A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things

Mehdi Hosseinzadeh[1,2] · Omed Hassan Ahmed[3] · Marwan Yassin Ghafour[4] · Fatemeh Safara[5] · Hawkar kamaran hama[6] · Saqib Ali[7] · Bay Vo[8] · Hsiu-Sen Chiang[9]

Abstract

Medical information systems such as Internet of Medical Things (IoMT) are gained special attention over recent years. X-ray and MRI images are important sources of information to be examined for a particular type of anomalies. Reports based on the images and laboratory examination results could be mined with machine learning techniques as well. Thyroid disease diagnosis is an important capability of medical information systems. The main objective of this study is to improve the diagnosis accuracy of thyroid diseases from semantic reports and examination results using artificial neural network (ANN) in IoMT systems. In order to improve generalization and avoid over-fitting of ANN during the training process, a set of multiple multilayer perceptron (MMLP) neural network with the back-propagation error ability is proposed in this paper. Moreover, an adaptive learning rate algorithm is used to deal with the slow convergence and the local minima problem of the back-propagation error algorithm. The proposed MMLP significantly increased the overall accuracy of thyroid disease classification. With MMLP with a set of 6 networks, an improvement of 0.7% accuracy is achieved compared to a single network. In addition, comparing to the standard back-propagation, by using an adaptive learning rate algorithm in the proposed MMLP, an improvement of 4.6% accuracy and the final accuracy of 99% have been obtained in IoMT systems. The proposed MMLP is compared to recent researches reported for thyroid disease diagnosis, and its superiority is shown.

Keywords Internet of medical things · Multiple multilayer perceptron · Adaptive learning rate · Back-propagation · Thyroid disease · Artificial neural network

✉ Bay Vo
  vd.bay@hutech.edu.vn

Extended author information available on the last page of the article

1 Introduction

Employing as Internet of Medical Things (IoMT) is on the rise in recent years. Diagnosis of thyroid disease is an important capability of medical information systems recently, because of the impact of thyroid on other human body organs [1]. The main objective of this study is to improve the diagnosis of thyroid diseases from the data collected from reports and examination results in IoMT. Artificial neural network (ANN) is one of the widely used machine learning techniques, in particular for medical systems. Both examination results and reports on medical images are the valuable sources of information to be mined by ANN for diagnosis purposes. In an ANN, the back-propagation error is a supervised learning algorithm that employs a gradient descent method. It is one of the most commonly used algorithms for training a multilayer perceptron (MLP) neural network [2]. Despite all the successes of the MLP with back-propagation, there are still some drawbacks including slow convergence and local minima problem in IoMT [3]. The convergence behavior is highly dependent on tuning the learning rate parameter of the back-propagation; with a large learning rate, the optimal solution may skip and the back-propagation may never converge at all. On the contrary, with a small learning rate the convergence takes a long time and there is a high risk to get trapped in local minima [4].

The impact of convergence behavior on neural network performance has been confirmed in previous researches [2–6]. In this study, we employ a set of multiple MLP (MMLP) neural network to classify thyroid disease dataset into three categories including: normal, hypothyroid, and hyperthyroid. In the proposed MMLP, we exploit back-propagation error to improve the network generalization behavior. An adaptive learning rate algorithm is used to improve the convergence rate of the back-propagation as well. Contributions of this paper can be summarized as follows:

- To diagnose three types of thyroid diseases using Multiple MLP.
- To improve the accuracy of classifying features selected from the row dataset.
- To use adaptive learning to improve the performance of ANNs.
- To achieve better results than previous researches using the proposed MMLP.

The reminder of this paper is organized as follows. Section 2 presents a brief review of previous a number of researches related to thyroid disease and neural network. Description of thyroid disease dataset and the process of feature selection and feature scaling are presented in Sect. 3. Moreover, the determination of the neural network model, the implementation of MMLP and the adaptive learning algorithm are explained in the same section in IoMT. In Sect. 4, results obtained by applying the proposed method for classifying the thyroid disease dataset are presented. In Sect. 5, the results of the proposed method are compared with that of other related papers and discussed. Section 6 gives a conclusion and suggests future works.

2 Related work

In this section, we provide a brief review on previous researches on thyroid disease and classification algorithms in IoMT environment. ANN has gained researchers' attention in numerous fields of studies, in particular in medical engineering and diseases diagnosis applications. Several studies have been reported on the improvement of the feed-forward neural networks and the back-propagation convergence behavior. We review look over different classification algorithms employed in recent literature. Researches on various adaptive learning rate are presented as well. In addition, different datasets used in recent studies on thyroid disease are reviewed. Thyroid disease dataset of UCI machine learning repository was the common dataset which most of the previous researches on thyroid disease are reported on.

Several researches have been conducted on MLP neural network with different configurations and different learning rate algorithm. Hamid et al. [7] employed a MLP with standard BP. As learning rate algorithm, they used adaptive momentum and best accuracy of 95.72%. Gorunescu et al. [8] also proposed MLP and momentum learning algorithm with accuracy of 83.64%. Belciug et al. [9] examined several ANN including probabilistics neural network, and the best accuracy of 95.69% was obtained.

Sundaram and Renupriya [10] have achieved the accuracy of 95.8% and 97.8% on the same thyroid disease dataset using a feed-forward neural network with the back-propagation and the Elman neural network, respectively. Liang et al. [4] proposed a feature selection method based on the random forest to improve the classification accuracy of the MLP. The same thyroid disease dataset was used, and the accuracy rate of 95.69% was achieved. In another study by Liang et al. [11], after normalizing the thyroid disease dataset, they achieved 95.34% of accuracy.

Chen et al. [12] proposed using deep learning for diagnosis of malignant thyroid glands. The accuracy of 95% was gained with deep learning which was superior compared to SVM, random forest and ANN.

Alkhasawneh et al. [13] proposed a Hybrid Cascade Forward Neural Network with Elman Neural Network (HECFNN) to classify six different benchmark datasets. One of them is UCI thyroid disease dataset; the authors gain results of 96.10% accuracy for thyroid disease classification. Ahmed et al. [2] presented a new hybrid system for intelligent decision making on thyroid diseases. The presented system is a hybrid of three data mining methods including k-nearest neighbor (kNN), linear discriminant analysis (LDA), weighed preprocessing, and adaptive neuro-fuzzy inference system (ANFIS) for the diagnosis of thyroid disorders. LDA is used for dimension reduction, KNN is used for preprocessing and ANFIS is used for classification. UCI dataset is used for evaluating the method, and the best accuracy of 98.5% is reported. In a similar work by Ahmed et al. [2, 14] the same project has been conducted with SVM, ANFIS, and KNN. In [15], the authors presented a method for diagnosis of thyroid disease in pregnant. In their proposed method, deep learning algorithm is used and 98.22% accuracy is achieved.

Several studies have been reported on the improvement of the feed-forward neural networks and the back-propagation convergence behavior. Kolbusz et al. [5]

proposed a back-propagation algorithm with adaptive learning algorithm to estimate classification output error. Takase et al. [3] investigated the impact of decreasing an increase in learning rate on error rate of neural network training. They have been proved that both increasing and decreasing the learning rate could reduce the error rate of a neural network learning results.

Iranmanesh and Mahdavi [16] proposed a high-speed learning technique using differential adaptive learning rate (DALRM) comparing to the Nguyen-widow weight initialization, the standard, and the optical back-propagation. An improvement of the back-propagation in the neighborhood-based feed-forward neural network has been proposed by [17] using adaptive learning parameter to accelerate the convergence and avoid local minima. Nawi et al. [18] proposed a method to improve the convergence behavior using the adaptive gain parameter of the transfer function on the classification of thyroid disease.

Jha et al. [19] published a comprehensive research on 32 classification methods used diagnosing 17 different diseases including thyroid diseases. The average accuracy of 79.92% is achieved. An image-based system is proposed based on ultrasound thyroid images [20]. The system is designed to diagnose thyroid nodules, where frequency domain and spatial domain information is employed. Convolutional neural network-based method is used to differentiate benign and malignant nodules. The accuracy of 90.88% is achieved by the system proposed. Such systems could be used in IOT-based healthcare system. An article is published on different issues of fog computing in healthcare systems by Al-khafajiy et al. [21]. A similar work was published by Fahim et al. [22] which is a smartphone application to enhance active lifestyle.

A parallel feed-forward neural network approach with a rule-based system using the back-propagation algorithm was proposed for prediction of Parkinson disease and they achieved an improvement of 8.4% with a set of 9 parallel networks [23]. In their proposed method, unknown data of each network is used in the training set of the next network. Zeiler [24] proposed a new adaptive learning rate (ADADELTA) based on the first-order information, which does not require manual adjustment and is sensitive to hyperparameters. In [25], authors proposed a stochastic gradient decent algorithm to optimize the accuracy of ANN and logistic regression. Different learning rates, such as Adadelta, Momentum, Adam, are examined, and the best accuracy of 98.6% was achieved through GAdam learning rate algorithm. Overall, in the methods reported for thyroid disease diagnosis with various classification techniques and learning rate algorithms, the best accuracy of 98.6% is achieved that needs to be improved.

3 Proposed method

MLP is a successful and therefore popular machine learning algorithm. As reviewed in literature review section, several ANNs were used in previous researches for thyroid disease diagnosis. The best accuracy reported was 98.6% [21] and therefore, an improvement is still necessary.

We examined an MLP with different configuration to improve the accuracy of thyroid disease detection which was reported before. However, desired accuracy was not achieved with a single MLP. Therefore, we tried to investigate the power of multiple MLP (MMLP) for the same goal. After training a set of multiple networks individually, each network is tested on the evaluation set with two measures: MSE of each individual MLP and average MSE for different combinations of MLPs. As will be shown later in this section, the accuracy is increased and MSE of validation set is decreased gradually by training more networks. However, the improvements of the networks will not improve after adding a certain number of networks. The best accuracy was obtained by 6 MLPs. The structure of the MMLP is illustrated in Fig. 1.

One of the important issues with MLP is its convergence rate. The impact of convergence behavior on neural network performance has been confirmed in previous researches [2–6]. In this paper, to employ an adaptive learning rate has been proposed to increase the generalization and convergence rate of the proposed MMLP in IoMT environment.

Therefore, an adaptive learning rate is used to overcome this problem; the value of the learning rate is changed in each epoch during the learning process. All the experiment steps are performed on the thyroid disease dataset taken from the UCI machine learning repository, which is a widely used benchmark dataset. The overview of the experiment processes is presented in Fig. 2.

3.1 Dataset

The dataset used for the experiment is a thyroid disease dataset taken from the UCI machine learning repository obtained by the Garvan institute [26]. The dataset
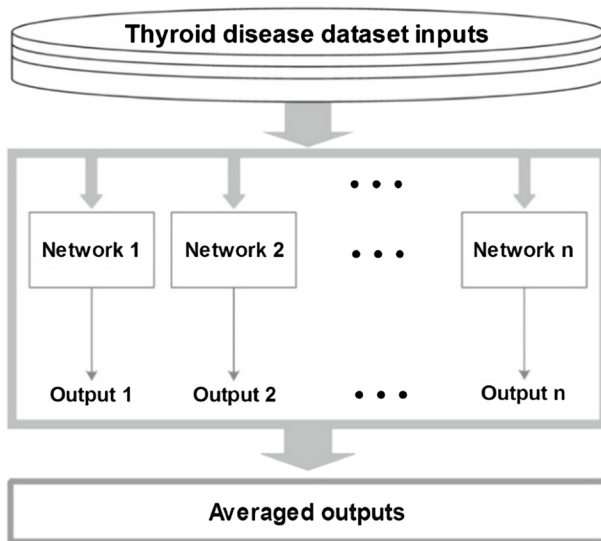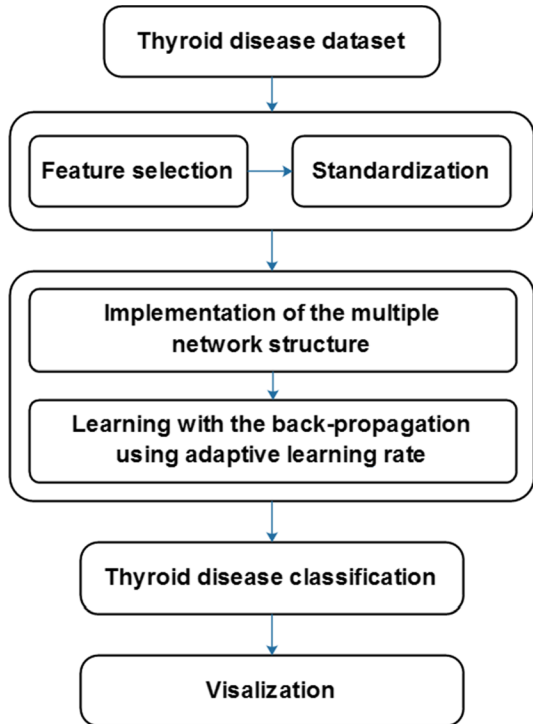

Fig. 1 The structure of the proposed MMLP

Fig. 2 Block diagram of the experiment processes

consists of 21 input variables of 7200 patients, with 15 binary and 6 continuous attributes without any missing data.

The target variables include 3 classes of patient thyroid gland function, which classifies a patient into a normal functioning, hyperthyroidism or hypothyroidism classes. From 7200 patients, 166 patients have normal thyroid function, 368 patients are hyperthyroid, and the reminder 6666 patients are hypothyroid. Table 1 shows the original thyroid disease dataset input and output attributes.

3.2 Feature selection

Feature selection, also known as attribute selection, is to select a subset of relevant features that affects the classification accuracy positively and to remove irrelevant variables [17]. In this study, different ranking methods of feature selection including Pearson's correlation, symmetrical uncertainty, one-R classifier and relief attribute evaluation have been used to select the most relevant features to improve the overall classification performance. In these methods, attributes are listed in order of their importance using the evaluation of their worth with respect to the class. After applying them, a score is calculated for each attribute based on its rank in each method. Then, a threshold is adjusted using these scores for selecting the top ones with high scores. The threshold is obtained by trial and error. Different thresholds are considered, and classification results are obtained.

Table 1 Dataset attributes description [26]

Attribute	Domain
Age	[0.01, 0.97]
Sex	[0, 1]
On thyroxine	[0, 1]
Query on thyroxin	[0, 1]
On antithyroid medication	[0, 1]
Sick	[0, 1]
Pregnant	[0, 1]
Thyroid surgery	[0, 1]
I131 treatment	[0, 1]
Query hypothyroid	[0, 1]
Query hyperthyroid	[0, 1]
Lithium	[0, 1]
Goiter	[0, 1]
Tumor	[0, 1]
Hypopituitary	[0, 1]
Psych	[0, 1]
Thyroid stimulating hormone (TSH)	[0.0, 0.53]
Triiodothyronine (T3)	[0.0005, 0.18]
Thyroid transcription factor (TT4)	[0.0020, 0.6]
Thyroxine utilization (T4U)	[0.017, 0.233]
Free thyroxine index (FTI)	[0.0020, 0.642]
Class normal	[0, 1]
Class hyperthyroid	[0, 1]
Class hypothyroid	[0, 1]

By applying the threshold, the features were divided into 2 categories as shown in Fig. 3; those features that were scored higher than the pointed threshold (green bars) were selected and other features were discarded. By performing this process, selecting important and relevant features would be more reliable. Finally, 12 relevant features have been selected for this experiment in terms of the values of Pearson's correlation, because the features selected by Pearson's correlation method produced the best classification results. Also, those selected features of patients are the most important risk factors for thyroid disease according to scientific researches.

Thyroid disease occurs most frequently in elderly women [27]. However, aging increases the risk of thyroid problems in both men and women. Thyroid stimulating hormone (TSH) is a pituitary hormone that stimulates the thyroid gland to produce Thyroxine (T4) and then converted to Triiodothyronine (T3). T4 and T3 are hormones that help control the body's metabolism. T3 and T4 levels are usually increasing or decreasing together. The two common thyroid disorders mentioned earlier are hyperthyroidism (overactive thyroid gland) and hypothyroidism (underactive thyroid gland) [28]. A higher than normal level of T4 is
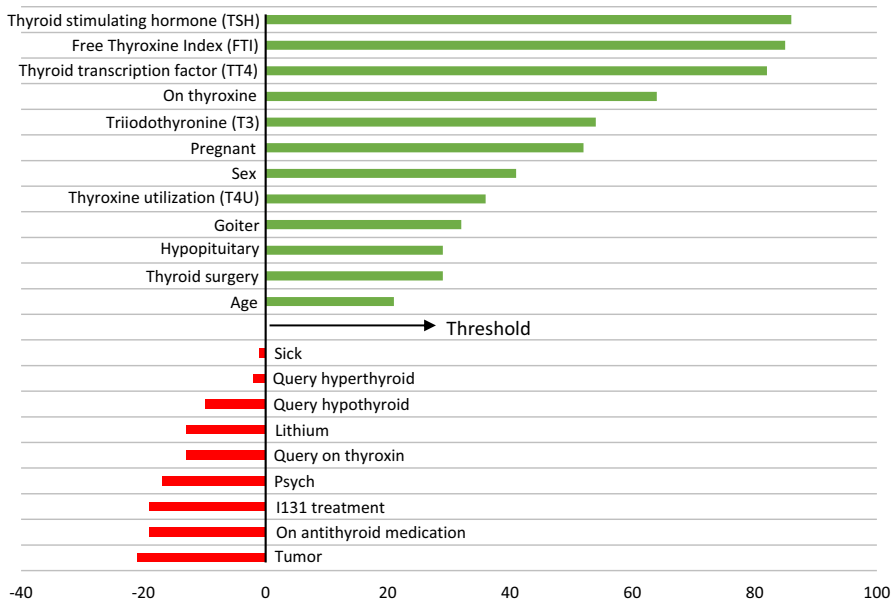
Fig. 3 The relative importance of features based on their scores to the pointed threshold

hyperthyroidism, and a lower than normal level of T4 is hypothyroidism. Also, during pregnancy high levels of T4 and T3 can exist.

3.3 Feature scaling

Feature scaling is an important data preparation step to rescale the value of features from dynamic range into a specific range. Standardization is a well-known method for scaling the value of features with different scales. In artificial neural network, standardizing the input vectors helps to have a fast and accurate convergence. It standardizes the value of features into a small specific range which can improve the accuracy of classification [29]. After the feature selection phase of this study, input variables have been standardized to map the values of each row in the data to have a mean of 0 and a standard deviation of 1.

3.4 Neural network classifier

In this paper, MLP has been used for the classification of thyroid disease which has a feed-forward structure, as illustrated in Fig. 4. This type of neural network consists of an input layer, one or more hidden layers and an output layer of interconnected neurons. This connection is between each neuron in one layer with every neuron on the next layer. The input signals are applied in a forward direction from the inputs to the outputs through the hidden neurons [30]. The number of neurons in the input layer is equal to the number of input variables of the dataset after the data
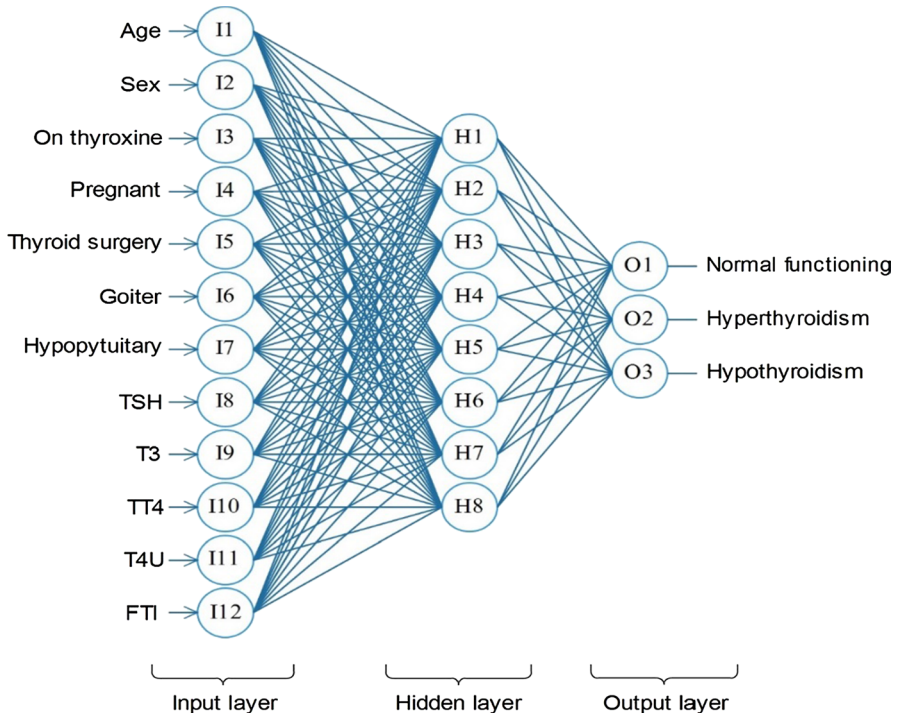
Fig. 4 The MLP structure with 12 inputs, 3 outputs, and 8 hidden neurons

preparation step, and the number of neurons in the output layer is equal to the number of classes in the dataset.

Feed-forward neural networks often have one or more hidden layers followed by an output layer of linear neurons. The hidden layer with a nonlinear transfer function allows the network to learn nonlinear and linear relationships between input and output vectors [31]. For determining the best type of the transfer functions for MMLP of this study, different transfer function has been compared, and hyperbolic tangent sigmoid transfer function for both the hidden layer and output layer has been selected.

Determining the number of hidden layers and the number of neurons in each layer is the two main issues. In this study, the MLP neural network with one hidden layer has been used. Classification accuracy may have increased with more than one hidden layer; however, too many epochs for training the network will be required (over-training) which result in over-fitting [32]. Over-fitting occurs when the model is overly complex and learns too accurate the relationship between inputs and outputs training data. In this regard, the model loses its generalization power over the test data and the error will be increased [33].

After determining the number of hidden layers, a determination of the number of neurons in each layer must be made. Determining the number of hidden neurons depends on the number of neurons of the input and output layers. With a large number of hidden neurons, the over-fitting problems may occur. On the contrary, using a

small number of hidden neurons will result in the under-fitting problem. The under-fitting problem occurs when the model is too simple which results in poor prediction accuracy on both training and test data [34].

The optimal number of hidden neurons can be determined using some rules of thumb that was proposed by many researchers [35, 36]. For this study, with $N_h = \frac{2}{3}(N_i)$ from [37], where N_h is the number of hidden neurons and N_i is the number of input neurons; the optimal number of hidden neurons has been obtained, and the MLP neural network with 8 hidden neurons has been used.

3.5 The proposed multiple MLP (MMLP)

In order to avoid over-fitting and improve the network generalization, using a set of multiple MLP feed-forward neural network (MMLP) has been proposed in this study. In this method, finding a network with high generalization power is highly possible. The optimal number of networks of this study has been determined by the trial-and-error approach, and the MMLP with 6 neural networks is designed and implemented.

After training all networks, each network is tested on the evaluation set with both individual MSE and the average MSEs of the networks. After training all networks, their MSEs compared to the MSE of their average. In this experiment, average MSE of 6 ANNs was less than all previous combinations examined. In addition adding one more ANN also increased the averaged MSE. Therefore, the final number of ANNs to be combined is determined as 6.

For this study, the dataset is randomly divided into training, test and validation sets. Training the network is done with 70% of the dataset and the remaining 30% is used for the evaluation.

One of the most common learning models for MLP is the back-propagation supervised learning technique, which uses a gradient descent method to minimize the mean squared error between the desired outputs and the actual outputs for each input vector. Although the back-propagation has been useful in the classification of biomedical data, the learning process may take a long time to converge and there is a high possibility of converging into a local minimum which is not a global solution [21].

3.6 Adaptive learning rate

The convergence rate of the back-propagation is highly dependent on the selection of the learning rate which is a control parameter to control the step size for each weight update. The value of the learning rate should be a value between 0 and 1. A large value of the learning rate may skip the optimal solution, and the convergence may never be achieved. On the other hand, a small learning rate will increase the total time to converge to the optimal value or trap in local minima [38, 39]. Therefore, selecting the optimal value of the learning rate is a major challenge regarding the convergence rate. To avoid these problems, an adaptive learning algorithm has been used, which adapts the value of the learning rate during the training process.

The most common learning rate includes: time-based, exponential, and step-based learning rate. Time-based learning rate modifies the learning rate regarding to the learning rate of the earlier time iteration. Exponential learning rate is similar to step-based learning rate; however, as a substitute of steps a decreasing exponential function is included. Step-based learning rate sets a new learning rate based on pre-determined steps. In this paper we employed step-based learning rate. The decreasing formula is defined as:

$$\eta_n = \eta_0 d^{\text{floor}\left(\frac{1+n}{r}\right)} \tag{1}$$

where η_n is the learning rate at iteration n, and η_0 is the initial learning rate, d is how much the learning rate should change at each decrease (for example 0.5 corresponds to a splitting) and r corresponds to the decrease rate, or how often the rate should be decreased (10 corresponds to a decrease every 10 iterations). The *floor* function here decreases the value of its input to 0 for all values smaller than 1.

At each epoch, if the performance of the network decreases, the learning rate will increase by a constant parameter and vice versa. In this procedure, the learning rate is increased whenever a large value could result in stable learning. When the learning rate is too large to guarantee a decrease in error, it is decreased until stable learning resumes [40].

4 Results

The results obtained by the MMLP with adaptive back-propagation comparing to the standard back-propagation are presented in this section. As mentioned earlier, the dataset that has been used for the experiment is a thyroid disease dataset with 21 input variables and 7200 records taken from the UCI machine learning repository [26]. At the data preparation step of this study, 12 attributes that were the most effective one on the classification accuracy were selected, then their values were standardized into a specific range. The comparison of the performance before and after the data preparation step is shown in Table 2.

MMLP with an adaptive learning algorithm for the classification of thyroid disease has been used. In order to determine number of ANNs to be combined and construct MMLP, different numbers of ANN are examined. The results achieved with each of the combinations (NN1, NN2, …, NN6, indicating one network, two

Table 2 The comparison of the performance of MMLP before and after data preparation step

	Initial dataset	After feature selection	After feature scaling
Accuracy	92.6%	93.9%	99.0%
Averaged Performance (MSE)	0.0375	0.0376	0.0119
R-value	0.90	0.92	0.98

networks,…, and 6 networks, respectively) are shown in Fig. 5. The best accuracy and performance of this study were obtained by a set of 6 networks with 12 input, 3 output and 8 hidden neurons. The optimal number of hidden neurons was obtained by the $N_h = \frac{2}{3}(N_i)$ formula, where N_h is the number of hidden neurons and N_i is the number of input neurons.

As illustrated in Fig. 5, with a set of 6 networks with improved back-propagation the best accuracy and performance have been achieved. According to the figure, the primary networks have low accuracy and poor performance compared to the final networks.

The lowest accuracy corresponds to the single network with a rate of 98.3% and MSE of 0.0133, whereas the highest results related to the set of 6 networks with an accuracy rate of 99.0% and MSE of 0.0108. Moreover, the accuracy of the test set, validation set and the overall accuracy of the improved and the standard back-propagation are presented in Table 3.

With the standard back-propagation, the accuracy of 94.4% and MSE of 0.0357 were achieved. As shown in Tables 3 and 4, an improvement of 4.6% accuracy, 0.0249 of MSE and 0.06 of R-value were obtained with MMLP. The R-value is the correlation between the outputs and targets. If this value is closer to 1, it indicates an appropriate fitting. The results of MMLP (improved BP) and the standard
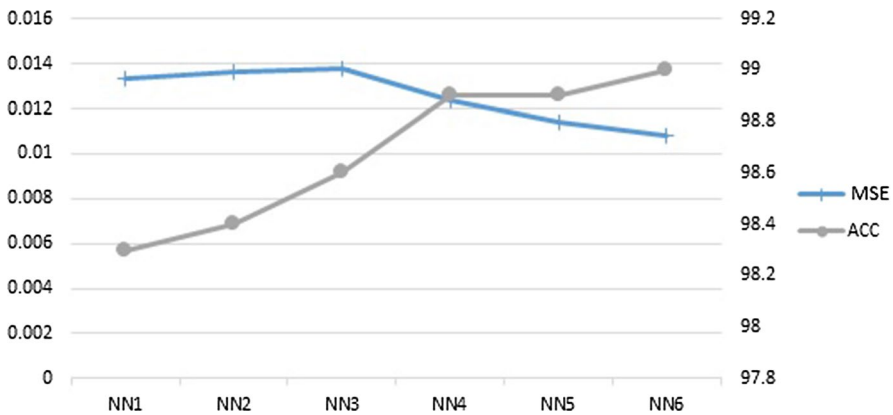

Fig. 5 The comparison of the different number of 6 networks regarding the accuracy and the MSE

Table 3 Accuracy of different sets for MMLP and the standard BP

	Training	Test	Validation	All
MMLP	99.2%	98.8%	98.6%	99.0%
Standard BP	94.5%	94.0%	94.3%	94.4%

Table 4 The MSE and the R-value of MMLP and the standard BP

	MSE	R-value
MMLP	0.0108	0.98
Standard BP	0.0357	0.92

back-propagation regarding the accuracy, MSE and *R*-value are proposed in Tables 3 and 4.

As explained in Sect. 3.1, the dataset is imbalanced as 6666 samples (92.6%) correspond to the hypothyroidism class. As the class distribution is imbalanced and the hypothyroidism class has much influence on accuracy than the other classes, the accuracy is not suitable enough for measuring the neural network performance. ROC (receiver operating characteristic) curve is another measure of performance that has been used in this study to compare the improved and the standard back-propagation, which is shown in Fig. 6. Whenever the ROC curve is closer to the upper left corner, the classification is more accurate regarding the high true positive rate (TPR or sensitivity). According to Fig. 6, the improved BP is highly accurate for all 3 classes, whereas the standard BP has a poor classification ability for all classes, especially for hyperthyroidism class.

As explained in Sect. 3.6, to avoid over-fitting the back-propagation with the adaptive learning rate has been used. The idea is the learning rate is increased whenever the new error is lower than the old error and it gets decreased until stable learning resumes. The value of the learning rate at epoch 818 equals 0.55, and the process of varying the learning rate is given in Fig. 7.

The comparison of MSE of the networks with respect to the learning rate is shown in Fig. 8. The best performance is related to MMLP with the set of 6 networks with the learning rate value of 0.55 according to Fig. 8. The performance has also examined with 7 and 8 networks; however, the results did not change considerably. Therefore, 6 networks are chosen as the final configuration.

As previously discussed, one major issue regarding the standard back-propagation is slow convergence. As shown in Fig. 9, the improved back-propagation using the adaptive learning rate saves the CPU time and the epoch cycles,
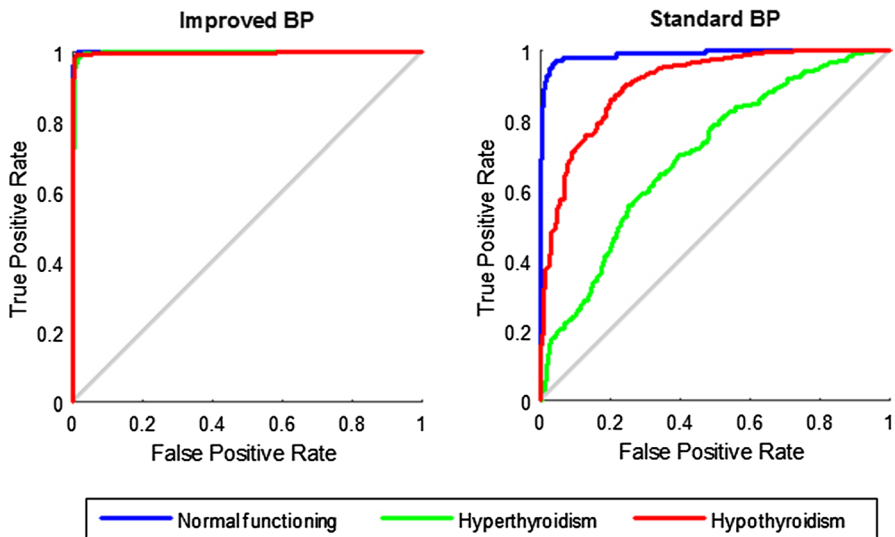

Fig. 6 Comparison of the ROC curve for MMLP (left) and the standard BP(right)
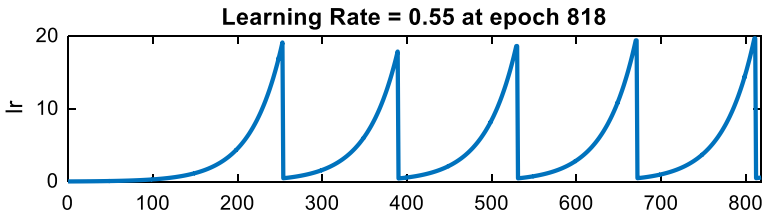
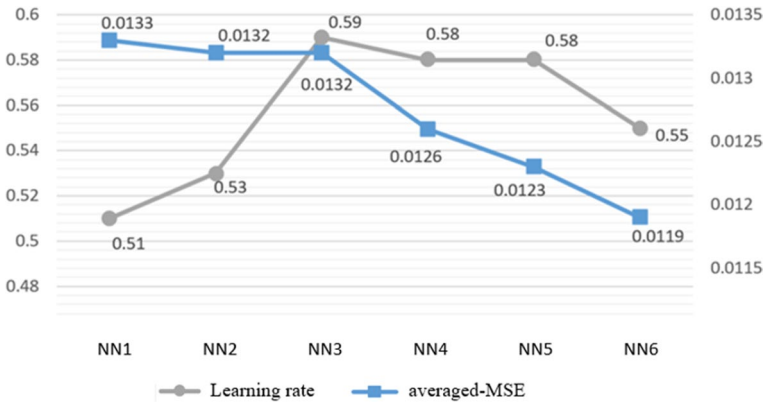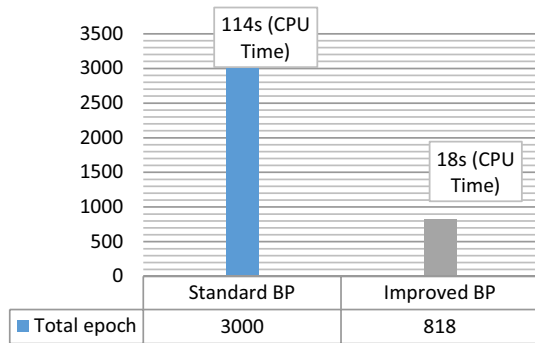**Fig. 7** The variation of learning rate during the MMLP training process

Fig. 8 The comparison of the effect of the different learning rate (Learning rate) values on averaged-MSE in the proposed MMLP

Fig. 9 The convergence time comparison for MMLP (the improved BP) and the standard BP

whereas the standard back-propagation uses all 3000 epochs, and there is a risk that convergence may not occur. Training the set of 6 networks using the standard back-propagation takes 114 s which takes 19 s for each network. MMLP, with

the set of 6 networks using an adaptive learning algorithm, trained in 18 s which takes almost 3 s for each network.

According to the results obtained from Fig. 10, the network errors have reduced significantly as regards the final MSE is small and the error of all sets decreased until epoch 818. With regard to the similar characteristics of the test set and the validation set error, it can be concluded that the learning process is well accomplished.

As shown in Fig. 10, over-fitting has not occurred as regards the test error did not increase before iteration 818 and the error of all sets reduced almost equally. Also, regarding the low training set error under-fitting has not occurred during the training process.

5 Discussion

To propose MMLP for improving the effectiveness of the MLP using a set of multiple networks and the back-propagation algorithm with an adaptive learning rate for classification of thyroid disease in IoMT environment is the main contribution of this study. The classification accuracy obtained by the proposed MMLP of this study and the other methods from the literature for the same dataset is summarized in Table 5. The thyroid disease dataset with 21 inputs and 7200 records was taken from the UCI machine learning repository [26], which includes three classes of patients: normal functioning, hyperthyroidism or hypothyroidism. First, the input attributes were reduced to 12 features using different ranking feature selection methods. The idea was finding the most relevant features using different methods and then selecting the most important of them. After that, the value of input attributes
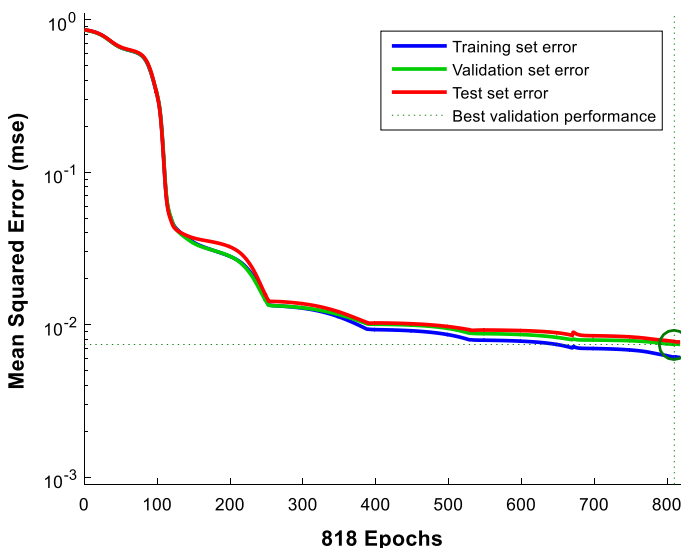

Fig. 10 Mean squared error (MSE) of the different sets using MMLP

Table 5 Classification accuracies obtained by different methods from the literature for the thyroid disease dataset

Study	Methods	Number of features	ACC (%)
[7]	MLP with standard BP	21	95.00
	The BP with adaptive gain The BP with adaptive momentum (with 21-5-3 architecture)		89.00
			95.72
[8]	MLP (after balancing dataset, with 12 hidden neurons, 0.4 and 0.9 initial learning rate and momentum)	21	83.64
	The BP/stimulus-sampling MLP		82.12
	Single hidden layer		
[41]	MLP (based on random forest feature selection)	2	95.69
	Single hidden layer		95.67
	RBF Network		93.75
	Probabilistic neural network		93.12
[11]	MLP	21	95.34
	Decision tree (after feature selection)		97.35
	Naïve Bayes		94.71
	RBF Network		94.27
[13]	Hybrid cascade forward neural Network with Elman neural Network	5	96.10
[25]	guided stochastic gradient descent	N/A	96.40
	GMomentum		96.80
	GAda-delta		93.00
	GRMSprop GAdam		98.40
			98.60
[5]	Elman neural network	21	97.80
	Feed-forward neural network (with different architecture)		95.80
[42]	Deep learning	N/A	95.00
	Four hidden layers		
	Compared to SVM, ANN, and Random forest		
[42]	MKSVM and IGWO	N/A	98.65
[43]	Multimodel ensemble (SVM, CART, DT, RF, GDBT)	11	98.64
[2]	LDA–kNN–ANFIS compared to C4.5, ANN, LVQ, and PNN	3	98.50
MMLP (proposed)	Multiple MLP neural network with the improved BP using an adaptive learning rate One hidden layer	21	99.20 (training)
			98.80 (test)
			98.80 (validation)
			99.00 (overall)

was standardized into a small specific range to done the training process faster and more accurate.

Table 5 presents a brief comparison of previous researches for improving the accuracy of thyroid disease diagnosis on UCI thyroid dataset. Belciug and

Gorunescu [57] proposed a feature selection method based on the random forests to improve the classification accuracy of the MLP. With their proposed method, only 2 attributes out of 21 attributes of the same thyroid disease dataset were selected and the accuracy rate of 95.69% was achieved. In another study [11], the authors after normalizing the thyroid disease dataset to be able to apply the MLP with the back-propagation, they achieved 95.34% of accuracy with 1500 epochs before feature selection. After detecting 3 irrelevant attributes and removing them, the accuracy with a rate of 94.71% was achieved by the MLP. According to Table 2 of this study, by selecting the most relevant features with the output classes and then standardized their values, the training process was done more accurately.

In [2] a new hybrid system is presented for intelligent decision making on thyroid diseases. The presented system is a hybrid of three data mining methods including *k*-nearest neighbor (kNN), linear discriminant analysis (LDA), weighed preprocessing, and adaptive neuro-fuzzy inference system (ANFIS) for the diagnosis of thyroid disorders. LDA is used for dimension reduction, KNN is used for preprocessing, and ANFIS is used for classification. UCI dataset is used for evaluating the method, and the best accuracy of 98.5% is reported. In a similar work by Ahmed et al. [2, 14], the same project has been conducted with SVM, ANFIS, and KNN.

Chen et al. [12] proposed using deep learning for diagnosis of malignant thyroid glands. The research was conducted on ultrasound reports. The accuracy of 95% was gained with deep learning which was superior in compare to SVM, random forest and ANN.

For modeling the neural network, the MLP with the back-propagation using the adaptive learning rate was used in this study. The number of hidden neurons was adjusted to 8 using $N_\mathrm{h} = {}^{2}\!/_{3}\left(N_\mathrm{i}\right)$ from [44], and the hyperbolic tangent sigmoid transfer function for both the hidden layer and the output layer was used. By adjusting these parameters, neither over-fitting nor under-fitting was occurred during the training process according to Fig. 10. Sundaram and Renupriya [10] have achieved the accuracy of 95.8% and 97.8% with the validation performance of 0.0267 and 0.0120 on the same thyroid disease dataset using a feed-forward neural network with the back-propagation and the Elman neural network, respectively. These results obtained by 2 hidden layers with 27 neurons in each layer with a soft max and a linear transfer function for the hidden layer and the output layer. According to the ROC plot, their neural network modeling has poor ability in the classification of hyperthyroidism and hypothyroidism classes compared to the normal functioning class.

The convergence behavior of the back-propagation is highly dependent on the selection of the learning rate; with a large learning rate value, the convergence may never be achieved and with a small learning rate value, the total epochs to converge to the optimal value will increase significantly and there is a risk to trap in local minima. To prevent these problems, the adaptive learning rate was used in this study. In this method, the learning rate will change during the training process. According to Fig. 9, by using the total 3000 epochs by the standard back-propagation there is no guarantee of convergence. Based on the same figure and Fig. 7, the improved back-propagation converged at epoch 818 in almost 3 s for each network with a learning rate of 0.55. Nawi et al. [18] proposed a method to change the gain parameter of the transfer function adaptively on classification problems including thyroid disease

dataset. They selected all 21 input attributes and 5 hidden neurons for the experiment. The standard gradient descent converged at epoch 3441 in 316.5 s, whereas the improved algorithm converged at epoch 1114 in 111.6 s. Hamid et al. proposed a method to improve the convergence behavior and local minima problem of the back-propagation using adaptive gain, adaptive momentum and adaptive learning rate [7]. The number of hidden neurons adjusted to 5 with 21 input variables and a 0.5 initial learning rate. The standard back-propagation used all 10,000 epochs in 1427.1 s with an accuracy rate of 95%, whereas the back-propagation with adaptive gain converged at epoch 1115 in 134.9 s with 89% of accuracy. The best results obtained by the back-propagation with adaptive gain, adaptive momentum and adaptive learning rate with 935 epoch of convergence in 10.5 s and accuracy of 95.72%. Hamid et al. [7] achieved an accuracy of 95%, 86.38% and 95.73% with the standard back-propagation, the back-propagation with adaptive gain and the back-propagation with adaptive momentum, respectively. The 21 input variables of the thyroid disease dataset with 5 hidden neurons were selected for the experiment. The standard back-propagation used all 3000 epochs in 145 s, whereas the back-propagation with adaptive gain converged at epoch 960 in 48 s. The back-propagation with adaptive momentum converged at epoch 1201 in 692.5 s.

The implementation of a multiple neural network approach was another method used in this study to improve the accuracy, generalization power and convergence. The number of networks obtained by trial-and-error and a set of 6 networks was trained to average their outputs. Related studies have been reported using a set of multiple neural networks for the classification of biomedical data. For instance, we proposed a parallel feed-forward neural network approach to predict Parkinson's disease. They achieved an improvement of 8.4% with a set of 9 parallel networks. In their proposed method, un-learned data of each network were used in the training set of the next network. As illustrated in Fig. 5 and Fig. 8 of this study, the best accuracy was obtained by the set of 6 networks with the rate of 99%.

6 Conclusion

Thyroid disease information, as an important part of the IoMT system, is studied in this paper. The MMLP is proposed which is a set of multiple MLP neural networks with the back-propagation learning algorithm using an adaptive learning rate. The MMLP is used for thyroid disease classification of the UCI thyroid disease dataset. As demonstrated earlier in this paper, the proposed method increased the overall accuracy and decreased overall MSE. To improve the network generalization and avoid over-fitting, a system with a set of 6 networks and the adaptive learning algorithms were used in the proposed MMLP, in which the accuracy rate of 99% was achieved. Moreover, the improved back-propagation converged more quickly compared to the standard back-propagation. According to the results, the proposed system has a remarkable ability to deal with over-fitting during the training process. Future works will focus on providing an effective method to find the global minima of the error function to guarantee convergence in the presence of local minima in IoMT environment. In addition, formal methods could

be examined to prove the effectiveness of the proposed method using mathematical modeling. Real thyroid dataset would be employed to examine the proposed method and confirm the capability of it to diagnose thyroid disease.

References

1. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M (Aug 2017) Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. J Digit Imag 30:477–486
2. Ahmad W, Ahmad A, Lu C, Khoso BA, Huang L (2018) A novel hybrid decision support system for thyroid disease forecasting. Soft Comput 22:5377–5383
3. Takase T, Oyama S, Kurihara M (May 2018) Effective neural network training with adaptive learning rate based on training loss. Neural Netw 101:68–78
4. Liang J, Xu Y, Bao C, Quan Y, Ji H (2019) Barzilai–Borwein-based adaptive learning rate for deep learning. Pattern Recogn Lett 128:197–203
5. Kolbusz J, Rozycki P, Lysenko O, Wilamowski BM. (2019) "Error Back Propagation Algorithm with Adaptive Learning Rate," In: 2019 International Conference on Information and Digital Technologies (IDT), pp 216–222
6. Nawi NM, Hussein AS, Samsudin NA, Hamid NA, Yunus M, Amin M et al (2017) The effect of pre-processing techniques and optimal parameters selection on back propagation neural networks. Int J Adv Sci Eng Inf Technol 7:770–777
7. Hamid NA, Nawi NM, Ghazali R, Salleh MNM (2012) Solving local minima problem in back propagation algorithm using adaptive gain, adaptive momentum and adaptive learning rate on classification problems. In: International Journal of Modern Physics: Conference Series, 2012, pp 448–455
8. Gorunescu F, Belciug S (2016) Boosting backpropagation algorithm by stimulus-sampling: application in computer-aided medical diagnosis. J Biomed Inform 63:74–81
9. Belciug S, Gorunescu F (2018) Learning a single-hidden layer feedforward neural network using a rank correlation-based strategy with application to high dimensional gene expression and proteomic spectra datasets in cancer detection. J Biomed Inform 83:159–166
10. Sundaram NM, Renupriya V (2016) Artificial neural network classifiers for diagnosis of thyroid abnormalities. Technology 206:211
11. Ioniţă I, Ioniţă L (2016) "Prediction of thyroid disease using data mining techniques. BRAIN Broad Res Artif Int Neurosci 7:115–124
12. Chen D, Niu J, Pan Q, Li Y, Wang M (2017) A deep-learning based ultrasound text classifier for predicting benign and malignant thyroid nodules. In: 2017 International Conference on Green Informatics (ICGI), 2017, pp 199–204
13. Alkhasawneh MS (2019) Hybrid cascade forward neural network with elman neural network for disease prediction. Arab J Sci Eng 44:9209–9220
14. Ahmad W, Huang L, Ahmad A, Shah F, Iqbal A (2017) Thyroid diseases forecasting using a hybrid decision support system based on ANFIS, k-NN and information gain method. J Appl Environ Biol Sci 7:78–85
15. Ozyilmaz L, Yildirim T (2002) Diagnosis of thyroid disease using artificial neural network methods. In: Proceedings of the 9th International Conference on Neural Information Processing, 2002, pp 2033–2036
16. Iranmanesh S, Mahdavi MA (2009) A differential adaptive learning rate method for back-propagation neural networks. World Acad Sci Eng Technol 50:285–288
17. Kathirvalavakumar T, Subavathi SJ (2009) Neighborhood based modified backpropagation algorithm using adaptive learning parameters for training feedforward neural networks. Neurocomputing 72:3915–3921

18. Nawi NM, Ransing RS, Salleh MNM, Ghazali R, Hamid NA (2010) An improved back propagation neural network algorithm on classification problems. In: Database Theory and Application, Bio-Science and Bio-Technology, pp 177–188
19. Jha SK, Pan Z, Elahi E, Patel N (2019) A comprehensive search for expert classification methods in disease diagnosis and prediction. Expert Syst 36:e12343
20. Nguyen DT, Pham TD, Batchuluun G, Yoon HS, Park KR (2019) Artificial intelligence-based thyroid nodule classification using information from spatial and frequency domains. J clin Med 8:1976
21. M. Al-Khafajiy, L. Webster, T. Baker, and A. Waraich, "Towards fog driven IoT healthcare: challenges and framework of fog computing in healthcare," In: Proceedings of the 2nd International Conference on Future Networks and Distributed Systems, 2018, pp. 1–7.
22. Fahim M, Baker T, Khattak AM, Alfandi O (2017) Alert me: Enhancing active lifestyle via observing sedentary behavior using mobile sensing systems. In: 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 2017, pp 1–4
23. Åström F, Koker R (2011) A parallel neural network approach to prediction of Parkinson's disease. Expert Syst Appl 38:12470–12474
24. Zeiler MD (2012) ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701
25. Sharma A (2018) Guided Stochastic gradient descent algorithm for inconsistent datasets. Appl Soft Comput 73:1068–1080
26. Lichman M (2013) UCI Machine Learning Repository. Irvine, University of California, Irvine, School of Information and Computer Sciences, Irvine
27. Bensenor IM, Olmos RD, Lotufo PA (2012) Hypothyroidism in the elderly: diagnosis and management. Clin Interv Aging 7:97
28. Kaur A, Singh N, Bahrdwaj A (2013) A comparison of supervised multilayer back propagation and unsupervised self organizing maps for the diagnosis of thyroid disease. Int J Comp Appl 82:39–43
29. Souri A, Hosseinpour S, Rahmani AM (2018) Personality classification based on profiles of social networks' users and the five-factor model of personality. Human-centric Comput Inf Sci 8:24
30. Nweke HF, Teh YW, Mujtaba G, Alo UR, Al-garadi MA (2019) Multi-sensor fusion based on multiple classifier systems for human activity identification. Human-centric Comput Inf Sci 9:34
31. Heidari AA, Faris H, Mirjalili S, Aljarah I, Mafarja M (2020) Ant Lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks. In: Nature-Inspired Optimizers, 2020, pp 23–46
32. Guliyev NJ, Ismailov VE (2018) Approximation capability of two hidden layer feedforward neural networks with fixed weights. Neurocomputing 316:262–269
33. Henríquez PA, Ruz GA (2018) A non-iterative method for pruning hidden neurons in neural networks with random weights. Appl Soft Comput 70:1109–1121
34. Qian X, Zhong X (2019) Optimal individualized multimedia tourism route planning based on ant colony algorithms and large data hidden mining. Multimedia Tools Appl 78(15):22099–22108
35. Taherkhani A, Belatreche A, Li Y, Maguire LP (2018) A supervised learning algorithm for learning precise timing of multiple spikes in multilayer spiking neural networks. IEEE Trans Neural Netw Learn Syst 29:5394–5407
36. Yakopcic C, Hasan R, Taha TM (2018) Flexible memristor based neuromorphic system for implementing multi-layer neural network algorithms. Int J Parallel Emerg Distrib Syst 33:408–429
37. Nagarajan HP, Mokhtarian H, Jafarian H, Dimassi S, Bakrani-Balani S, Hamedi A, et al. (2019) Knowledge-based design of artificial neural network topology for additive manufacturing process modeling: a new approach and case study for fused deposition modeling. J Mech Des 141:1–12
38. Rene ER, López ME, Park HS, Murthy D, Swaminathan T (2012) ANNs for identifying shock loads in continuously operated biofilters: application to biological waste gas treatment. In: Handbook of Research on Industrial Informatics and Manufacturing Intelligence: Innovations and Solutions, 2012, pp 72–103
39. Sheel S, Varshney T, Varshney R (2007) Accelerated learning in MLP using adaptive learning rate with momentum coefficient. In: 2007 International Conference on Industrial and Information Systems, pp 307-310
40. Seth N, Ubrani A, Mane S, Kazi FA (2019) Comparative analysis of major Jacobian and gradient backpropagation optimizers of ANN on SVPWM. In: Soft Computing and Signal Processing, 2019, pp 345–357
41. Belciug S, Gorunescu F (2016) Improving performance in neural networks using feature selection based on random forests : application to automated medical diagnosis. Int J Inf Technol Knowl 10(1):33–46

42. Shankar K, Lakshmanaprabu S, Gupta D, Maseleno A, De Albuquerque VHC (2020) Optimal feature-based multi-kernel SVM approach for thyroid disease classification. J Supercomput 76:1128–1143
43. Yadav DC, Pal S (2019) Thyroid prediction using ensemble data mining techniques. Int J Inf Technol 1–11
44. Boger Z, Guterman H (1997) Knowledge extraction from artificial neural network models. In: 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, 1997, pp 3030–3035

Affiliations

Mehdi Hosseinzadeh[1,2] · Omed Hassan Ahmed[3] · Marwan Yassin Ghafour[4] · Fatemeh Safara[5] · Hawkar kamaran hama[6] · Saqib Ali[7] · Bay Vo[8] · Hsiu-Sen Chiang[9]

Mehdi Hosseinzadeh
hosseinzadeh.m@iums.ac.ir

Omed Hassan Ahmed
omed.hassan@uhd.edu.iq

Marwan Yassin Ghafour
marwan.ghafour@uoh.edu.iq

Fatemeh Safara
fsafara@iiau.ac.ir

Hawkar kamaran hama
Hawkar.hama@univsul.edu.iq

Saqib Ali
saqib.ali@ieee.org

Hsiu-Sen Chiang
hschiang@nutc.edu.tw

[1]   Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

[2]   Health Management and Economics Research Centre, University of Medical Sciences, Tehran, Iran

[3]   Department of Information Technology, College of Science and Technology, University of Human Development, Sulaymaniyah, Iraq

[4]   Department of Computer Science, College of Science, University of Halabja, Halabja, Iraq

[5]   Department of Computer Engineering, Islamshahr Branch, Islamic Azad University, Islamshahr, Iran

[6]   Computer Science Department, College of Basic Education, University of Sulaimani, Sulaymaniyah, Iraq

[7]   Department of Information Systems, College of Economics and Political Science, Sultan Qaboos University, Al Khoudh, Muscat, Oman

[8]   Faculty of Information Technology, Ho Chi Minh City University of Technology (HUTECH), Ho Chi Minh City, Vietnam

9　Department of Information Management, National Taichung University of Science & Technology, 404, Taichung City, Taiwan