

An Efficient Ada Max based Parameter Tuned Deep Neural Network for Medical Data Classification

R. Raja¹, B. Ashok²

¹Assistant Professor, ThiruKolanjiappar Govt. Arts College, Viruthachalam, India.

²Assistant Professor, PSPT MGR Govt. Arts and Science College, Sirkali, India.

[¹rajamanira2000@gmail.com](mailto:rajamanira2000@gmail.com), [²ashok.au@gmail.com](mailto:ashok.au@gmail.com)

Abstract

Medical data classification involves the application of intelligent algorithms to examine the medical dataset for the detection of diseases. This paper concentrates on a medical data classification process to determine the existence of particular diseases for diagnostics and prognostics. The proposed model uses an AdaMax based deep neural network (DNN) model, called AM-DNN for medical data classification. The presented AM-DNN model comprises different processes namely preprocessing, classification, and parameter optimization. The presented model preprocesses the medical data in the initial phase to transform it into a compatible format. In addition, DNN based classification process gets executed to allocate the proper class label of DNN. Besides, AM optimizer is applied to fine-tune the parameters of DNN model. The application of AM helps to improvise the efficiency of the DNN model. For assessing the simulation performance of the AM-DNN model, a series of experiments were performed. The obtained outcomemakes sure that the AM-DNN model has resulted in a maximum accuracy of 0.9275, 0.8945, and 0.9333 on the applied chronic kidney disease (CKD), diabetes, and heart disease datasets.

Keywords: Medical data classification, deep learning, AM optimizer, CKD, classification

1. Introduction

Medical analysis plays an effective role in enhancing the patient's lifetime and maintain the health state to a greater extent which depends upon the patient's clinical and non-clinical profile. It is assumed to be an essential process in medical sector to diagnose the disease from examined symptoms [1]. But it is a complicated process for medical experts to offer an accurate health report after examining the person with massive number of attributes. Due to the existence of dense medical information derived from clinicians and clinical models, medical diagnosis ends up with imprecise, irregular, and unreliable patient details which are highly essential for clinical diagnosing issues. With the complication of diverse diseases and

inexistence of knowledge regarding a problem, valid details are inaccessible due to the presence of irregular medical information. Therefore, uncertainty is one of the essential factors in clinical diagnosing process [2].

In general, clinical information illustrates the exclusive features like noise derived from human and logical errors, imputed values, sparseness, and so forth. Here, data quality plays an important role in mining the final outcome [3]. Briefly, Neurocognitive infections are examined as dreadful diseases in hospitals. A transparent and evident disease diagnosis is extremely significant and correlated to diverse results with enhanced management of unpredicted neuropsychiatric representations are assumed to be major medical target [4]. In order to overcome these issues, different types of corrective solutions are prioritized. Several diseases diagnosing principles are dependent on consecutive examinations which are common and distinguishing clinical analyzing operations. Hence, it is named sequential diagnosis which contributes to solving multifaceted decision issues [5]. Similarly, supervised learning approaches are involved in improving the health state within a limited time period.

[6] deployed an exponential programmed outline of rule based classifier (FRBCSs) from clinical data with the help of Multi-Objective Evolutionary Optimization Algorithms (MOEOAs). It is generated in a solitary run and gathered of arrangements (clinical FRBCSs) implied by several accuracy levels. [7] defined various Data Mining (DM) models applied for cancer analysis. The lung disorder pathology is allocated by means of pathology defines the size and degree of initial tumor and cancer spreading rate (metastasis). Observing the lung infection pathology is considered to be the important process on grounds which is applied for monitoring the patient's health state which is applicable for physicians to provide appropriate treatment. [8] illustrated the application of Support Vector Machines (SVM) variable tuning plan that applied traditional Fruit Fly Optimization Algorithm (FOA) called FOA-SVM and the procedure is related to clinical analysis. The developed FOA-SVM is a combination of FOA and SVM. Additionally, the sufficiency and efficiency of FOA-SVM can be applied over 4 clinical datasets.

[9] deployed a model for reducing the response time prior to reduce the clinical centres of anti-haemorrhagic therapies and dispense the sensitive outcome of intracranial leakage. [10] developed a framework to diagnose Parkinson's sickness (PD) with the help of Magnetic Resonance Imaging (MRI) details. Especially, Joint Feature-Sample Selection (JFSS) approach was presented initially for selecting unique subset of samples and properties to

consider dependable diagnosing process. Hence, the selected features are considered to be major portion of PD representation that differentiate important and fundamental imaging biomarkers for PD. Here, a significant classifier has been developed to eliminate the noise subset of features as well as samples and applied the classification method. Similarly, noise removal is important in developing appropriate data. Unlike the conventional process, de-noising is computed in unsupervised fashion and processed by de-noising operation in developing and testing details with diagnostic accuracy.

Medical Images (MI) are relevant wellsprings of data to distinguish and diagnose wider range of ailments and differences. Because of the essential, this works is carried out on Breast Ultrasound (BUS) that is a basic subclass for mammography to examine common bosom values for females globally. Followed by, in order to update data security, picture consistency, authenticity, and ingredient validation in e-wellbeing scenarios, MI watermarking is widely applied and principle intension is to incorporate persistent meta-data within MI and the frequent picture retains better data quality. [11] illustrated the analysis of 2 watermarking models as Specific Spread range in Discrete Cosine Transform (SS-DCT) as well as High-Capacity Data-Hiding (HCDH) method; thus the watermarked BUS images make sure to adequate for Computer-Aided Diagnosis (CADx) approach, where 2 basic outcomes are sore division as well as classification. [12] depicted a new model for examining lung cancer previously under the application of Cuckoo Search (CS) optimizer and SVM classifier. [13] implied a model for selecting better set of features with the help of Salp Swarm Algorithm (SSA) relied on Particle Swarm Optimization (PSO) effectively.

This paper focuses on a medical data classification process to determine the presence of specific diseases for diagnostics and prognostics. The proposed model uses an Adamax (AM) based deep neural network (DNN) model, called AM-DNN for medical data classification. The presented AM-DNN model comprises different processes namely preprocessing, classification, and parameter optimization. The presented model preprocesses the medical data in the initial phase to transform it into a compatible format. In addition, DNN based classification process gets executed to allocate the proper class label of DNN. Also, AM optimizer is applied to fine-tune the parameters of DNN model. The application of AM benefits to improvise the effectiveness of the DNN method. For evaluating the simulation performance of the AM-DNN method, a series of investigations are accomplished.

2. The Proposed AM-DNN Model

Fig. 1 briefs the overall working procedure contained in the AM-DNN model. The figure demonstrated that the input medical data is initially preprocessed to improve the data quality. Followed by, the preprocessed data is fed into the DNN based classification module to determine the appropriate class labels. At the same time, the AM optimizer is applied to fine-tune the parameters of the DNN model.

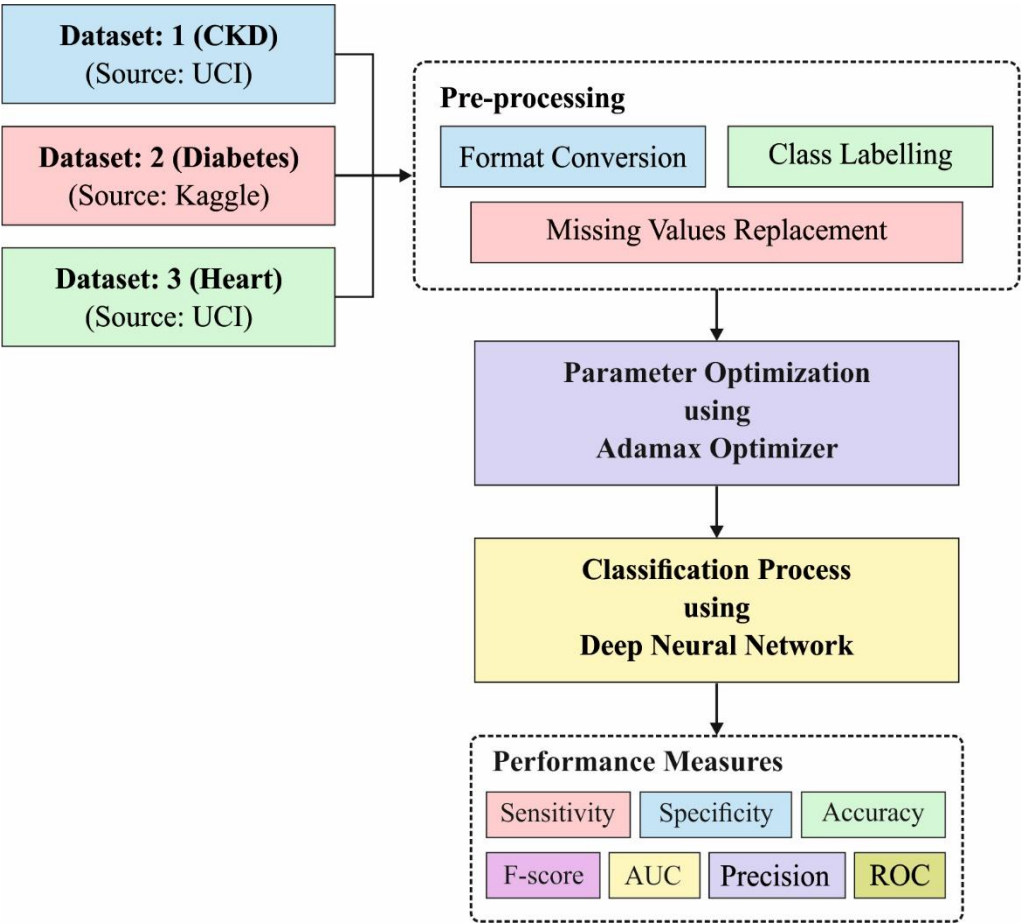


Fig. 1. Working process of AM-DNN model

2.1.Data Pre-processing

Here, the input clinical data undergoes pre-processing for the purpose of enhancing the data quality in 3 formats. Initially, data conversion is performed in which the input data is in .xls format and it is changed into .csv format. Then, a class labeling task is processed where the data samples are assigned to respective class labels. Finally, replacement of missing value is performed with the help of k-nearest neighbors (KNN) approach. In this approach, KNN is an effective and elegant model which saves the previous cases and classifies new cases

according to the similarity metrics. At this point, KNN approach has been applied as a material for data imputation [14]. Also, the strategy behind K-NN is given below:

- Estimate the parameter k: The measure of parameter k is fixed as 5. When k is allocated with low value, then high noise may exist and reduce the classification accuracy. Meantime, when k is assigned with minimum value, which results in limited noise and maximum classification accuracy.
- Evaluate the Euclidian distance amongst missing values and measure with the help of Eq. (1).

$$d_{(x,y)} = \sqrt{\sum_{j=1}^s (x_j - y_j)^2} \quad (1)$$

Where $d_{(x,y)}$ signifies Euclidian distance, J implies data parameter with $j = 1, 2, 3, \dots, s$, s means the data dimensions, x_{aj} signifies value from j - attribute in conjunction with missing value, and y_{bj} refers the score from j - attribute without missing data.

- According to the distance information derived, minimum Euclidian distance upon the variable k is processed as imputed value in missing details. Therefore, the measure of imputation is computed under the application of Weight Mean Estimation mechanism as applied in Eq. (2).

$$x_j = \frac{\sum_{k=1}^K w_k v_k}{\sum_{k=1}^K w_k} \quad (2)$$

Where x_j implies the Weight Mean Estimation, K refers the parameter value of $k = 5$, w_k indicates nearest neighbor observation value, and v_k means the value from full set of variables with dropped values relevant to parameter k. Here, the expression of w_k is estimated using Eq. (3).

$$w_k = \frac{1}{d_{(x,y)}^2}, \quad (3)$$

Where $d_{(x,y)}$ represents Euclidian distance of each parameter k.

2.2.Data Classification

In general, DL models are suitable for gaining highly-dimensional properties from the input dataset. Followed by, features gathered from DNN are utilized for improvising the performance of classifiers. The prominently applied Deep learning (DL) approach is DNN classifier which has been developed by integrating heap of auto-encoder (AE) systems with the application of softmax (SM) classification method [20-25].

2.2.1. AE Network

Normally, AE is involved in input, hidden, and output layers. Here, AE is trained in unsupervised approach for creating an input finally with limited error. Therefore, the output is identical to input. The main aim of training the AE is for embedding the input with feature spaces and contributes to reducing dimension when compared to input space. So, dimensions of a code space can be decided as higher than the input space for improving the classification accuracy in specific events. As a result, AE manages to provide best illustration of input vector by replacing an appropriate code.

Fig. 2 portrays the system of AE where the number of neurons from final layer is identical to input values. The M dimension input vector is illustrated by $u^{(1)}, u^{(2)} \dots u^{(Z)}$. Here, T refers the value of input vectors. A left part of AE is named as encoder, where input is considered as input for AE with final result is defined as an outcome of hidden layer. Also, encoder changes the input vector as a code with efficient input vectors. Finally, input, as well as output connected of encoder, could be implied as $c = g_E(W, b; u)$ and demonstrated in Eq. (4):

$$c = f(b + W^Z u)$$
 (4)

Where f means an activation function of encoded neurons.

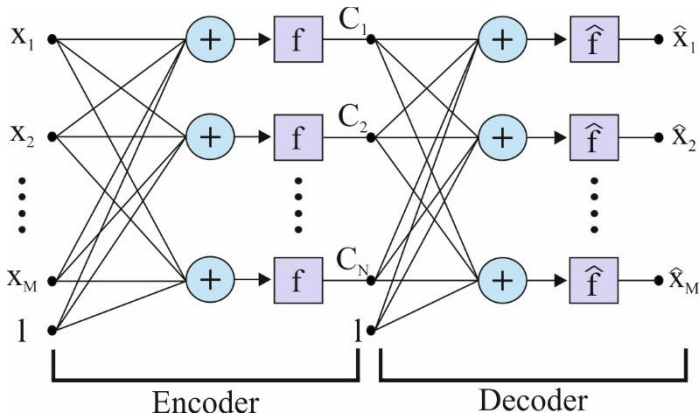


Fig. 2. Network structure of AE

Actually, weight of an encoder can be represented by W matrix which correlates the inputs of hidden layers and b vector of neuron bias. Hence, vector u and vector carementioned as input and output of encoder.

Then, right side part of AE is termed as decoding unit or decoder in which the outcome of hidden layer (c) is fed as input, and result \hat{u} refers the result of AE. The decoder is comprised of weight matrix \hat{W} and \hat{b} vector converts the applied code vector to original input vector that combines low errors [15]. Hence, the association from input and output of encoder is shown below:

$$\hat{u} = \hat{f}(\hat{b} + \hat{W}c) \quad (5)$$

During this model, \hat{f} signifies an activation function of decoding neurons. The input as well as output relation of decoder is represented as $\hat{u} = g_D(\hat{W}, \hat{b}; c)$. The network architecture of AE is illustrated in Fig. 3 and simulation result of AE is depicted as $\hat{u} = g_{AE}(W, b, \hat{W}, \hat{b}; u)$.

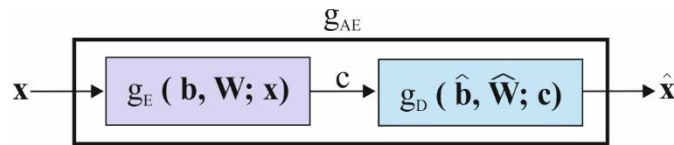


Fig. 3. Layers in Cascading Encoder/Decoder

An objective process of AE is defined as shown in the following:

$$E_{\text{sparse}} = E_Z + \beta \sum_{q=1}^N KL(\rho || \hat{\rho}) \quad (6)$$

In previous cost, function is operated in 2 parts. Firstly, E_Z refers the objective function of NN. The β implies the weight of sparsity penalty in Eq. (6):

$$E_Z = \frac{1}{Z} \sum_{k=1}^Z e_k^2 + \frac{\lambda}{2} (\|W\| + \|\hat{W}\|) \quad (7)$$

Where λ signifies a regularization term used for eliminating over-fitting issues. Hence, error vector is considered as difference amongst the chosen outputs and ground result as demonstrated as depicted below:

$$e_k = \|u^{(k)} - \hat{u}\| \quad (8)$$

Where $k = 1, 2, \dots, Z$. It can be simple for observing that E_Z is a function which represents the inner weight of AE as depicted in the following:

$$E_Z = E_{AE}(W, b, \hat{W}, \hat{b}). \quad (9)$$

In secondary part of Eq. (6), is illustrated below:

$$KL(\rho \parallel \hat{\rho}_q) = \rho \log \frac{\rho}{\hat{\rho}_q} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_q} \quad (10)$$

Where ρ denotes the sparsity measure and $\hat{\rho}$ means the value as depicted in Eq. (11):

$$\hat{\rho}_j = \frac{1}{Z} \sum_{p=1}^Z f_q(u^{(i)}) \quad (11)$$

AE units are inter-linked for developing Stacked Autoencoder (SAE) system.

2.2.2. SAE Network

The encoder is comprised of multiple AE and connected for making SAE as depicted in Fig. 4. By reforming the input-output relationship of AE SAE system with L cascaded AEs are derived as depicted in the following:

$$g_{SAE} = g_E^1 \circ g_E^2 \circ \dots \circ g_E^L \quad (12)$$

The SAE approach is built by using encoding part of trained AE. Decoding parts of AEs are not employed in developing SAE as it can be applied for AE training as shown below.

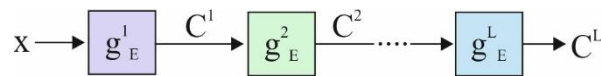


Fig. 4. Network structure of SAE

2.3. Parameter Optimization of DNN

In order to optimize the parameters of DNN, AM technique is employed. In this approach, NN training is one of the common optimization problems along with non-convex objective function and minimization issues $\min_{\theta} J(\theta; \mathcal{D}_{\text{train}})$. In case of training process, model attributes θ have been maximized iteratively to mitigate the cost of training data $\mathcal{D}_{\text{train}}$. Followed by, bold symbols like θ has been applied vector quantities and regular values for scalar quantity. The typically used termination criteria for iterative training is a predefined

value passed through readily available training data, named epochs. Each epoch has several iterations.

There are different types of optimization approaches available in this study which are used for NN training process which varies in upgrading the network variables. The performance efficiency can be measured using 2 measures namely:

- Speed of convergence: Time required for a model to achieve best value.
- Generalization: Performance of a model on newly arrived data.

Optimization of DNN has massive number of challenges. Initially, extremely non-convex objective function with enormous suboptimal local minima as well as saddle points. Alternatively, high-dimension of search space and proper measures for hyperparameters. Also, conventional and adaptive optimization methods are generally employed for optimizing NN process.

The Adam optimization [16] method was developed for combining the advantages of Nesterov momentum, AdaGrad, and RMSProp methodologies. Then, weights are maximized on the basis of Eq. (13):

$$w_t^i = w_{t-1}^i - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \cdot \hat{m}_t \quad (13)$$

where:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (14)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (15)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)G \quad (16)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)[G]^2 \quad (17)$$

$$G = \nabla_w C(w_t) \quad (18)$$

where η denotes the learning rate hyperparameter, w_t implies the weights at step t , $C(.)$ defines the cost function, and $\nabla_w C(w_t)$ means the gradient of weight attributes w_t for image x and respective label y , β_i is applied for selecting the volume of data required from existing

update, in which $\beta_i \in [0, 1]$, m_t defines the running average of gradients and named as the primary moment, v_t signifies the running average of squared gradients and is named as secondary moment. When the primary and secondary moments are allocated as 0, then it is biased to resolve the zero-biased issue and moments are bias-corrected by classifying the concerned β .

Adamaxis an extended version of Adam approach, in which distributed variance intends to ∞ . Also, weights are maximized according to Eq. (19):

$$w_t^i = w_{t-1}^i - \frac{\eta}{v_t + \epsilon} \cdot \hat{m}_t \quad (19)$$

where:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (20)$$

$$v_t = \max(\beta_2 \cdot v_{t-1}, |G_t|) \quad (21)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) G \quad (22)$$

$$G = \nabla_w C(w_t) \quad (23)$$

where η denotes the learning rate hyperparameter, w_t defines the weights at step t , $C(\cdot)$ implies the cost function, and $\nabla_w C(w_t)$ signifies the gradient of weight parameters w_t for image x and corresponding label y . β_i is employed for selecting quantity of details required from old update, where $\beta_i \in [0,1]$. m_t and v_t represents primary and secondary moment.

Algorithm 1:Pseudocode of AdaMax

η : Learning score

$\beta_1, \beta_2 \in [0, 1)$: Exponential decay values for moment candidates

$C(w)$: Cost function with variable w

w_0 : Primary parameter vector

$m_0 \leftarrow 0$

$u_0 \leftarrow 0$

$i \leftarrow 0$ (Stimulate time step)

while w is not converged do

$i \leftarrow i + 1$

$$m_i \leftarrow \beta_1 \cdot m_{i-1} + (1 - \beta_1) \cdot \frac{\partial C}{\partial w}(w_i)$$

$$u_i \leftarrow \max\left(\beta_2 \cdot u_{i-1}, \left|\frac{\partial C}{\partial w}(w_i)\right|\right)$$

$$w_{i+1} \leftarrow w_i - (\eta/(1 - \beta_1^i)) \cdot m_i/u_i$$

end while

return w_i (final variables)

3. Experimental Validation

The proposed AM-DNN model has been simulated using Python 4.6.5 tool and the results are determined under three datasets. Table 1 shows the dataset description. Firstly, the CKD dataset includes a maximum of 400 instances with 24 features and 2 class labels. In addition, a set of 250 samples fall into positive class labels and remaining 150 instances come under negative class labels. Secondly, the Diabetes dataset is composed of maximum of 768 instances with 8 features and 2 class labels. Also, a set of 268 samples belongs to positive class label and remaining 500 instances lie in negative class label. Thirdly, the Heart disease dataset contains a maximum of 270 instances with 13 features and 2 class labels. Additionally, a set of 120 samples comes under positive class label and remaining 150 instances fall into negative class label.

Table 1 Dataset Description

Description	CKD	Diabetes	Heart Disease
No. of Instances	400	768	270
No. of Attributes	24	8	13
No. of Class	2	2	2
No. of Positive Samples	250	268	120

No. of Negative Samples	150	500	150
Data source	[17]	[18]	[19]

Fig. 5 showcases three confusion matrices generated by the AM-DNN model on the applied three benchmark datasets. Fig. 5a illustrates that the AM-DNN model has categorized a set of 137 images under false class and 234 images under true class effectively. Concurrently, Fig. 5b implies that the AM-DNN method has classified a set of 472 images under false class and 215 images from true class significantly. In line with, Fig. 5c signifies that the AM-DNN approach has divided a set of 142 images from false class and 110 images under true class effectively.

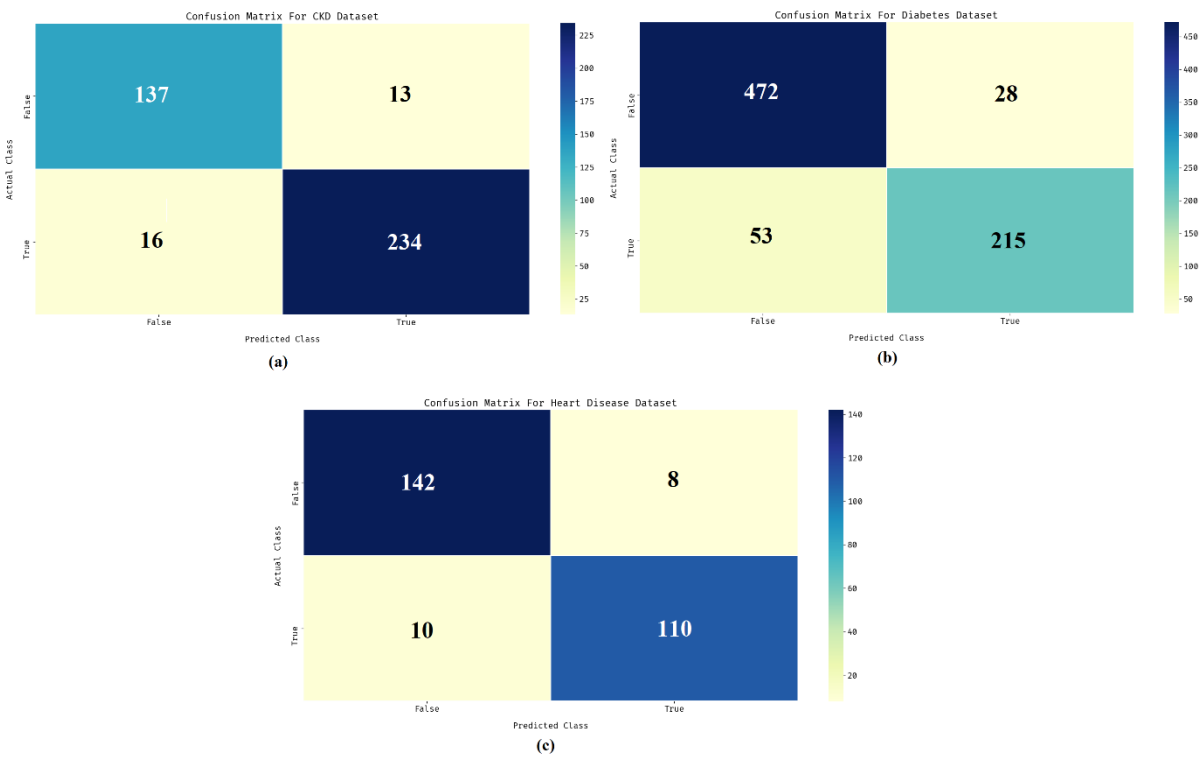


Fig. 5. Confusion Matrix a) CKD Dataset b) Diabetes Dataset c) Heart Disease Dataset

Table 2 and Fig. 6 investigates the results obtained by the AM-DNN model interms of different measures. On the applied test CKD dataset, the AM-DNN model has resulted in an improved sensitivity of 0.9133, specificity of 0.9360, precision of 0.8954, accuracy of 0.9275, F-score of 0.9043, and AUC of 0.9247. At the same time, on the given test Diabetes dataset, the AM-DNN method has resulted in enhanced sensitivity of 0.9440, specificity of 0.8022, precision of 0.8990, accuracy of 0.8945, F-score of 0.9210, and AUC of 0.8731. On the other hand, on the applied test Heart disease dataset, the AM-DNN framework has

provided maximum sensitivity of 0.9467, specificity of 0.9167, precision of 0.9342, accuracy of 0.9333, F-score of 0.9404, and AUC of 0.9317.

Table 2 Result Analysis of Proposed AM-DNN Model

Sensitivity	Specificity	Precision	Accuracy	F-Score	AUC
CKD Dataset					
0.9133	0.9360	0.8954	0.9275	0.9043	0.9247
Diabetes Dataset					
0.9440	0.8022	0.8990	0.8945	0.9210	0.8731
Heart Disease Dataset					
0.9467	0.9167	0.9342	0.9333	0.9404	0.9317

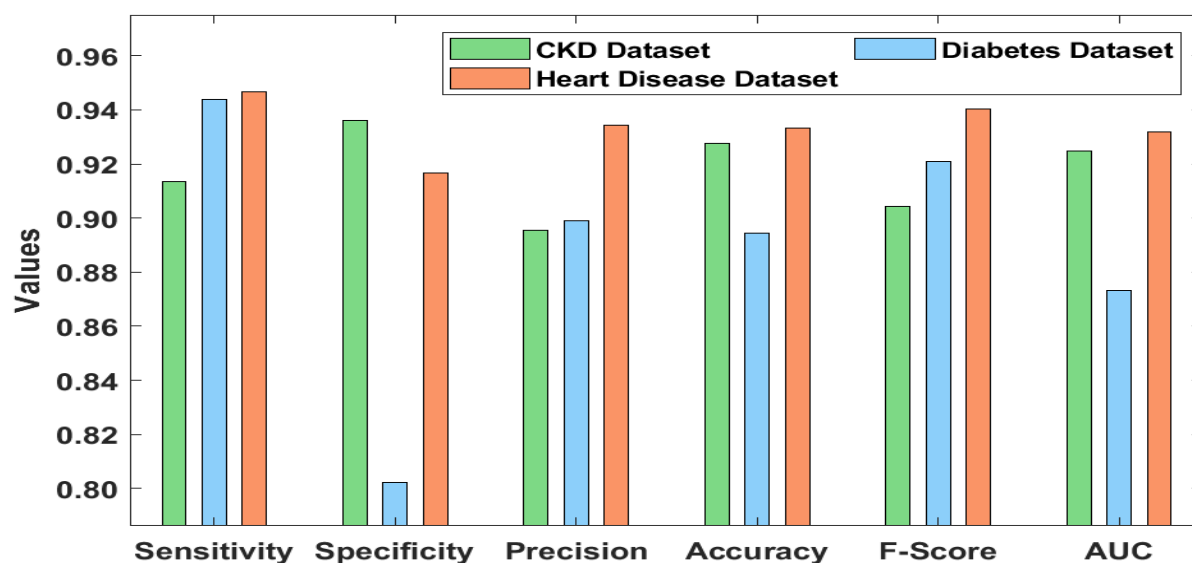


Fig. 6. Result analysis of AM-DNN model

A comparative results analysis of the AM-DNN model with state of art methods takes place on CKD dataset is depicted in Table 3 and Fig. 7. From the outcome, it is evident that the Olex-GA model is appeared as the least performer by achieving a lower sensitivity of 0.8, specificity of 0.666, and accuracy of 0.75. Besides, the LR model has resulted in a slightly enhanced classification outcome over the Olex-GA model with the sensitivity of 0.83, specificity of 0.82, and accuracy of 0.82. Accordingly, the XGBoost model has retained a

moderate result with a sensitivity of 0.83, specificity of 0.83, and accuracy of 0.83. Followed by, the PSO algorithm has surpassed the earlier ones with the sensitivity of 0.88, specificity of 0.8461, and accuracy of 0.875. Moreover, the ACO algorithm has accomplished manageable results with a sensitivity of 0.8, specificity of 0.666, and accuracy of 0.75. Eventually, the DT model has tried to portray near optimum performance with the sensitivity of 0.9038, specificity of 0.8928, and accuracy of 0.9. But the presented AM-DNN model has resulted in a maximum sensitivity of 0.9133, specificity of 0.936, and accuracy of 0.9275.

Table 3 Classification results analysis of existing with proposed AM-DNN Model on CKD

Dataset

Methods	Sensitivity	Specificity	Accuracy
AM-DNN	0.9133	0.9360	0.9275
Decision Tree	0.9038	0.8928	0.9000
ACO	0.8888	0.8461	0.8750
PSO	0.8800	0.8000	0.8500
XGBoost	0.8300	0.8300	0.8300
Logistic Regression	0.8300	0.8200	0.8200
OlexGA	0.8000	0.6666	0.7500

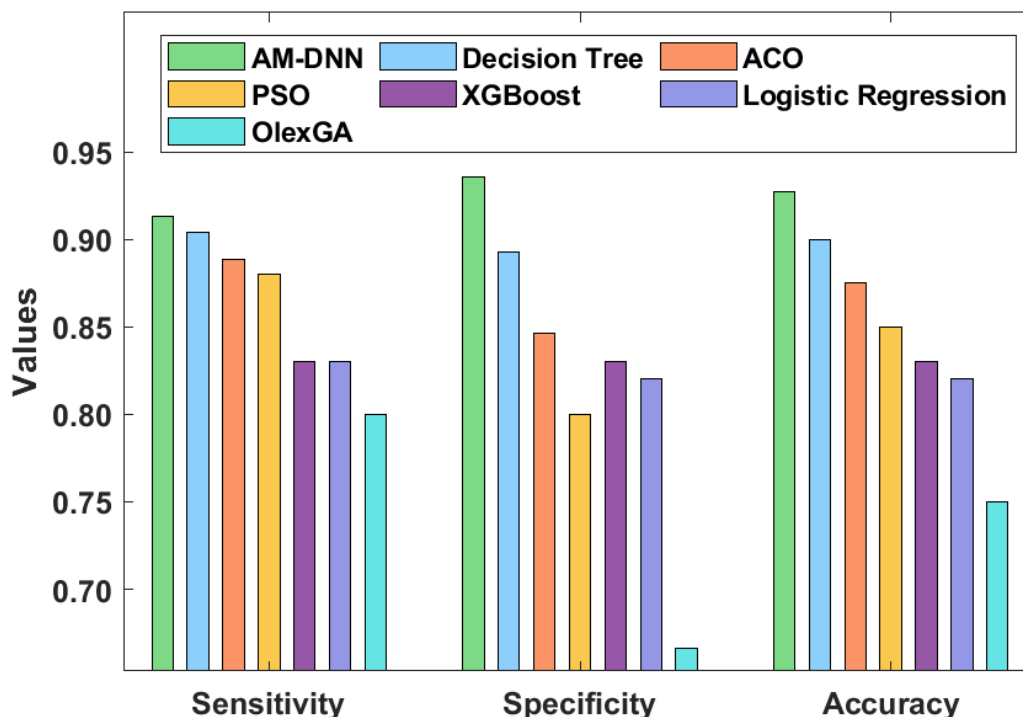


Fig.7. Comparative analysis of AM-DNN model on CKD dataset

A comparative results analysis of the AM-DNN method with state of art models takes place on Diabetes dataset is illustrated in Table 4 and Fig. 8. From the outcome, it is apparent that the DT model is appeared as a minimum performer by achieving a low precision of 0.8140 and sensitivity of 0.7902. Also, the Voted perceptron scheme has resulted in moderate classification result over the DT model with a precision of 0.844 and sensitivity of 0.6804. Followed by, the LogitBoost model has maintained a considerable result with a precision of 0.8460 and sensitivity of 0.7761. Next, the GBT technology has performed well the previous ones with a precision of 0.87 and sensitivity of 0.9089. Additionally, the LR approach has attained moderate outcome with a precision of 0.88 and sensitivity of 0.7927. However, the newly presented AM-DNN approach has provided a higher precision of 0.899 and sensitivity of 0.9440.

Table 4 Classification results analysis of existing with proposed AM-DNN Model on Diabetes Dataset

Methods	Precision	Sensitivity	Accuracy	F-score
AM-DNN	0.8990	0.9440	0.8945	0.9210
GBT	0.8700	0.9089	0.8867	0.9134
LR	0.8800	0.7927	0.7721	0.8341
Voted Perceptron	0.8440	0.6804	0.6679	0.7837
LogitBoost	0.8460	0.7761	0.7408	0.8095
DT	0.8140	0.7902	0.7382	0.8019

From the figure, it can be clear that the Voted perceptron model is considered to be poor performer by accomplishing least accuracy of 0.6679 and F-score of 0.7837. On the other hand, the DT model has gained acceptable classification outcome than the Voted perceptron model with an accuracy of 0.7382 and F-score of 0.8019. Next, the LogitBoost technology has retained a reasonable outcome with an accuracy of 0.7408 and F-score of 0.8095. Afterward, the LR mechanism has outperformed the previous ones with an accuracy of 0.7721 and F-score of 0.8341. In addition, the GBT scheme has attained reliable results with an accuracy of 0.8867 and F-score of 0.9134. Therefore, the newly developed AM-DNN technology has resulted in supreme accuracy of 0.8945 and F-score of 0.9210.

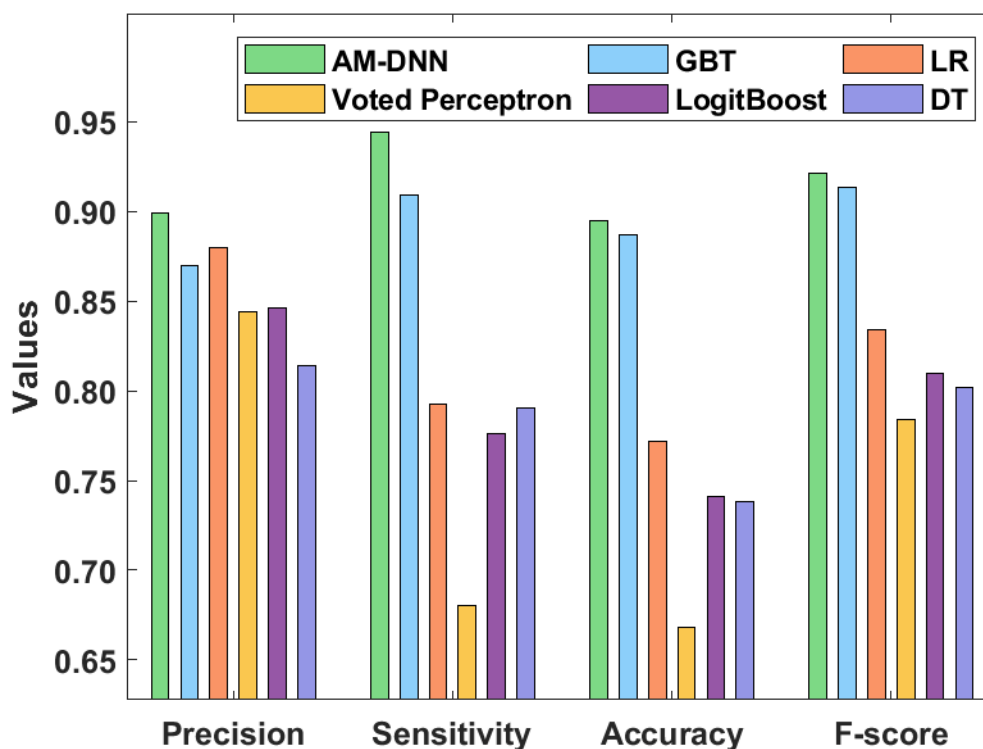


Fig. 8. Comparative analysis of AM-DNN model on Diabetes dataset

A comparative results analysis of the AM-DNN method with state of art methods takes place on Heart disease dataset is illustrated in Table 5 and Fig. 9. From the outcome, it is eminent that the RT approach is referred as poor performer by accomplishing minimum sensitivity of 0.7295, specificity of 0.7905, and precision of 0.7416. Additionally, the J48 framework has resulted from a moderate classification outcome when compared with RT model with the sensitivity of 0.7394, specificity of 0.7881, and precision of 0.7333. Followed by, the NBTree technology has reserved a considerable outcome with the sensitivity of 0.7964, specificity of 0.8089, and precision of 0.75. Afterward, the RF algorithm has achieved well than classical models with the sensitivity of 0.8034, specificity of 0.8300, and precision of 0.7833. Furthermore, the RBFNetwork technique has attained reasonable outcome with a sensitivity of 0.8291, specificity of 0.8497, and precision of 0.8033. However, the projected AM-DNN model has ended up with high sensitivity of 0.9467, specificity of 0.9167, and precision of 0.9342.

From the figure, it can be clear that the RT model is referred as least performer by gaining low accuracy of 0.7629 and F-score of 0.7355. Then, the J48 approach has provided moderate classification outcome over the RT mechanism with an accuracy of 0.7666 and F-score of 0.7364. Then, the NBTree approach has conserved a considerable outcome with an accuracy

of 0.8037 and F-score of 0.7725. On the other side, the RF algorithm has outperformed the previous ones with an accuracy of 0.8185 and F-score of 0.7932. Moreover, the RBFNetwork algorithm has attained manageable results with an accuracy of 0.8407 and F-score of 0.8186. But the presented AM-DNN method has offered a superior accuracy of 0.9333 and F-score of 0.9404.

Table 5 Classification results analysis of existing with proposed AM-DNN Model on Heart Disease Dataset

Methods	Sensitivity	Specificity	Precision	Accuracy	F-Score
AM-DNN	0.9467	0.9167	0.9342	0.9333	0.9404
J48	0.7394	0.7881	0.7333	0.7666	0.7364
Random Tree	0.7295	0.7905	0.7416	0.7629	0.7355
RBFNetwork	0.8291	0.8497	0.8033	0.8407	0.8186
NBTree	0.7964	0.8089	0.7500	0.8037	0.7725
Random Forest	0.8034	0.8300	0.7833	0.8185	0.7932

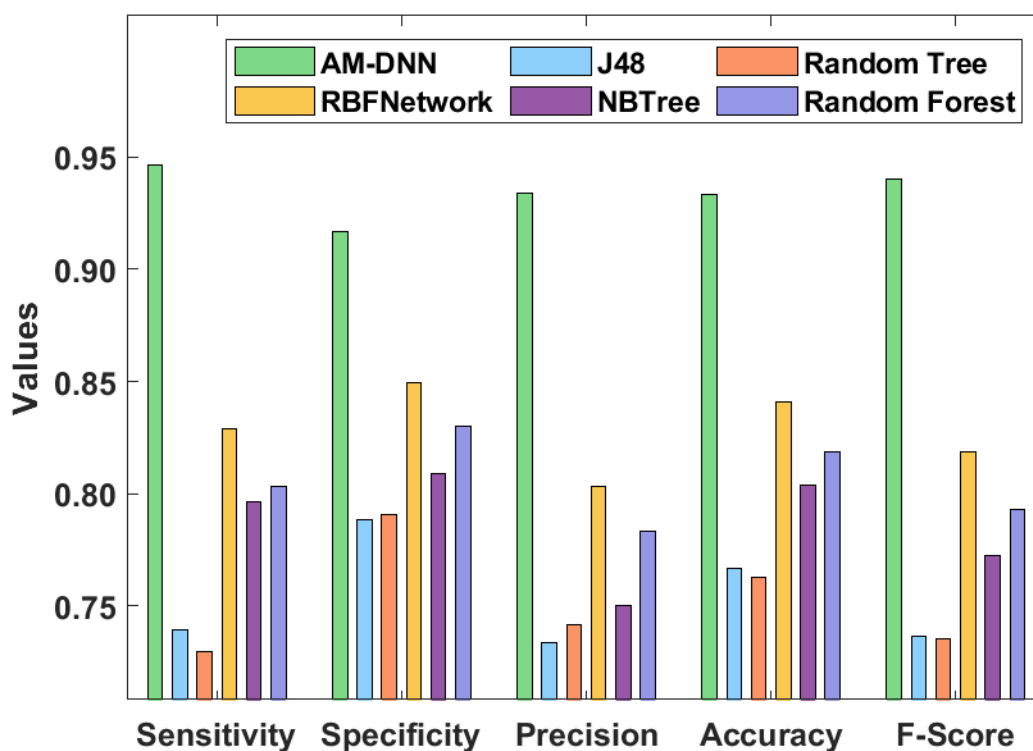


Fig. 9. Comparative analysis of AM-DNN model on Heart disease dataset

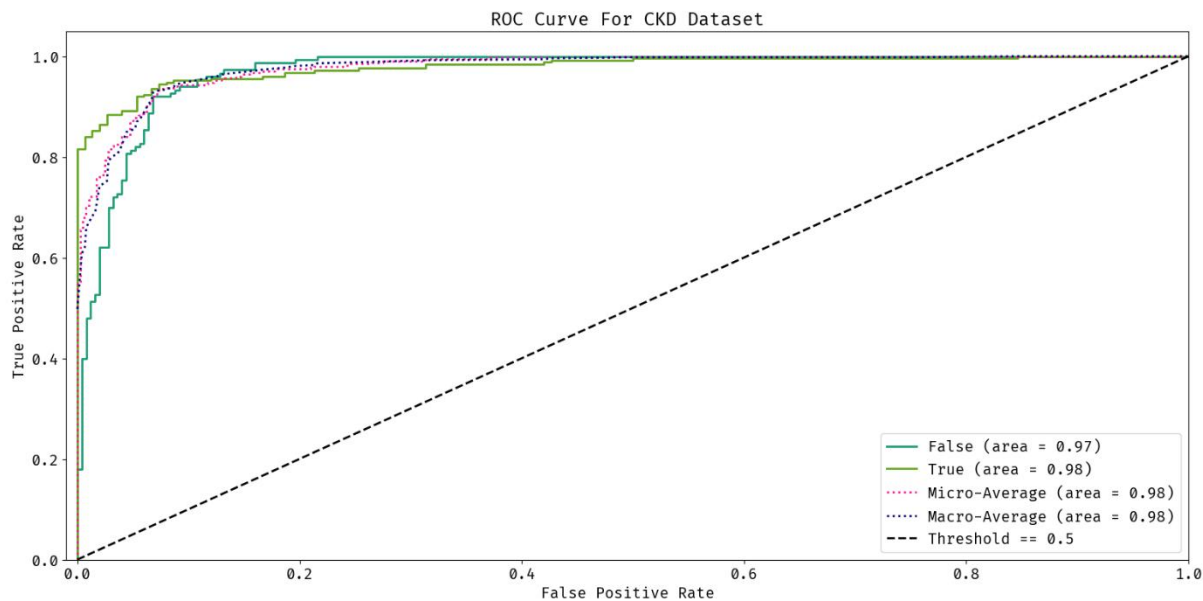


Fig. 10.ROC Analysis of CKD Dataset

Fig. 10 depicts the ROC analysis of the AM-DNN model on the applied CKD dataset. The figure depicted that the AM-DNN model has achieved an effective outcome by attaining a maximum ROC of 0.98.

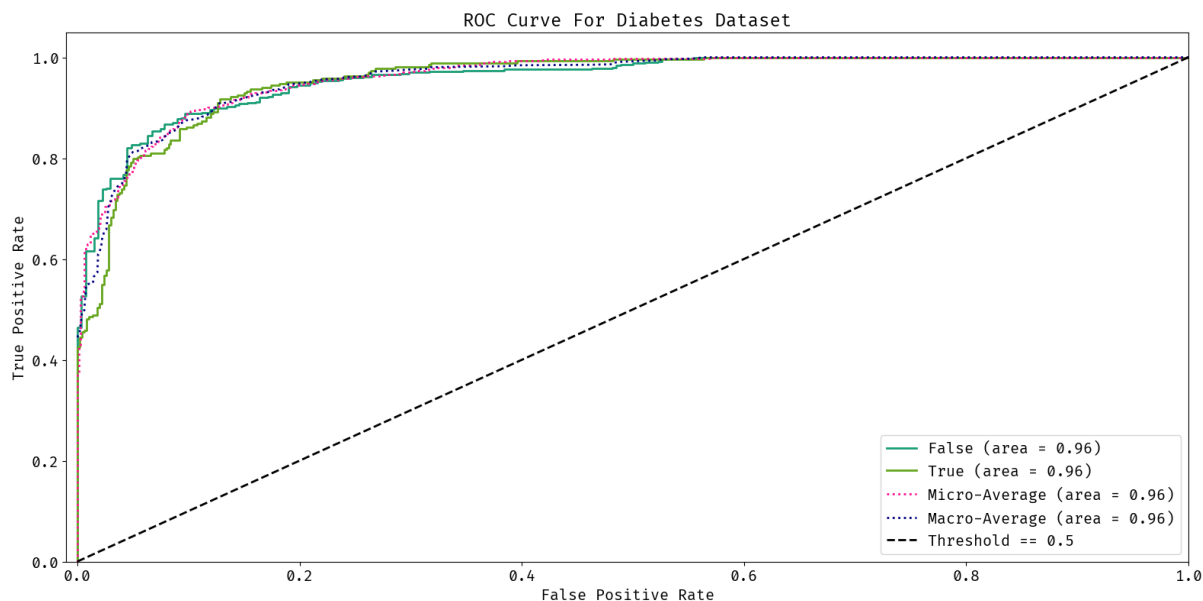


Fig. 11.ROC Analysis of Diabetes Dataset

Fig. 11 demonstrated the ROC analysis of the AM-DNN method on the given Diabetes dataset. The figure portrayed that the AM-DNN approach has accomplished efficient results by achieving a maximum ROC of 0.96.

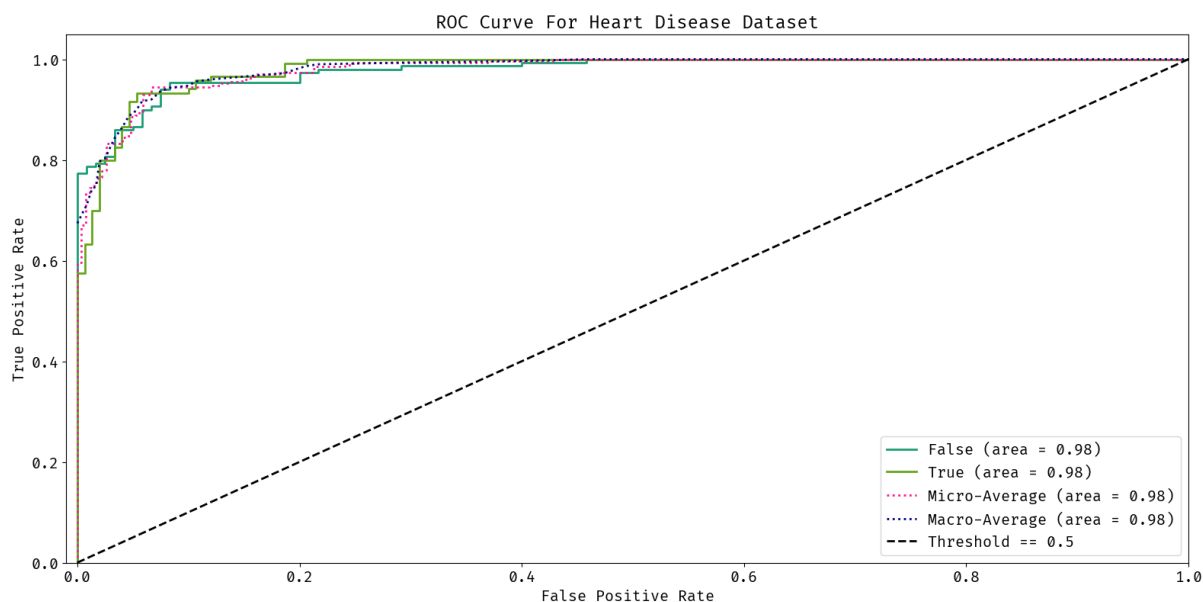


Fig. 12. ROC Analysis of Heart Disease Dataset

Fig. 12 projected the ROC analysis of the AM-DNN approach on the applied Heart disease dataset. The figure illustrated that the AM-DNN model has gained productive results by reaching high ROC of 0.98.

4. Conclusion

This paper has presented a new AM-DNN method for medical data classification. The presented AM-DNN model comprises different processes namely preprocessing, classification, and parameter optimization. Primarily, the input medical data is initially preprocessed to improve the data quality. Followed by, the preprocessed data is fed into the DNN based classification module to determine the appropriate class labels. At the same time, the AM optimizer is applied to fine-tune the parameters of the DNN model. The application of AM helps to improvise the efficiency of the DNN model. For assessing the simulation performance of the AM-DNN model, a series of experiments were performed. The obtained outcomes make sure that the AM-DNN method has resulted in a maximum accuracy of 0.9275, 0.8945, and 0.9333 on the applied chronic kidney disease (CKD), diabetes, and heart disease datasets.

References

- [1] Thong NT (2015) Intuitionistic fuzzy recommender systems: an efective tool for medical diagnosis. Knowl Based Syst 74:133–150

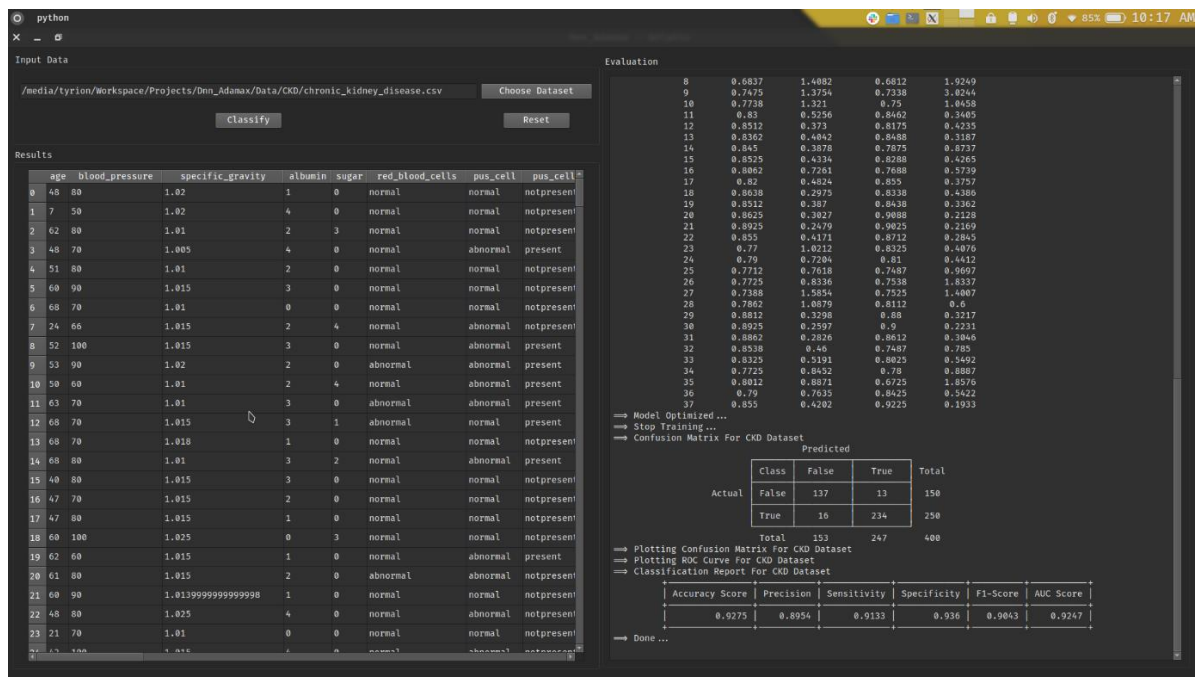
- [2] Wójtowicz A, Patryk Ż, Anna S, Krzysztof D (2016) Solving the problem of incomplete data in medical diagnosis via interval modeling. *Appl Soft Comput* 47:424–437
- [3] AlMuhaideb S, Menai ME (2016) An individualized preprocessing for medical data classification. *ProcedComputSci* 82:35–42
- [4] Leonard M, O’Connell H, Williams O, Awan F, Exton C, O’Connor M, Adamis D, Dunne C, Cullen W, Meagher DJ (2016) Attention, vigilance and visuospatial function in hospitalized elderly medical patients: Relationship to neurocognitive diagnosis. *J Psychosom Res* 90:84–90
- [5] Kurzyński M, Majak M, Żołnierak A (2016) Multiclassifier systems applied to the computer-aided sequential medical diagnosis. *Biocybern Biomed Eng* 36(4):619–625
- [6] Gorzałczany MB, Rudziński F (2016) Interpretable and accurate medical data classification-a multi-objective genetic-fuzzy optimization approach. *Expert Syst Appl* 71:26–39
- [7] Yang H, Chen Y-P (2015) Data mining in lung cancer pathologic staging diagnosis: correlation between clinical and pathology information. *Expert Syst Appl* 42(15):6168–6176
- [8] Shen L, Chen H, Zhe Y, Kang W, Zhang B, Li H, Yang B, Liu D (2016) Evolving support vector machines using fruit fly optimization for medical data classification. *Knowl Based Syst* 96:61–75
- [9] Castellano NN, Gazquez JA, García RM, Salvador A-E, Fernandez-Ros M, Manzano-Agugliaro F (2015) Design of a real-time emergency telemedicine system for remote medical diagnosis. *BiosystEng* 138:23–32
- [10] Adeli E, Feng S, Le A, Chong-Yaw W, Guorong W, Tao W, Dinggang S (2016) Joint feature-sample selection and robust diagnosis of Parkinson’s disease from MRI data. *Neuro Image* 141:206–219
- [11] Garcia-Hernandez JJ, Gomez-Flores W, Rubio-Loyola J (2016) Analysis of the impact of digital watermarking on computer-aided diagnosis in medical imaging. *ComputBiol Med* 68:37–48
- [12] Prabukumar M, Agilandeewari L, Ganesan K (2019) An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. *J AmbIntell Hum Comput* 10:267–293

- [13] Ibrahim RA, Ahmed A, Ewees DO, Elaziz MA, Songfeng Lu (2019) Improved salp swarm algorithm based on particle swarm optimization for feature selection. *J Ambient Intell Hum Comput* 10:3155–3169
- [14] Pan, R., Yang, T., Cao, J., Lu, K. and Zhang, Z., 2015. Missing data imputation by K nearest neighbours based on grey relational structure and mutual information. *Applied Intelligence*, 43(3), pp.614-632
- [15] Badem, H., Basturk, A., Caliskan, A. and Yuksel, M.E., 2017. A new efficient training strategy for deep neural networks by hybridization of artificial bee colony and limited–memory BFGS optimization algorithms. *Neurocomputing*, 266, pp.506-526.
- [16] Yi, D., Ahn, J. and Ji, S., 2020. An Effective Optimization Method for Machine Learning Based on ADAM. *Applied Sciences*, 10(3), p.1073.
- [17] https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease
- [18] <https://www.kaggle.com/uciml/pima-indians-diabetes-database>
- [19] [http://archive.ics.uci.edu/ml/datasets/statlog+\(heart\)](http://archive.ics.uci.edu/ml/datasets/statlog+(heart))
- [20] Irina ValeryevnaPustokhina, Denis Alexandrovich Pustokhin, Deepak Gupta, Ashish Khanna, K. Shankar, GiaNhu Nguyen, “An Effective Training Scheme for Deep Neural Network in Edge Computing Enabled Internet of Medical Things (IoMT) Systems”, *IEEE Access*, Volume. 8, Issue. 1, Page(s): 107112-107123, December 2020.
- [21] Lakshmanaprabu S.K, SachiNandanMohanty, Sheeba Rani S, SujathaKrishnamoorthy, Uthayakumar J, K. Shankar, “Online clinical decision support system using optimal deep neural networks”, *Applied Soft Computing*, Volume 81, Page(s): 1-10, August 2019.
- [22] Lakshmanaprabu S.K, SachiNandanMohanty, K. Shankar, Arunkumar N, Gustavo Ramireze, “Optimal deep learning model for classification of lung cancer on CT images”, *Future Generation Computer Systems*, Volume 92, Pages 374-382, March 2019.
- [23] Denis A. Pustokhin, Irina V. Pustokhina, Phuoc Nguyen Dinh, Son Van Phan, GiaNhu Nguyen, Gyanendra Prasad Joshi & Shankar K. (2020) An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19, *Journal of Applied Statistics*, DOI: 10.1080/02664763.2020.1849057
- [24] Le, DN., Parvathy, V.S., Gupta, D. et al. IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and

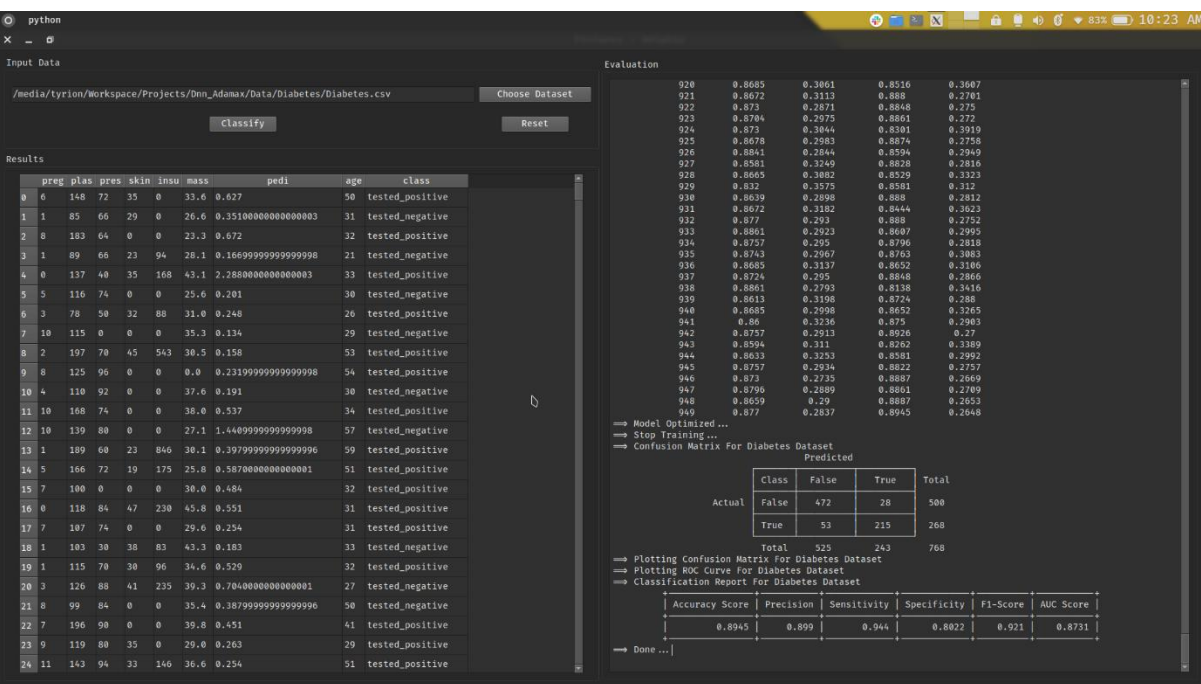
classification. Int. J. Mach. Learn. & Cyber. (2021). <https://doi.org/10.1007/s13042-020-01248-7>

[25] Shankar, K., Perumal, E. A novel hand-crafted with deep learning features based fusion model for COVID-19 diagnosis and classification using chest X-ray images. Complex Intell. Syst. (2020). <https://doi.org/10.1007/s40747-020-00216-6>

Appendix-I CKD Dataset



Appendix-II Diabetes Dataset



Appendix-III Heart Disease Dataset

