International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

Optimal Deep Learning based Data Classification Model for Type-2 Diabetes Mellitus Diagnosis and Prediction System

M. Ganesan, N. Sivakumar, M. Thirumaran, R. Saravanan

Abstract: ***In recent days, deep learning models become a significant research area because of its applicability in diverse domains. In this paper, we employ an optimal deep neural network (DNN) based model for classifying diabetes disease. The DNN is employed for diagnosing the patient diseases effectively with better performance. To further improve the classifier efficiency, multilayer perceptron (MLP) is employed to remove the misclassified instance in the dataset. Then, the processed data is again provided as input to the DNN based classification model. The use of MLP significantly helps to remove the misclassified instances. The presented optimal data classification model is experimented on the PIMA Indians Diabetes dataset which holds the medical details of 768 patients under the presence of 8 attributes for every record. The obtained simulation results verified the superior nature of the presented model over the compared methods.***

Index Terms: ***Classification; Medical data; Deep Learning; Multilayer Perceptron.***

I. INTRODUCTION

Diabetes is metabolic disease type that is categorized by the use of hyperglycemia. With the dysfunction, different organs might be failed and long-term damage, the diabetes chronic hyperglycemia is linked mainly with kidneys, heart, eyes, blood vessels and nerves. In the diabetes development, many pathogenic procedures are included [1]. With continuous insulin deficiency, these vary from β-cells autoimmune destruction of pancreas towards abnormalities which gives the insulin action resistance. Insulin deficiency causes insufficient secretion of insulin or/and reduced responses of tissue towards insulin at one or additional points in composite hormone action pathways. In similar patient, impairments and defects of insulin secretion coexist frequently and it is regularly uncertain that abnormality is alone or is the major hyperglycemia cause.

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author
M. Ganesan*, Research Scholar, Dept. of CSE, Pondicherry University, Puducherry, India. ganesan@smvec.ac.in
N. Sivakumar, Assistant Professor, Dept. of CSE, Pondicherry Engineering College, Puducherry, India, sivakumar11@pec.edu,
M. Thirumaran, Assistant Professor, Dept. of CSE, Pondicherry Engineering College, Puducherry, India, tirumaran@pec.edu
R. Saravanan Assistant Professor, Dept. of IT, Sri Manakula Vinayagar Engineering College, Puducherry, India, r.saravanan26@gmail.com

Polydipsia, polyphagia, polyuria, weight loss, and blurred vision are the marked hyperglycemia symptoms.

The chronic hyperglycemia is accompanied by the impairment of susceptibility and growth. Without or with ketoacidosis syndrome, life-threatening and acute uncontrolled diabetes consequence is hyperglycemia.

With the probable vision loss, long-term diabetes complication involve retinopathy; peripheral neuropathy with amputations, Charcot joints and foot ulcers risks and autonomic neuropathy causes genitourinary, sexual dysfunction, gastrointestinal, cardiovascular symptoms. The diabetes patients comprise an increased peripheral arterial, cerebrovascular disease and atherosclerotic cardiovascular incidences. The lipoprotein metabolism abnormalities and hypertension are frequently found in diabetes patients. The complete insulin secretion deficiency is known as type 1 diabetes. The people with enhanced developing risk of this diabetes type can be recognized by autoimmune pathologic procedure serological occurrence that occurs at pancreatic islets and through genetic markers. A hyperglycemia degree is adequate to provide functional modifications and pathologic in different target tissues; however, with no clinical symptoms might be exist in earlier to diabetes detection.

Over time, the hyperglycemia degree might change based on the degree of underlying disease procedure. A disease procedure might exist however it does not comprise progressed in causing hyperglycemia. The impaired glucose tolerance (IGT) or/and impaired fasting glucose (IFG) can be caused through the similar disease procedure without satisfying the diabetes diagnosis condition. The sufficient glycemic control might be attained with exercise, oral glucose-lowering or/and weight reduction in few diabetes individuals. Hence, the individuals do not need insulin. However, another person with few remaining insulin secretions need exogenous insulin for sufficient glycemic control might live without it. The person with wide β-cell destruction and without remaining secretion of insulin needs insulin for endurance.

At present worldwide, there exist a total count of adult diabetic patients is 4.15 hundred million. The cost spent for worldwide diabetes treatment cost and difficulties is $673 billion cost in 2015. Over in globe, a person might be diagnosed as diabetic in each 3 second and a patient dies each 7 second due to diabetic complications. Additionally, numerous individuals comprise glucose tolerance in abnormal rate and without treatment or intervention would comprise diabetes even after 5 years.

The diabetes is a public health issue in worldwide [2, 3]. The procedure of mining and extricating data which comprise of a possible rate from a huge data sums through employing certain techniques.

Nowadays, technique of data mining had been extensively employed and given to numerous scientific as well as business domains. The treatment is digitized gradually with consequent computer technology development and huge sum of medical data had been preserved accurately in medical institutions. These resources of medical data find useful in several research domains [4, 5]. In medical database, techniques of data mining can be applied to the huge amount of historical data and derives precious rules for medical diagnosis [6]. The treatment, prevention and diagnosis can be aided through data mining [7-10]. Numerous data mining methods might be employed to cluster or classify the diseases. However, selection of features intended to simplify or enhance the computational efficacy of the model [11, 12]. For the diabetes disease classification, few researchers employs the technique of data mining with clinical test data like adiponectin, BMI and HbA1c and it attains the expected outcome [13]. The data mining application in clinical medicine is constrained relatively due to the human body complexity. The model of medical diagnosis analysis is constructed through the technique of data mining comprise of AI nature: it has the capability of excluding the human factors interference and it comprise firm objectivity, it might automate and standardize the medical diagnosis procedure gradually [7, 14, 15]. When compared to the individual physician conclusion with constrained case in career, the conclusion derived through the model of clinical diagnosis is given depending on the data of clinical detection of about several patients in decades. Therefore, the data mining application towards the facility of clinical medical disease offers wide prospects relatively for this application.

In addition, classification acts as a significant role in disease treatment problems of clinical diagnosis. By employing the dataset of Pima Indian diabetes, several researches had been performed over the data classification in diabetes. The researchers have employed various techniques for the accuracy and classification problem that had been enhanced by employing the California University dataset, Irvine (UCI) machine learning repository. By employing Least Square - Support Vector Machine (LS-SVM) and Generalized Discriminant Analysis (GDA), a cascade learning model is projected. They gained with the -fold cross-validation, the accuracy of 78.21% by employing LS-SVM and accuracy of 79.16% attained through it [9]. By employing Adaptive Neuro-Fuzzy Inference System (ANFIS) and Principal Component analysis (PCA) over the dataset of Pima Indian diabetes, an expert system had been built that gains the 89.47% as classification accuracy. For diabetes diagnosis, a common regression neural network is constructed and attained 80.21% as classification accuracy. A method depending on multilayer neural network is projected that attains 77.08% as classification accuracy. By employing L-BFGS method, an easier training method for DNN classifier had been projected. It attains 77.09% as classification accuracy. With the image dataset, a deep convolutional neural network depended method is projected for the diabetic retinopathy classification cases. By employing LR and k-means algorithm [16-18], a prediction system is constructed. Patient's diabetes mellitus are categorized by employing the approaches of machine learning. By employing Hoeffding Tree algorithm, it achieves recall and precision rate of 0.775 and 0.770 correspondingly. Through employing PCA, NN and SOM, accuracy enhancing model is projected.

In this paper, we employ an optimal deep neural network (DNN) based model for classifying diabetes disease. The DNN is employed for diagnosing the patient diseases effectively with better performance. To further improve the classifier efficiency, multilayer perceptron (MLP) is employed to remove the misclassified instance in the dataset. Then, the processed data is again provided as input to the DNN based classification model. The use of MLP significantly helps to remove the misclassified instances. The presented optimal data classification model is experimented on the PIMA Indians Diabetes dataset which holds the medical details of 768 patients under the presence of 8 attributes for every record. The obtained simulation results verified the superior nature of the presented model over the compared methods.

II. PROPOSED METHOD

An optimal DNN depended model is projected through the motivation of attractive deep networks features by employing stacked autoencoders for data classification of diabetes that enhances the entire performance measure of the problem. By employing the softmax layer and stacked autoencoders, the DNN classifier is constructed for diabetes dataset. Eight attributes are comprised through the dataset and the class attribute are given in subsequent parts. In addition, the misclassified instances in the dataset are removed by the use of multilayer perceptron (MLP). Once the misclassified instances are eliminated by MLP, DNN offers better classification performance. A set of eight attributes are given as input to the input layer. Two autoencoders stacked layers are constructed through DNN. With every 20 neurons, the network comprises of two hidden layers. With the final hidden layer, the softmax layer is included for the procedure of classification. For the applied records, the output layer would provide non-diabetic and diabetic class probabilities. Table 1 shows the attributes employed for the model simulation.

Table 1 Parameters used for simulation

Parameters	Values
L2Weight Regularization	0.01
Sparsity Regularization	4
Sparsity Proportion	0.05
Maximum epoch	1000
Learning rate	0.01
Loss function	Cross entropy
Training Algorithm	Scaled Conjugate Gradient

A. Training of layers

For autoencoder training $\{x_{(1)}, x_{(2)} \dots x_{(N)}\}$, input vectors denoted through N are subjected for training procedure. The input reconstruction is performed through the auto encoder training as given below.

$$x' = f_D(W', b'; f_E(W, b; x)) \tag{1}$$

Eq. (1) can be denoted by

$$x' = f_{AE}(W, b, W', b'; x) \tag{2}$$

where, f_{AE} denotes the function that maps the input towards the output in the autoencoder. Through reducing the main function that is provided through sum error function, the autoencoder is trained.

$$E_{Total} = E_{MSE} + E_{Reg} + E_{sparsity} \tag{3}$$

Let E_{Reg}, $E_{sparsity}$ and E_{MSE} represents the regularization factor, sparsity factor and mean square error. E_{MSE} is estimated through

$$E_{MSE} = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \tag{4}$$

Let the error is denoted through e_i that is variation among the original output, the examined output is denoted by $x\,(i\,)$. e_i is the error rate that might be estimated through

$$e_i = ||x(i) - x'(i)|| \tag{5}$$

In the training data set, the deep network learns each point which tends to model overfitting. This is the problem of deep networks as it gives reduced model performance over a novel testing data. E_{Reg} is the regularization factor that is assumed as main function that might be estimated to avoid this problem using,

$$E_{reg} = \frac{\lambda}{2}\left(\sum_{i=1}^{C} ||w_i|| + \sum_{i=1}^{D} ||w_i'||\right) \tag{6}$$

For the model regularization, λ is denoted. For the interesting features learning, sparsity constraint enables the system from the data. $E_{sparsity}$ is the sparsity factor that might be estimated,

$$E_{sparsity} = \beta \sum_{i=1}^{c} KL(\rho||\rho_j) \tag{7}$$

The sparsity weight term refers to β and Kullback–Leibler divergence refers the $KL(\rho||\rho_j)$ is derived through

$$KL(\rho||\rho_j) = \rho \, log\frac{\rho}{\rho_j} + (1-\rho)\frac{1-\rho}{1-\rho_j} \tag{8}$$

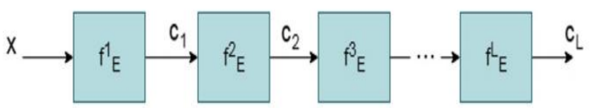

Fig. 1. Stacked autoencoder with L layers

Let ρ is the sparsity parameter constant and the average activation rate is demonstrated through ρ_j for the j^{th} neuron that might be estimated by

$$\rho_j = \frac{1}{T}\sum_{i=1}^{T} f^j(x_{(i)}) \tag{9}$$

In the autoencoder hidden layer, activation function is demonstrated through $f^j(x_{(i)})$ of jth neuron.

In the autoencoder hidden layer, activation function is demonstrated through $f^j(x_{(i)})$ of jth neuron.

B. The Stacked Autoencoder

Through encoder layer cascading, the deep network by employing the autoencoders is built and it is demonstrated in Fig. 1. The map autoencoders can be defined as

$$f_{SAE} = f_E^1 \circ f_E^2 \circ F_E^3 \dots \circ f_E^L \tag{10}$$

Let f_{SAE} is the stacked autoencoder function that is demonstrated above. The encoder function is used in every stacked autoencoder layer. In each and every layer, it is significant to denote the decoder function.

C. The softmax layer

Softmax classifier is one of the multiclass classifier that employs LR that classifies the data. The supervised learning technique is employed by softmax layer that employed expanded LR in order to multiple class classification. For softmax classifier, LR is the fundamental component. The softmax classifier estimates the every class probability in multiclass classifier problem with that the data is classified. Therefore, the sum of class probability would be equivalent to one other. For searching the probabilities of class, the softmax function carry out the procedure of exponentiation and normalization. f_{SC} is the function of softmax layer and is used with the stack autoencoder.

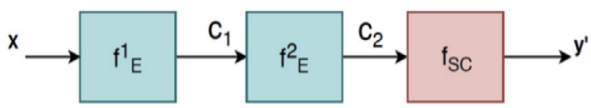

Fig. 2. A DNN framework with stacked autoencoder cascaded with softmax layer.

D. Fine tuning

The subsequent model training stage is known as fine tuning when the entire network layers are trained. In the procedure of classification, fine tuning is the end phase that is employed to enhance the model performance. The model is tuned finely to reduce the error in classification with supervised learning. The complete network is used for training as same as the multilayer perceptron training procedure using training data set.

E. Training method for DNN classifier

Through the cascading stacked auto encoder, the DNN classifier might be produced with the softmax classifier. Two or numerous autoencoders layers are comprised through the stacked autoencoder.

The DNN classifier with SAE comprises of two auto encoders is demonstrated in Fig. 3. Let $\{x_{(1)}, x_{(2)} \dots x_{(D)}\}$ is the input to the DNN and the respective output class attributes are $\{y_{(1)}, y_{(2)} \dots y_{(N)}\}$.

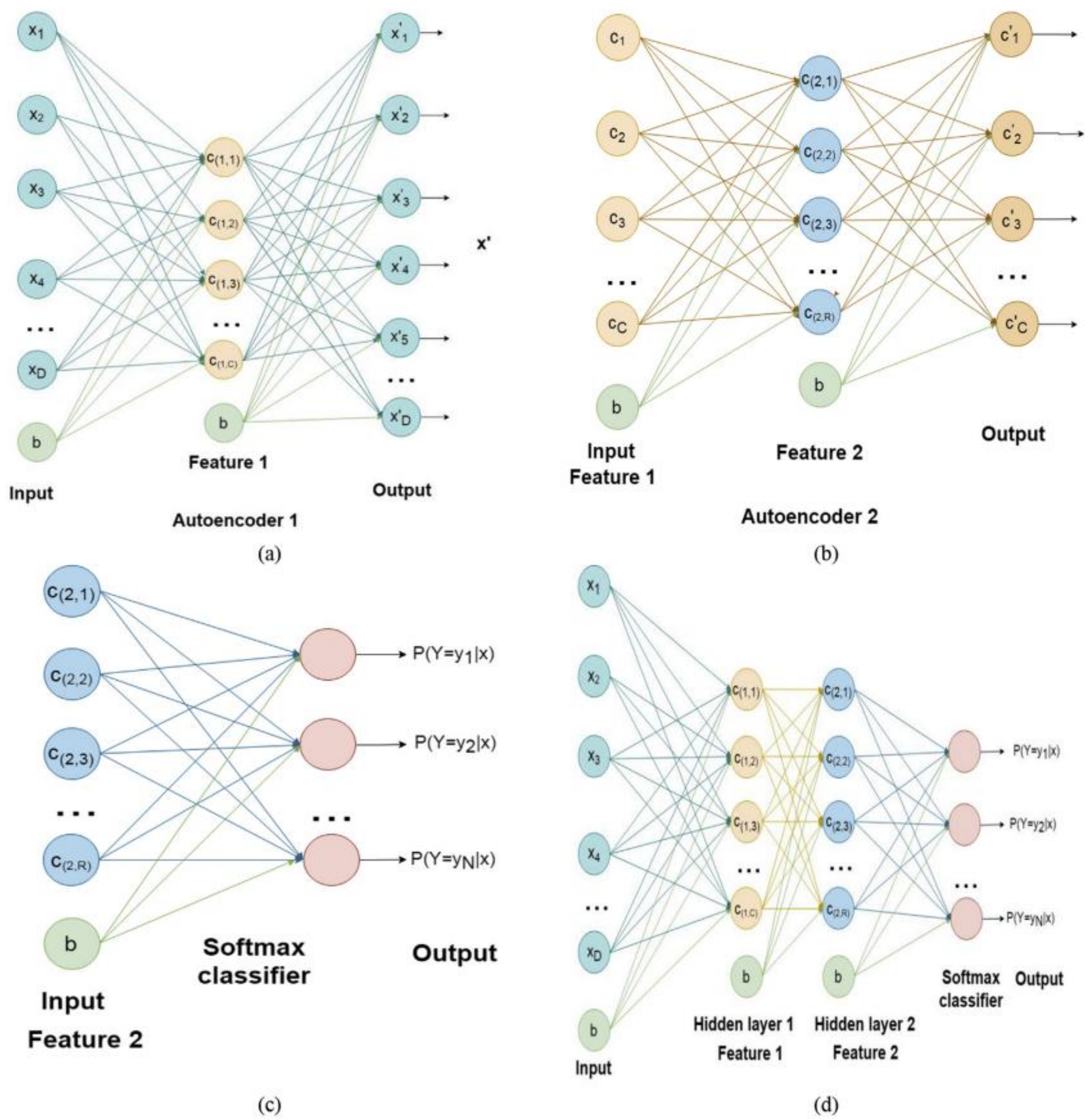

Fig. 3. (a) Network of autoencoder 1 (b) Network of autoencoder 2 (c) Softmax classifier (d) DNN.

The DNN classifier training is done by

1. With actual input vector $\{x_{(1)}, x_{(2)} \dots x_{(D)}\}$, the primary autoencoder layer is trained with the similar goal vector. Through $\{c_{(1,1)}, c_{(1,2)} \dots c_{(1,c)}\}$ features extraction, this layer attempts for input reconstruction with the autoencoder structure as demonstrated in Fig. 3.
2. The primary autoencoder layer output vector is considered as input vector towards the $\{c_{(1,1)}, c_{(1,2)} \dots c_{(1,c)}\}$ and creates $\{c_{(2,1)}, c_{(2,2)} \dots c_{(2,R)}\}$ output vector. The subsequent autoencoder layer attempts to rebuilt the $c_{(1,i)}; i = 1,2, \dots C$ as input.
3. With the layer of softmax classifier, stacked autoencoder is cascaded. Through assuming the subsequent autoencoder layer's output, this layer is trained and the actual class variables $\{y_{(1)}, y_{(2)} \dots y_{(N)}\}$.
4. Backpropagation is used to enhance the DNN classification performance that is denoted as fine tuning. In a supervised way, fine tuning is carried out through network retention with training data.

When the DNN performs the classification process, the results are validated in a clear way. In the next case, MLP is applied to remove the misclassified instances. MLP is the neural network based classifier derived from loading the input layer with the input vector and then propagates the actions in a feed forward way using the weighted links in the entire network (Hornik et al., 1989). For an input w_k, the state of i^{th} neuron (s_i) is equated in Eq. (11).

$$s_i = f\left(w_{i,0} + \sum_{j \in P_i} w_{i,j} \times s_j\right) \quad (11)$$

where, $\boldsymbol{f}$ is the activation function, $\boldsymbol{P_i}$ is the set of nodes reaching node $\boldsymbol{i}$, $\boldsymbol{w_{i,j}}$ is the weight of the connection among the nodes $\boldsymbol{i}$ and $\boldsymbol{j}$. MLP make use of an iterative function for learning process that begins with the random weights. A training algorithm is applied for controlling the weights to a particular target values.

The training will be stopped in case the error slope reaches to zero. MLP is useful due to the ability of solving problems in a stochastic way that offers approximate solutions for highly complex problems like fitness approximation.

III. RESULT ANALYSIS

The dataset from Kaggle is used to examine the projected method efficiency [19]. For the prediction of diabetes, the PIMA Indian diabetes dataset is usually employed. The total number of instances in the pima Indian database is 768. The total number attributes exist in the dataset is 8. The number of classes in prediction of diabetes is 2 such as positive and negative. The total percent of positive instances are 34.90% and the total percent of negative instances are 65.10%. The dataset description is given in table 2.

Table 2 Dataset Description

Description	Pima Indian Diabetes
No. of Instances	768
No. of Attributes	8
No. of Class	2
Positive Samples (%)	34.90%
Negative Samples (%)	65.10%
Data sources	[19]

The attribute description is provided in table 3. For instance, age and preg are the attributes that evaluates number of times got pregnant and age of the person. All the features that are taken into account are of numerical type. The statistical results such as Min, max, standard deviation and mean are given in the table 3. The frequency distribution of the given dataset is demonstrated in Fig. 4.

Table 3 Statistical Analysis of Pima Indian Diabetes Dataset

Attributes	Description	Type	Min.	Max.	Mean	Std. Dev
Preg	Number of times pregnant	Numerical	0	17	3.84	3.37
Plas	Plasma glucose concentration	Numerical	0	199	120.89	31.97
Pres	Diastolic blood pressure	Numerical	0	122	69.10	19.36
Skin	Triceps skin fold thickness	Numerical	0	99	20.54	15.95
Insu	2-hour serum insulin	Numerical	0	846	79.79	115.24
BMI	Body mass index	Numerical	0	67.1	31.99	7.88
Pedi	Diabetes pedigree function	Numerical	0.078	2.42	0.472	0.331
Age	Age	Numerical	21	81	33.241	11.76

Table 4 Confusion Matrix of Various Classifier Algorithms with Proposed Method for Diabetes Dataset

Experts	Proposed		LR		Voted Perceptron (VP)		Logit Boost		DT	
	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
Negative	456	44	440	60	462	38	423	77	407	93
Positive	109	159	115	153	217	51	122	146	108	160

The confusion matrix obtained through different classifier methods are demonstrated in Table 4 over the applied dataset. The confusion matrix represented that that the DNN model properly classifies 456 negative instances as negative and 159 positive instances as position. At the same time, it is noted that the DT model shows poor classification where it classifies 407 negative instances as negative and 160 positive instances as positive.

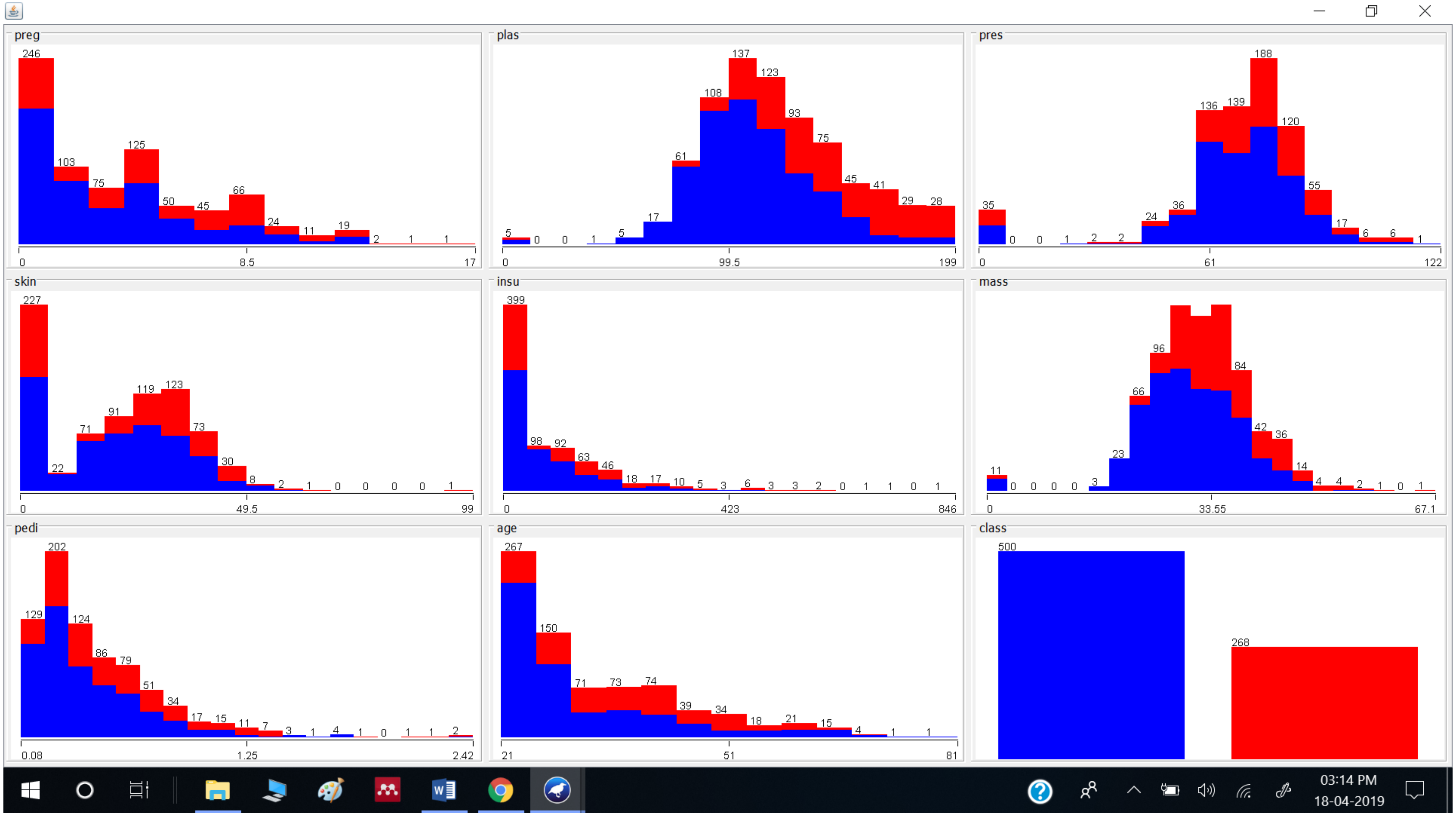

Fig. 4. Frequency Distribution of Pima Indian Diabetes Dataset for all Attributes

Table 5 Performance Evaluation of Different Classifiers with Proposed Method on Diabetes Dataset

Classifiers	Precision	Recall	Accuracy	F-score	Error Rate	MCC	Kappa
Proposed	91.20	80.71	80.08	85.63	0.19	0.54	53.54
LR	88.00	79.27	77.21	83.41	0.22	0.48	47.34
VP	92.40	68.04	66.79	78.37	0.33	0.17	13.52
LogitBoost	84.60	77.61	74.08	80.95	0.25	0.41	40.66
DT	81.40	79.02	73.83	80.19	0.26	0.42	41.64

A collection of performance evaluating parameters are employed to examine the projected method outcomes like precision, accuracy, error rate, kappa, recall, f-score and MCC as shown in Table 5. The projected method is compared with the existing methods like DT, LogitBoost, Voted Perception (VP) and LR. On comparing the classifier results of diverse methods, it is noted that the VP offers poor classification over all the other classifier models. At the same time, the DT shows somewhat effective classifier results over VP. However, it exhibits its inefficiency to outperform all the other methods. Simultaneously, it is revealed that the LogitBoost manages well over the VP and DT, but, it does not outperform the LR and proposed model. At the same time, the LR shows competitive performance with the higher classifier results. But, the presented DNN model shows effective classification and shows superior results with the maximum precision of 91.20, recall of 80.71, accuracy of 80.08, F-score of 85.63, MCC of 0.54 and kappa value of 53.54 respectively.

Table 6 Confusion Matrix Before vs. After Misclassified Instances Removed (MIR)

	After MIR-Proposed		Before MIR-Proposed	
Experts	**Negative**	**Positive**	**Negative**	**Positive**
Negative	373	12	456	109
Positive	12	194	44	159
Total	385	206	500	268
Total Instances	**591**		**768**	

The confusion matrix obtained using the proposed classifier method before and after MIR is demonstrated in table 6 over the applied dataset. The proposed method after MIR shows reduced count of false negative rate as 12 instances. And, the proposed method after MIR shows decreased number of false positive rate as it gives only 12 instances. At the same time, it is noted that a maximum of 373 negative instances is properly classified as negative and the maximum of 194 positive instances as positive. These values depicted that the false positive as well as false negative rate gets reduced by the inclusion of MIR.

To examine the projected method before and after the application of MIR, a set of performance evaluating parameters are employed like precision, accuracy, error rate, kappa, recall, f-score and MCC which is demonstrated in table 7. The comparison of different methods interms of precision and recall are shown in Fig. 5.

Table 7 Performance Evaluation Before vs. After Misclassified Instances Removed (MIR) on Diabetes Dataset

Classifiers	Precision	Recall	Accuracy	F-score	Error Rate	MCC	Kappa
After MIR-Proposed	96.88	96.88	95.94	96.88	0.04	0.91	91.06
Before MIR-Proposed	91.20	80.71	80.08	85.63	0.19	0.54	53.54

In terms of precision, the proposed method before MIR shows worst performance with minimal classification performance of 91.20. But, the after MIR, the proposed method exceeds the before MIR-proposed method by attaining the precision rate of 96.88. For recall, 80.71 value is the attained rate by before MIR-proposed whereas the proposed method after MIR outperformed the before MIR method through attaining higher recall value of 96.88.

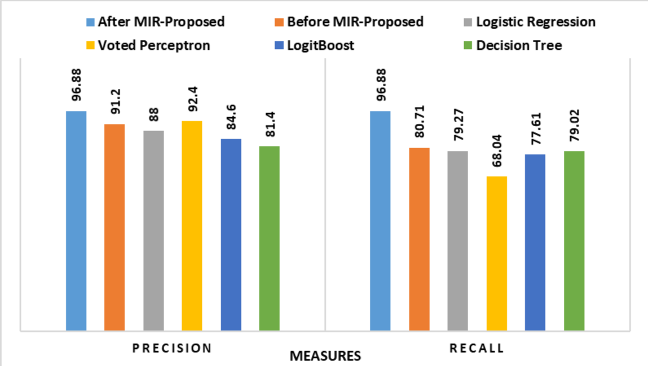

Fig. 5. Comparative results interms of precision and recall

In addition, the comparison of different methods interms of accuracy and F-score are shown in Fig. 6. From the figure, it is demonstrated that the after MIR proposed model shows effective classification by attaining an accuracy of 95.94 and F-score of 96.88 which is considerably high compared to the Before MIR proposed model which shows slightly lower classifier results with the accuracy value of 80.08 and F-score value of 85.63.

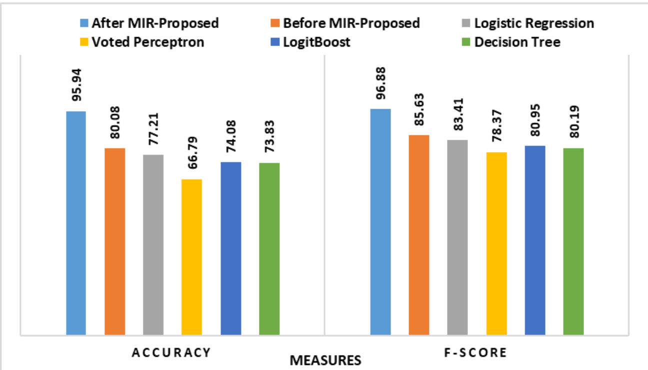

Fig. 6. Comparative results interms of accuracy and F-score

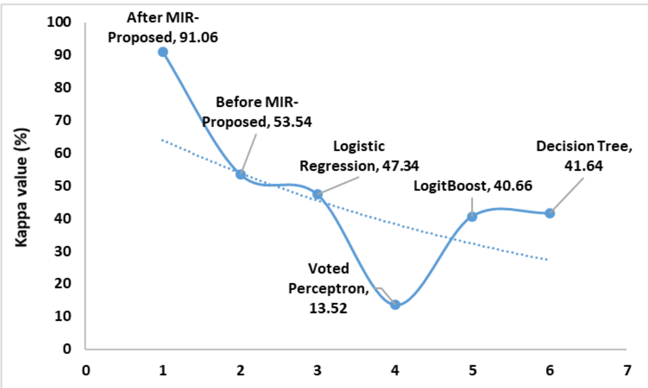

Fig. 7. Comparative results interms of kappa value

Fig. 7 shows the results attained by diverse models interms of kappa value. The values present in the figure clearly indicates that a maximum of 91.06 kappa value is attained by the presented after MIR method. In addition, the before MIR proposed model shows the kappa value of 53.54 which is higher than the values obtained by other methods except the after MIR proposed model. In line with, the VP model shows ineffective classification with the minimum kappa value of 13.52. These values define the superiority of the proposed model over the other methods with the maximum classification results. A detailed comparative analysis with the recent methods on the applied dataset is tabulated in Table 8 as well as in Fig. 8. From the table, it is evident that the maximum classification performance is attained by the proposed after-MIR model with the maximum classifier accuracy of 95.94.

Table 8 Comparison with Recent Methods with Proposed for Applied Dataset in terms of Accuracy

Classifiers	Accuracy	Reference
After MIR-Proposed	95.94	This Paper
K-means with Logistic Regression	95.42	Han et al. (2018)
Fuzzy Neural Classifier (FNC)	94.50	Priyan et al. (2018)
HPM	92.38	B.M. Patil
AMMLP	89.93	Alexis Marcano-Cedeno
J48 (pruned)	89.30	Aliza Ahmad
J48 (unpruned)	86.60	Aliza Ahmad
Hybrid Model	84.50	Humar Kahramanli
MLP	81.90	Aliza Ahmad
Logistic	78.20	Weka
J48	76.70	Weka
SGD	76.60	Weka
ELM	75.72	Rojalina Priyadarshini
NaiveBay	74.90	Weka
BayesNet	74.70	Weka
CART	72.80	Ster & Dobnikar
KNN	67.60	Statlog

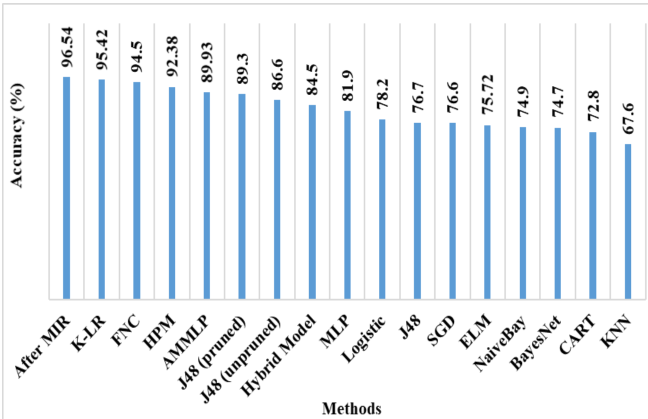

Fig. 8. Comparative accuracy analysis with recently proposed methods

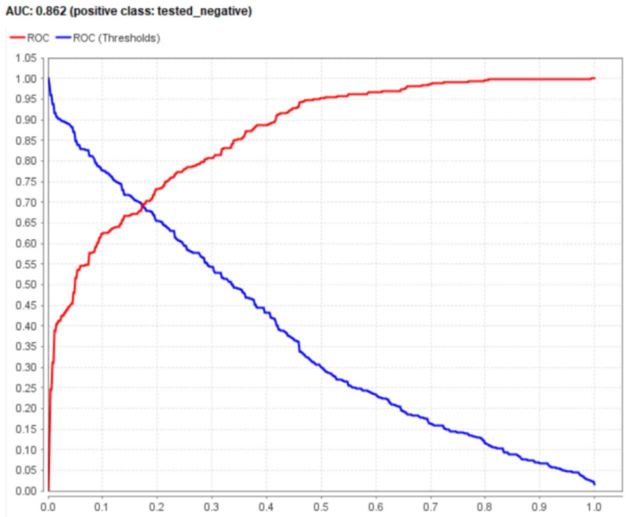

Fig. 9. AUC of DNN model Before MIR

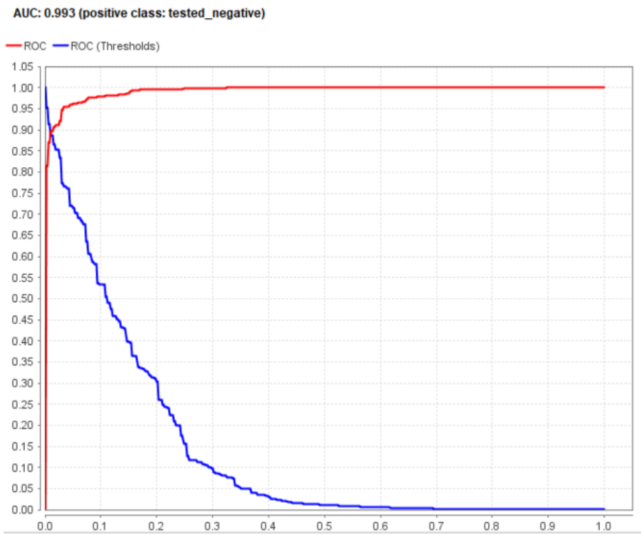

Fig. 10. AUC of DNN model After MIR-Proposed

To check the classification model performance, Area under curve (AUC) is used. At different threshold setup, it is performance measurement of the classifier. The predicting model is evaluated by the AUC, the model is superior at predicting 0s as 0s and 1s as 1s. When the AUC is the superior, the developed model is better at differentiating the diseased and non-diseased patients. From the Fig. 9, it is evident that the presented model achieves an AUC of 0.862 which is effective on the applied dataset. But, on observing the Fig. 10, it is evident that the AUC after MIR is 0.993 which is higher than AUC before MIR. From these values, it is evident that the presented optimal DNN classifier is found to be effective over the compared methods.

IV. CONCLUSION

In this paper, an optimal DNN depended model is projected through the motivation of attractive deep networks features by employing stacked autoencoders for data classification of diabetes that enhances the entire performance measure of the problem. By employing the softmax layer and stacked autoencoders, the DNN classifier is constructed for diabetes dataset. In addition, the misclassified instances removal MIR process takes place MLP. Once the misclassified instances are eliminated by MLP, DNN offers better classification performance. For the prediction of diabetes, the PIMA Indian diabetes dataset is usually employed. An extensive experimental outcome ensured the superior performance of the presented model over traditional as well as recently presented methods. In future, the performance can be further enhanced by the use of feature selection methods.

REFERENCES

1. Chen H, Tan C, Lin Z, Wu T. The diagnostics of diabetes mellitus based on ensemble modeling and hair/urine element level analysis. Comput Biol Med. 2014;50:70–75.
2. Mohamed EI, Linder R, Perriello G, Di ND, Poppl S, De AL. Predicting type 2 diabetes using an electronic nose-based artificial neural network analysis. Diabetes, nutrition & metabolism. 2002;15(4):215–221.
3. Polat K, Gunes S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digit Signal Process. 2007;17(4):702–710.
4. World Health Organization. Global Statistics Reports on NCDs. 2016; 2016http:// www.who.int/diabetes/global-report/en/, Accessed date: 10 September 2018. 5.
5. Cheruku R, Edla and DR, Kuppili V, Sm-ruleminer. Spider monkey based rule miner using novel fitness function for diabetes classification. Comput Biol Med. 2017;81:79–92.
6. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313(5786):504–507. https://doi.org/10.1126/science. 1127647.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436. https:// doi.org/10.1038/nature14539.
8. Kayaer K, Yıldırım T. Medical diagnosis on pima indian diabetes using general regression neural networks. Proceedings of the International Conference on Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP). 2003; 2003:181–184 Istanbul, Turkey, June 26–29.
9. Caliskan A, Yuksel ME, Badem H, Basturk A. Performance improvement of deep neural network classifiers by a simple training strategy. Eng Appl Artif Intell. 2018;67:14–23https://doi.org/10.1016/j.engappai.2017.09.002.
10. Duch W. Dataset used for classification comparision of results. http://fizyka.umk.pl/ kis-old/projects/datasets.html; 2018, Accessed date: 10 September 2018.
11. Ng A. Sparse autoencoder. https://web.stanford.edu/class/cs294a/ sparseAutoencoder_2011new.pdf; 2011, Accessed date: 10 September 2018.
12. Le QV, Ngiam J, Coates A, Lahiri A, Prochnow B, Ng AY. On optimization methods for deep learning. Proceedings of the 28th International Conference on International Conference on Machine Learning. Omnipress, Bellevue. 2011; 2011:265–272 Washington, USA - June 28 - July 02.
13. Poultney C, Chopra S, Cun YL, et al. Efficient learning of sparse representations with an energy-based model. Advances in Neural Information Processing Systems,. 2007; 2007:1137–1144.
14. Uci machine learning repository. available at http://archive.ics.uci.edu/ml/; 2018 [Online].
15. Mohamadi H, Habibi J, Abadeh MS, Saadi H. Data mining with a simulated annealing based fuzzy classification system. Pattern Recogn. 2008;41(5):1824–1833https://doi. org/10.1016/j.patcog.2007.11.002.
16. Wan S, Liang Y, Zhang Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. Comput Electr Eng. 2018;72:274–282.
17. Wu H, Yang S, Huang Z, He J, Wang X. Type 2 diabetes mellitus prediction model based on data mining. Informatics in Medicine Unlocked. 2018;10:100–107
18. Mercaldo F, Nardone V, Santone A. Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia Computer Science. 2017;112:2519_2528. https://doi.org/10.1016/j.procs.2017.08.193.
19. https://www.kaggle.com/uciml/pima-indians-diabetes-database.