Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

A Comparative Study of Legendre Neural Network and Chebyshev Functional Link Artificial Neural Network for Diabetes Data Classification

Swati Das
B.tech / M.Tech in Computer Science Engineering
Faculty ,Rourkela Institute of Management Studies,Odisha, India,

***Abstract*— Data mining plays an important role in data classification technology. As diabetes is an ongoing research project in medical science, analyzing diabetes data has become increasingly important in the near future. Better and faster is a more efficient data analysis method to get results more accurate. Our proposed work is based on the classification of the most effective diabetes data in current medical science research .We worked on two advanced neural networks, namely the Legendre Neural Network (LeNN) and Chebyshev Functional Link based ANN (CHFLANN) and compared their performance in terms of accuracy and F-Measure for a diabetic sample collected from the UCI database. By performing simulations in the MATLAB environment and analyzing the results of the Legendre Neural Network-based architecture provides better performance compared to the Chebyshev-based method.**

Keywords— ***Diabetes data, LeNN, CHFLANN, FLANN, Artificial neural network (ANN)***

I. INTRODUCTION

The current area of interest and research work in the health care sector is the prevention of various diabetes-related diseases. Numerous data mining methods have been proposed and performance analysis has been done for the identification of major causes of diabetes when considering several data sets. Data mining techniques are applied to existing diabetic records and for decades of analysis. Data mining can be referred to as simultaneously extracting logical data and analyzing and summarizing useful information that can be used to predict future data or experiments. Development of data mining techniques using various predictive algorithms predicts and calculates the error data of sugar data that can also be used in patient safety research. Our aim is to create different testing environment using strategies like Legendre neural network and Chebyshev functional link artificial neural network for predicting the class value for the data set. In [1] a method of data mining considering an analytical problem of health care system in the New Orleans Area with 30,386 diabetic patients was performed with respective results. In [2] simulated treatment data can predict errors of omission in clinical patient data and developed the potential for wide use in identifying decision strategies leading encounter-specific treatments errors in chronic disease care. In [3] various wireless channel equalization techniques for communication system has been analyzed and compared such as LeNN, MLP and FLANN considering a set of database and showed LeNN gives better performance. In [4] an investigation and forecasting based approach has been done for the price fluctuation by an improved LeNN algorithm. In the predictive modeling, the investor decides their investing positions by analyzing the historical data on the stock market. In [5] a Functional Link Artificial Neural Network (FLANN) based approach has been performed for task for classification and an extensive simulation study is performed to demonstrate the effectiveness of the classifier. In [6] an approach has been established for the comparison analysis on Decision Tree, Multi-Layer Perceptron (MLP) and Chebyshev functional link artificial neural network (CFLANN) in terms of their classification accuracy and elapsed time for credit card fraud detection.

With reference to the above proposed works and considering a new problem statement for diabetes data analysis, we have decided to compare the results for LeNN with respect to CHFLANN.

In the Performance analysis we will evaluate the specificity, sensitivity, recall, precision accuracy and f-measure for the PIMA Indian Diabetes dataset. Meanwhile, we will calculate the mean square error curve for both the processes .The details for the proposed technique has been given below along with the appropriate mathematical equations.

II. DATA MINING TECHNIQUES USED IN THE STUDY

In recent days a number of applications of data mining techniques have been found in the diabetes data. Legendre Neural Networks (LeNN) and Chebyshev Functional Link Artificial Neural network (CHFLANN) are the two techniques commonly used in this field. The working principle of these techniques has been described below.

A. LEGENDRE NEURAL NETWORK

The Legendre Neural Network (LeNN) structure is similar to FLANN (Functional neural network link). In FLANN, trigonometric functions are used in expansion functions and LeNN uses Legendre orthogonal function.[8] The Legendre polynomial involves less computation compared to that of trigonometric functions. Therefore, LeNN offers faster training compared to FLANN.

The Legendre polynomials are given by Ln(X), where n is the order whereas $-1 < x < 1$ will be the argument of the polynomial.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

The zero and the first order Legendre polynomial are, respectively given by

$L_0(x) = 1$ and $L_1(x) = x$ (1)

The higher order polynomials are given by

$L_2(x) = 1/2(3x^2-1)$
$L_3(x) = 1/2(5x^3 -3x)$ (2)
$L_4(x) = 1/8(35x^4-30x^2+3)$

The recursive formula to generate higher order Legendre polynomials is expressed as

$L_{n+1}(x) = 1/n+1\ [(2n + 1)\ xL_n(x) - nL_{n-1}]$ (3)

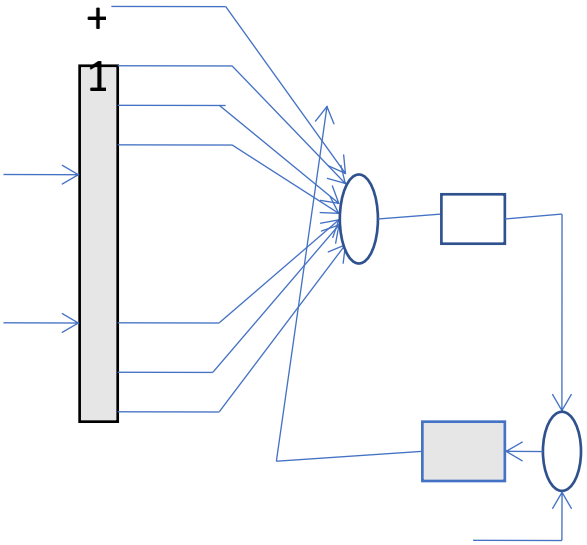

Fig I. Architecture of Legendre neural network

The above figure shows the scheme for the LeNN showing the functional Expansions as well as the learning components. The eight attribute input pattern is enhanced to 17 numbers of LeNN functional expansion. Then we calculate the Means Square Error and further other parameters like class value, accuracy and f-measure calculated for the result analysis.

After getting final expansion for LeNN, the average of the components of the enhanced input pattern is obtained using the below formula

$$\text{Weighted Sum} = \sum wjLXj \quad (4)$$

The error obtained by comparing the output with desired output is used to update the weights of the network structure by a weight updating algorithm. The Back Propagation algorithm, which is used to train the network, becomes very simple because of absence of any hidden layer its hidden layer. In every iteration the gradient of the cost function w.r.t the weights is determined and the weights are incremented by a fraction of the negative gradient

$E\ k = 1/2[d\ k - y\ k]^2 = 1/2\ e\ k^2$ (5)

Where,
Ek is the error rate
dk is the desired value
yk is output of output of the network
ek is the error at kth step

$w\ j, k+1 = w\ j, k + \alpha\ ek\ (1 - y\ k)\ 2\ L\ (X)$ (6)

Where,
w j, k+1 is the updated coefficient
w j, k is the old co efficient
Alpha is the learning rate
L(X) is the expanded unit

ALGORITHM

STEP1- Normalize the input features of the diabetes data set
STEP2- Normalization of dataset is stored in x.
STEP3- Expansion of normalized data using Legendre polynomial of order 3 and output is stored in X.
STEP4- Divide the input matrix into training and testing.
STEP5- Random initialization of weights i.e. 0.1 to 1.
STEP6- Take first sample values from training matrix and multiplies the weights with them. Display the sum of the multiplied values and store it and repeat the same for all the samples. Again, multiply weight (w) with testing matrix set and stored it in testing output matrix.
STEP7- Find class value of testing output matrix.

If class value> 0.5
Class value is 1
Else
Class Value is 0

B. CHEBYSHEV FUNCTIONAL LINK ARTIFICIAL NEURAL NETWORK

CHFLANN (Chebyshev Functional Link Neural Network) is a one layer neural network in which the previous input pattern extends to a higher magnitude through a set of Chebyshev orthogonal functions. Chebyshev polynomials are set by polynomials denoted by CHp (X), where p denotes the polynomial order. [7] These polynomials are obtained as solutions to the Chebyshev equation. The zero and the first order of Chebyshev polynomials are given respectively

$Ch_0(x) = 1$ and $Ch_1(x) = x$. (7)

The higher order polynomials are
$Ch_2(x) = 2x^2-1$
$Ch_3(x) = 4x^3-3x$
$Ch_4(x) = 8x^4-8x^2+1$

The recursive formula for generating higher order Chebyshev polynomial is given as

$$Ch\ p+1\ (x) = 2xChp\ (x) - Chp-1\ (x) \quad (8)$$

After getting final expansion for LeNN, the weighted sum of the components of the enhanced input pattern is obtained using the below formula

$$\text{Weighted Sum} = \sum wjCHXj \quad (9)$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

The error obtained by comparing the output with desired output is used to update the weights of the network structure by a weight updating algorithm. The Back Propagation algorithm, which is used to train the network, becomes very simple because of absence of any hidden layer its hidden layer. In every iteration the gradient of the cost function w.r.t the weights is determined and the weights are incremented by a fraction of the negative gradient

$$E\ k = 1/2[d\ k - y\ k]^{2} = 1/2\ e\ k^{2} \quad (10)$$

Where,
Ek is the error rate
dk is the desired value
yk is output of output of the network
ek is the error at kth step

$$w\ j,\ k+1 = w\ j,\ k + \alpha\ ek\ (1 - y\ k)\ 2\ CH(X) \quad (11)$$

Where,
w j, k+1 is the updated coefficient
w j,k is the old co efficient
Alpha is the learning rate
CH(X) is the expanded unit

III. SIMULATION STUDIES

A. Input Dataset

Diabetes data classification plays a major role in disease classification and analysis. Forwarding patient information. In this paper the data sets were based on the PIMA Indian Diabetes database of the UCI repository. The data set description is given in Table I.

TABLE I
DESCRIPTION OF DATASET

PIMA INDIAN DIABETES DATABASE	
No of Rows/instances	768
No of attributes plus class level	8 plus 1
No of rows- training	512
No. of rows - testing	256

Description or the Attributes in the Database:

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (mu Uml)
6. Body mass index (weight in kg (height in m) ^2)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1)

Class Distribution (class value 1 is interpreted as tested positive for diabetes)

Class Value Number of instances
0 500
1 268

B. Performance Metrics

For analysing the results and performance metrics of LeNN and CHFLANN the f-measure and accuracy has been considered for training and testing of input dataset.

With the specification of the diabetic data classification precision is the fraction of the relevant data instances and yet the recall is the relevant instances obtained. In both mathematical details we have calculated the f-ratio which is the harmonic mean of the precision and recall and accuracy which is the arithmetic mean of precision and recall.

The accuracy and f measure of the proposed works were evaluated in terms of confusion matrix values. These are TP (true positives), FN (false negatives), FP (false positives) and TN (true negatives) .The input data contain category level attributes such as 0 or negative (neg) and 1 or positive (pos) eventually we obtain the matrix of confusion based on the classified result shown in table II.

TABLE II
CONFUSION MATRIX

Actual Result	Classified Result	Value
1	1	TP
1	0	FN
0	0	TN
0	1	FP

The accuracy was calculated by using the following formula:

Specificity=TP/pos
Sensitivity=TN/neg
Precision=TP/TP+FP
Recall=TP/FN
Accuracy= sensitivity× pos/ (pos + neg) + specificity× neg / (pos+ neg) (12)
F-measure= (2×precision×recall)/ (precision+ recall)

IV. EXPERIMENTAL RESULTS

A. Description

In our proposed work to analyze the performance of LeNN and CFLANN, the PIMA Indian Diabetes database is divided into training and testing set of 512 and 256 attribute data respectively. Simulation was performed for the data by transferring each set of training and tests to LeNN and CFLANN 10 times taking into account the 3rd order polynomial equation with different learning rates (alpha) values and threshold values. After that the best accuracy is evaluated for a particular threshold value and learning rate, alpha. Finally the performance achieved with the best network size and the threshold value of both networks have been compared.

B. Experiment with LeNN

In the LeNN based diabetes data classification for the data mining operations, a better accuracy has been observed for the 0.5 threshold value and 0.04 learning rate. The detailed description has been given below in the table.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

The Accuracy value for the training dataset and testing dataset found to be 0.8086 and 0.7695 respectively showing satisfied result.

TABLE III
PERFORMANCE ANALYSIS (ACCURACY MEASURE) FOR DIABETES DATASET IN LeNN

Order	Learning rate	Threshold	Training accuracy	Testing accuracy
3	**0.08**	**0.5**	0.7617	0.8086
			0.7617	0.8047
			0.7695	**0.8086**
			0.7676	0.8047
	0.08	0.6	0.7559	0.7969
			0.7559	0.7930
			0.7439	0.7930
			0.7559	0.7930
	0.08	0.7	0.7423	0.7742
			0.7431	0.7816
			0.7412	0.7816
			0.7423	0.7812

TABLE IV
PERFORMANCE ANALYSIS (BEST LEARNING VALUE) FOR DIABETES DATASET IN LeNN

Order	Learning rate	Threshold Order	Training accuracy	Testing accuracy
3	**0.02**	**0.5**	**0.7695**	**0.8164**
	0.04	0.5	0.7695	0.8086
	0.06	0.5	0.7715	0.8086
	0.08	0.5	0.7695	0.8085

B. Experiment with CHFLANN

In the CHFLANN based diabetes data classification for the data mining operations, a better accuracy has been observed for the 0.5 threshold value and 0.08 learning rate. The detailed description has been given below in the table V.
The accuracy value for the training dataset and testing dataset found to be 0.7559 and 0.8086 respectively showing satisfied result.

TABLE V
PERFORMANCE ANALYSIS (ACCURACY MEASURE) FOR DIABETES DATASET IN CHFLANN

Order	Learning rate	Threshold	Training accuracy	Testing accuracy
3	**0.08**	**0.5**	0.7520	0.8086
			0.7559	0.8086
			0.7559	**0.8086**
			0.7578	0.8047
	0.08	0.6	0.7500	0.8086
			0.7500	0.8047
			0.7500	0.8086
			0.7480	0.8086
	0.08	0.7	0.7500	0.8047
			0.7520	0.8008
			0.7539	0.7969
			0.7520	0.8008

TABLE VI

PERFORMANCE ANALYSIS (BEST LEARNING VALUE) FOR DIABETES DATASET IN CHFLANN

Order	Learning rate	Threshold Order	Training accuracy	Testing accuracy
3	**0.02**	**0.5**	**0.7715**	**0.8164**
	0.04	0.5	0.7578	0.7969
	0.06	0.5	0.7422	0.7969
	0.08	0.5	0.7520	0.8047

C. Result Analysis

In our project work we have considered two different emerging techniques viz. LeNN and CHFLANN for the evaluation of the accuracy and f-measure for the Diabetes Data classification. The below table shows the comparative analysis for LeNN and CHFLANN for 2nd order and 3rd order polynomial equation. The analysis shows LeNN has the better performance with respect to the accuracy and f-measure for the diabetes dataset. The MSE (Mean Square Error) for LeNN and CHFLANN for the dataset have been represented in fig II and fig III respectively.

TABLE VI

PIMA INNDIAN DIABETES DATA CLASSIFICATION COMPARISON FOR LeNN & CFLANN					
ORDER		TRAINING DATASET		TESTING DATASET	
		ACCURACY	F-MEASURE	ACCURACY	F-MEASURE
LeNN	2	0.7734	0.9057	0.8008	0.9592
CHFLANN		0.7676	0.9032	0.8047	0.9899
LeNN	3	0.7695	0.8972	0.8086	0.9897
CHFLANN		0.7617	0.8858	0.7930	0.9505

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

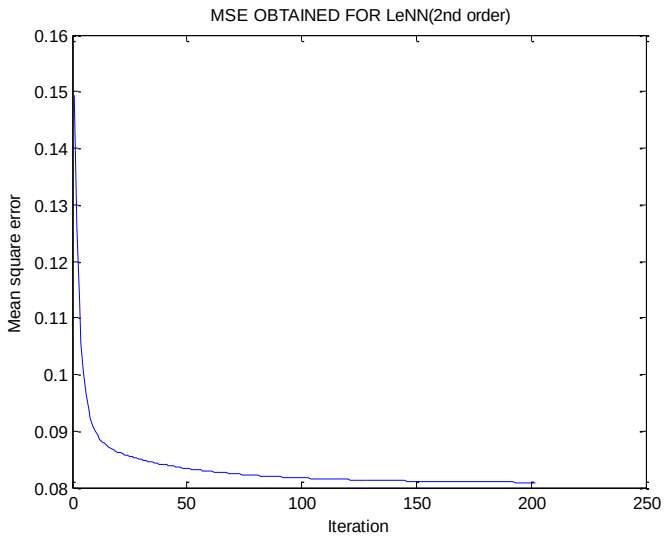

Fig II.Plot MSE vs. ITERATION for 2nd order LeNN

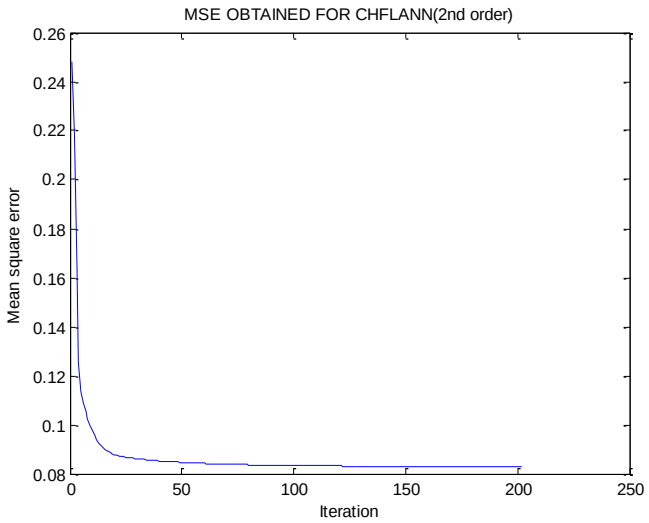

Fig III.Plot MSE vs. ITERATION For 2nd order CHFLANN

V. CONCLUSION

This study clearly shows the comparative performance of LeNN and CHFLANN over the PIMA Indian Diabetes database for diabetes data classification. The results view in the LeNN-set data worked better than the CHFLANN analysis and the f-measurement accuracy.

While analyzing the effect of both LeNN and CHFLANN processes, the 2nd order polynomial equation shows a significant difference in the results comparison compared to the 3rd order. So LeNN shows better performance in 2nd order calculations. However the f-measure gives better results for the classification of the 3rd order data as compared to the 2nd order data in the LeNN based method.

REFERENCES

[1] Joseph L. Breault, Colin R. Goodall, Peter J. Fos, "Data mining a diabetic data warehouse", *Artificial Intelligence in Medicine* vol.26, pp.37-54, 2002.

[2] E. G. Yildirim, A. Karahoca, T. Uar, "Dosage planning for diabetes patients using data mining methods", *Procedia Computer Science*, vol.3, pp.1374-1380, 2011

[3] Jagdish C. Patra, Pramod K. Meher, Goutam Chakraborty, "Non linear channel equalization for wireless communication systems using Legendre neural networks", *Signal Processing*, vol.89, pp.2251-2262, 2009.

[4] Fajiang Liu, Jun Wang, "Fluctuation prediction of stock market index by Legendre neural network with random time strength function", *Neurocomputing*, vol.83, pp.12-21, 2012.

[5] B.B. Misra, S.Dehuri, "Functional Link Artificial Neural Network for Classification Task in Data Mining", *Journal of Computer Science*, vol.3 (12), pp.948-955, 2007.

[6] Mukesh Kumar Mishra,Rajashree Dash, "A comparative study of Chebhyshev Functional Link Artificial Neural Network, Multi Layer Perceptron and Decision Tree for Credit Card Fraud Detection", *International Conference on Information Technology*,2014.

[7] [7] Emirhan Gulcin Yildirim, Adem Karahoca, Tamer Uçar, "Dosage planning for diabetes patients using data mining methods", *Procedia Computer Science*, vol.3, pp.1374-1380, 2011.

[8] Abdullah A. Aljumah, Mohammed Gulam Ahamad, Mohammad Khubeb Siddiqui*, "Application of data mining: Diabetes health care in young and old patients", *Journal of King Saud University-Computer and Information Sciences*, vol.25, pp.127-136, 2013.

[9] Mishra K. Sudhansu, Panda Ganpati, Meher Sukadev "Chebyshev Functional Link Artificial Neural Network for Denoising of Image Corrupted by Salt and Pepper", International journal of recent trends in Engineering, vol.1 (1), pp.413-417, 2009.

[10] Patra C. Jagdish, Kot C. Alex," Nonlinear Dynamic System Identification Using Chebyshev Functional Link Artificial Neural Networks",IEEE Transactions on Systems,pp.5-7,2002.

[11] Patra C. Jagdish, Thanh C. Nguyen, Meher K. Pramod,"Computationally Efficient FLANN-Based Intelligent Stock Price Prediction System",Proceedings of International joint Conference on Neural Networks,IEEE,pp.2431-2437,2009.

[12] https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/