For fileID = 1: Number of data attributes
7. [beta, FitInfo] = Regression
8. Plot Regression plot
12. End
13. Apply Levenberg-Marquardt backpropagation Training = beta, Fit Info
14. net = cascadeforwardnet(beta, Fit Info); // Neural Network
15. net = train(net,x,t);
16. Set parameters trainParam.epochs = 1000; trainParam.goal = 0.01;trainParam.mu = 0.01;
16. perf = perform(net,y,t)
18. end .

Algorithm of Linear Regression

regression(x,y,n,al,a0,syx,r2)

sumX = 0, sumxy =0 st = 0,

sumY = 0, sumX2 = 0, sr = 0

For i = 1:n

sumX = sumX+x_i;

sumY = sumY+y_i;

sumXY = sumXY+$x_i * y_i$,

sumX2 = sumX2+$x_i * y_i$

end

$xm = sum\ x/n$

$ym = sum\ y/n$

$a1 = \frac{n*sum\ xy - sum\ x*sum\ y}{n*sum\ x2 - sum\ x*sum\ x}$

$a0 = ym - a1 * xm$

$for\ i = 1{:}n$

$st = st + (y_i - y_m)^2$

$st = sr + (y_i - a1 * xi - a0)^2$

end

$syx = (sr/(n-2))^{0.5}$

$r2 = (st - sr)/st$

end

V. SIMULATION AND RESULT

4.1 Result Parameters and Simulation Tool

The result of proposed method for development of middle ware using machine learning technique for diabetic type 2 shown in this section, simulation of our proposed method and result calculation. We have done proposed work with the help the MATLAB R 2015a (8.1.0.602) software and simulate our whole proposed methodology in data analysis. Basic configuration of our system is: Processor: Intel (R) Quad Core (VM) i3 – 3110 Central Processing unit @, 2.40 GHz with 4GB RAM: System type: 64-bit Operating System.

5.2.1 Accuracy

The accuracy is the ratio of addition of number of correct production (TP+TN) and total number of production ($\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}$) .

$$Acc = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

where: TP = True positive; FP = False positive; TN = True negative; FN = False negative

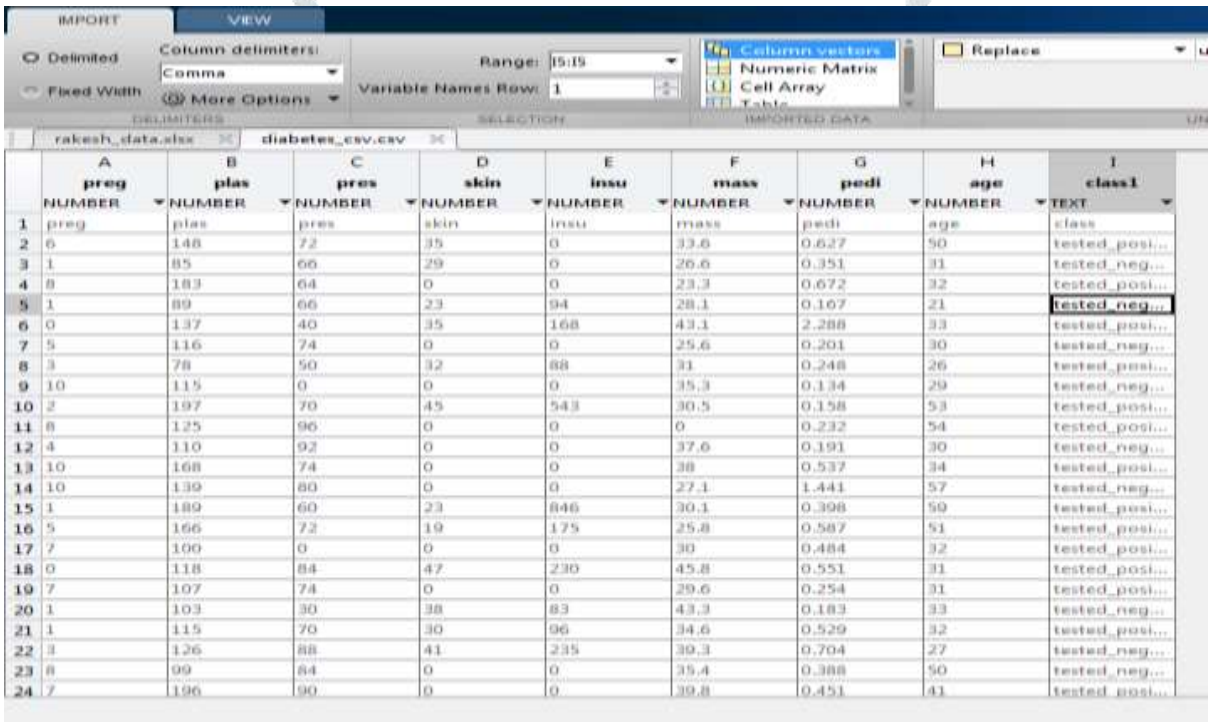
Data Sets

There are different dietetic patient data set are taken for performing proposed work. First discuss about the **Pima Indian Diabetes Dataset**. This dataset is originally from the National Institute of hereditary disease and organic process and internal organ Diseases. the target of the dataset is to diagnostically predict whether or not or not a patient has polygenic disease, supported bound diagnostic measurements enclosed within the dataset.Many constraints were placed on the choice of those instances from a bigger information. particularly, all patients here are females a minimum of twenty one year's recent of Pima Indian heritage.

Table 2 Pima Indian Diabetes Dataset.

Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
11	145	69	32	30	30.6	0.624	47	1
3	82	63	27	20	23.6	0.354	34	0
10	180	61	20	15	20.3	0.675	35	1
3	86	63	20	91	25.1	0.164	24	0
5	134	43	38	165	40.1	2.285	36	1
7	113	71	21	110	22.6	0.204	33	0
9	75	53	35	91	34	0.245	29	1
6	118	45	20	60	32.3	0.131	26	0
8	194	73	42	540	33.5	0.155	56	1
10	122	99	40	160	30.5	0.230	57	1
6	113	95	43	543	40.6	0.189	33	0

10	165	77	40	540	35	0.534	37	1
9	137	83	45	232	30.1	1.444	60	0
6	186	63	26	843	33.1	0.395	56	1
7	169	75	22	178	28.8	0.584	54	1
9	103	78	25	181	33	0.487	35	1
5	115	87	50	233	45.5	0.554	34	1
7	110	77	53	236	26.6	0.251	34	1
3	106	33	35	80	40.3	0.180	30	0
9	112	73	33	93	31.6	0.526	35	1
10	123	85	44	231	36.3	0.701	30	0
8	102	87	47	234	32.4	0.382	53	0
5	193	93	50	237	36.8	0.454	44	1
9	116	83	38	240	30	0.260	32	1
7	140	91	30	143	33.6	0.251	54	1
10	122	73	30	112	31.4	0.202	44	1
7	143	73	33	115	39.0	0.254	46	1
6	100	69	18	143	23.5	0.483	25	0



	A preg NUMBER	B plas NUMBER	C pres NUMBER	D skin NUMBER	E insu NUMBER	F mass NUMBER	G pedi NUMBER	H age NUMBER	I class1 TEXT
1	preg	plas	pres	skin	insu	mass	pedi	age	class
2	6	148	72	35	0	33.6	0.627	50	tested_posi...
3	1	85	66	29	0	26.6	0.351	31	tested_neg...
4	8	183	64	0	0	23.3	0.672	32	tested_posi...
5	1	89	66	23	94	28.1	0.167	21	tested_neg...
6	0	137	40	35	168	43.1	2.288	33	tested_posi...
7	5	116	74	0	0	25.6	0.201	30	tested_neg...
8	3	78	50	32	88	31	0.248	26	tested_posi...
9	10	115	0	0	0	35.3	0.134	29	tested_neg...
10	2	197	70	45	543	30.5	0.158	53	tested_posi...
11	8	125	96	0	0	0	0.232	54	tested_posi...
12	4	110	92	0	0	37.6	0.191	30	tested_neg...
13	10	168	74	0	0	38	0.537	34	tested_posi...
14	10	139	80	0	0	27.1	1.441	57	tested_neg...
15	1	189	60	23	846	30.1	0.398	59	tested_posi...
16	5	166	72	19	175	25.8	0.587	51	tested_posi...
17	7	100	0	0	0	30	0.484	32	tested_posi...
18	0	118	84	47	230	45.8	0.551	31	tested_posi...
19	7	107	74	0	0	29.6	0.254	31	tested_posi...
20	1	103	30	38	83	43.3	0.183	33	tested_neg...
21	1	115	70	30	96	34.6	0.529	32	tested_posi...
22	3	126	88	41	235	39.3	0.704	27	tested_neg...
23	8	99	84	0	0	35.4	0.388	50	tested_neg...
24	7	196	90	0	0	39.8	0.451	41	tested_posi...

Fig. 7 Shows the Diabetes Dataset in MATLAB

4.2 Result Discussion

1 Regression output

In this section the result or outcomes of proposed method. after data cleaning and pre processed of the above data set apply linear regression on this data. In the below figure 8 shows the regression output, in this plot X axis shows the value of Lembda that is log decimal value , in the Y axis shows the expected normal value of input data.

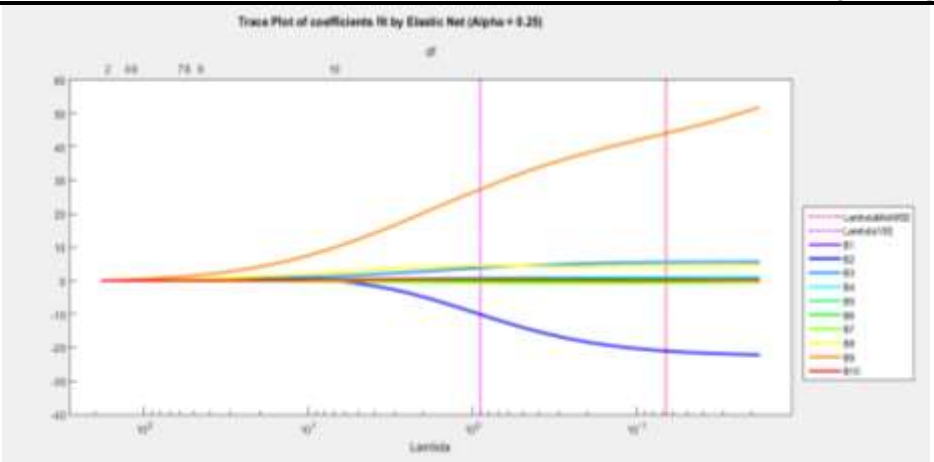


Fig. 8 Regression output of tested data 1

4.3 Cascaded Feed Forward Neural Network

In the below figure 9. shows the cascaded feed forward neural network,

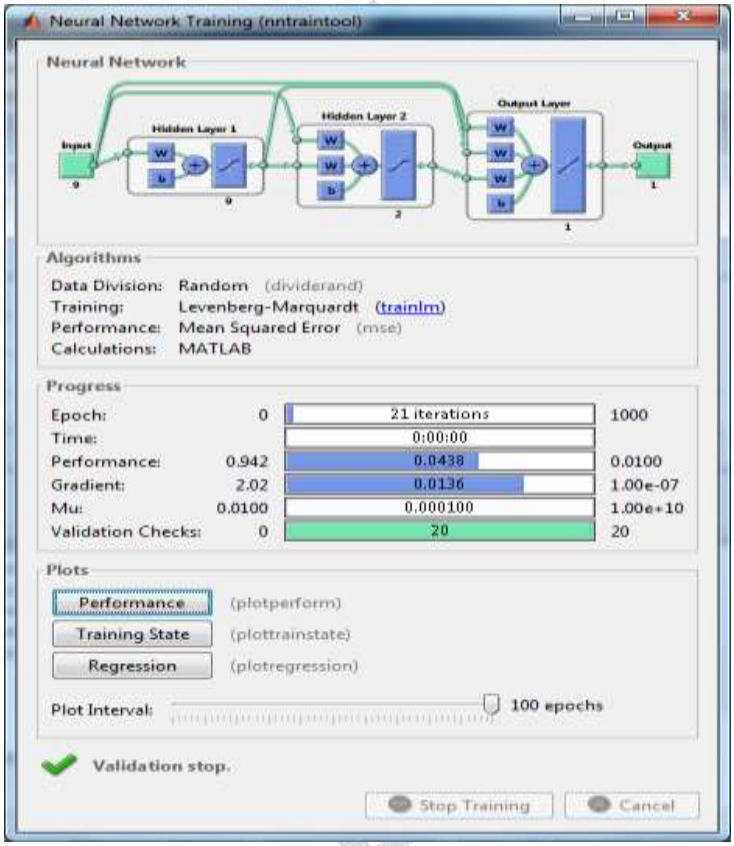


Fig.9 Neural network output

4.4 Result Comparison

In the above section discuss the result compression of proposed method. Now compare the resultant output with different machine learning methods that is shown in below. In the below result table 3. shows the accuracy of classification of different methods using machine learning approach. The resultant accuracy of proposed work is near around 79% (without regression).

Table 3. The values of accuracy of classification made on pima Indian diabetes dataset

Method	Accuracy(%)
Proposed Method	79.6
Discrim	77.5
MLP	73.8
Logdisc	78.2
SMART	76.8
Bayesnet	74.7
NaivebAY	74.9

RandomForent	76
J48	76.7
SGD	76.6
SMO	77
Backiprop	75.2
RBF	75.7
LMT	76.6

In the below figure 10 shows the accuracy of the comparison of different methods such as Discrim, MLP, Logdisc, SMART, Bayesnet, NaivebAY, Random Forent, J48, SGD, SMO, Backiprop, RBF, LMT , in the graphical form in figure 10. In all the methods apply direct classification techniques on different machine learning platform such as Weka, R , R studio and Python. Our proposed Cascaded method shows better accuracy near around 80%.

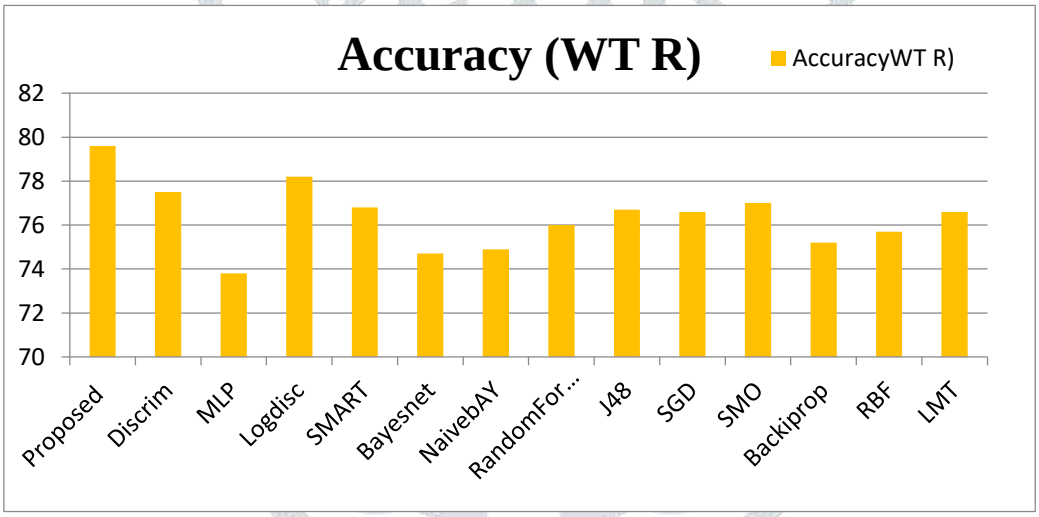


Fig. 10 Shows the accuracy comparison with different methods

V.CONCLUSION & FUTURE ENHANCEMENT

In the research work present a hybrid model that is combination of linear regression model and cascaded feed forward network with back propagation training. The detection of the diabetic type 2 patient shows better accuracy as compare to other previous methods. The outcome proposed method shown that can be detected with an awfully high accuracy, up to 96.54 % in the different diabetic patient datasets. The proposed method also shows lower complexity and better timing due to regression model soft threshold values. The result comparison of proposed method shown in the different tabular form in the table 2 and 3.

Future Work

In future try to improve the accuracy of the diabetic patient detection and also focus on different techniques of detection a well as other part detection not only in the type 2 detection as well as type -1.

Try to implement presented method on single chip with the filed preamble gate array technology. FPGA implementation of presented method is not a easy task. Focus on new vegetable indices that is useful for different type of other weed detection in plants In future world is totally digital, whole agriculture will be digitalized.

In future also try to enhance the accuracy with the help of different regression techniques such non linear regression techniques.

REFERENCES

1. Wu, Han, et al. "Type 2 diabetes mellitus prediction model based on data mining." Informatics in Medicine Unlocked 10 (2018).
2. Kavakiotis, Ioannis, Olga Tsave, AthanasiosSalifoglou, NicosMaglaveras, IoannisVlahavas, and IoannaChouvarda. "Machine learning and data mining methods in diabetes research." Computational and structural biotechnology journal (2017).

3. Songthung, Phattharat, and KunwadeeSripanidkulchai. "Improving type 2 diabetes mellitus risk prediction using classification." In Computer Science and Software Engineering (JCSSE), 2016 13th International Joint Conference on, pp. 1-6. IEEE, 2016.
4. Panwar, Madhuri, Amit Acharyya, Rishad A. Shafik, and Dwaipayan Biswas. "K-nearest neighbor based methodology for accurate diagnosis of diabetes mellitus." In Embedded Computing and System Design (ISED), 2016 Sixth International Symposium on, pp. 132-136. IEEE, 2016.
5. Saxena, Kanak, and Richa Sharma. "Diabetes mellitus prediction system evaluation using c4. 5 rules and partial tree." In Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on, pp. 1-6. IEEE, 2015.
6. ZahirniaKiarash, Mehdi Teimouri, RohallahRahmani, and Amin Salaq. "Diagnosis of type 2 diabetes using cost-sensitive learning." In Computer and Knowledge Engineering (ICCKE), 2015 5th International Conference on, pp. 158-163. IEEE, 2015.
7. Lee, Bum Ju, Boncho Ku, Jiho Nam, Duong Duc Pham, and Jong Yeol Kim. "Prediction of fasting plasma glucose status using anthropometric measures for diagnosing type 2 diabetes." IEEE journal of biomedical and health informatics 18, no. 2 (2014).
8. Choi, SooBeom, Won Jae Kim, Tae KeunYoo, JeeSoo Park, Jai Won Chung, Yong-ho Lee, EunSeok Kang, and Deok Won Kim. "Screening for prediabetes using machine learning models." Computational and mathematical methods in medicine 2014 (2014).
9. Velu, C. M., and K. R. Kashwan. "Visual data mining techniques for classification of diabetic patients." In Advance Computing Conference (IACC), 2013 IEEE 3rd International, pp. 1070-1075. IEEE, 2013.
10. Dalakleidi, Kalliopi V., KonstantiaZarkogianni, Vassilios G. Karamanos, Anastasia C. Thanopoulou, and Konstantina S. Nikita. "A hybrid genetic algorithm for the selection of the critical features for risk prediction of cardiovascular complications in Type 2 Diabetes patients." In Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on, pp. 1-4. IEEE, 2013.
11. OlokobaAbdulfatai B., Olusegun A. Obateru, and Lateefat B. Olokoba. "Type 2 diabetes mellitus: a review of current trends." Oman medical journal 27, no. 4 (2012).
12. Marinov, Miroslav, Abu Saleh Mohammad Mosa, IllhoiYoo, and Suzanne Austin Boren. "Data-mining technologies for diabetes: a systematic review." Journal of diabetes science and technology 5, no. 6 (2011).
13. Sacks, David B., Mark Arnold, George L. Bakris, David E. Bruns, Andrea Rita Horvath, M. Sue Kirkman, AkeLernmark, Boyd E. Metzger, and David M. Nathan. "Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus." Clinical chemistry 57, (2011).
14. Zhang, Ping, Xinzhi Zhang, Jonathan Brown, DorteVistisen, Richard Sicree, Jonathan Shaw, and Gregory Nichols. "Global healthcare expenditure on diabetes for 2010 and 2030." Diabetes research and clinical practice 87, no. 3 (2010): .
15. B. M. Patil, R. C. Joshi and DurgaToshniwal, “Association rule for classification of type -2 diabetic patients”, Second International Conference on Machine Learning and Computing, 2010.
16. Rodbard, Helena W., Andrew J. Green, Kathleen M. Fox, and Susan Grandy. "Trends in method of diagnosis of type 2 diabetes mellitus: results from SHIELD." International journal of endocrinology 2009 (2009).
17. RajeebDey , VaibhavBajpai, Gagan Gandhi and BarnaliDey, “Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus”, IEEE Region 10 and the Third international Conference on Industrial and Information Systems,Dec 2008.
18. BalakrishnanSarojini, RamarajNarayanaswamy, Nickolas Savarimuthu, and Rita Samikannu. "SVM ranking with backward search for feature selection in type II diabetes databases." In Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on, pp. 2628-2633. IEEE, 2008.
19. Liang Lily, VinayMandal, Yi Lu, and Deepak Kumar. "Multi-dimensional cluster misclassification test for pathway differential analysis of diabetes." In Computer and Computational Sciences, 2007. IMSCCS 2007. Second International Multi-Symposiums on, pp. 84-91. IEEE, 2007.
20. Aekplakorn, Wichai, PongamornBunnag, Mark Woodward, PiyamitrSritara, SayanCheepudomwit, SukitYamwong, Tada Yipintsoi, and RajataRajatanavin. "A risk score for predicting incident diabetes in the Thai population." Diabetes care 29, no. 8 (2006).
21. Lindström, Jaana, and JaakkoTuomilehto. "The diabetes risk score." Diabetes care 26, no. 3 (2003): 725-731.
22. Tuomilehto, Jaakko, JaanaLindström, Johan G. Eriksson, Timo T. Valle, Helena Hämäläinen, PirjoIlanne-Parikka, SirkkaKeinänen-Kiukaanniemi et al. "Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance." New England Journal of Medicine344, no. 18 (2001): 1343-1350.
23. Pan, Xiao-Ren, Guang-Wei Li, Ying-Hua Hu, Ji-Xing Wang, Wen-Ying Yang, Zuo-Xin An, Ze-Xi Hu et al. "Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and Diabetes Study." Diabetes care 20, no. 4 (1997): 537-544.
24. G.V. Kass, “An Exploratory Technique for Investigating Large Quantities of Categorical Data”, journal, Wiley for the Royal Statistical Society, 1980.