Expert Systems With Applications 140 (2020) 112875

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Comparing of deep neural networks and extreme learning machines based on growing and pruning approach

Kemal Akyol

Department of Computer Engineering, Kastamonu University, Kastamonu 37100, Turkey

ARTICLE INFO

Article history:
Received 8 June 2019
Revised 30 July 2019
Accepted 15 August 2019
Available online 16 August 2019

Keywords:
Deep Neural Networks
Extreme Learning Machines
Growing and pruning
Parkinson
Self-care activities

ABSTRACT

Recently, the studies based on Deep Neural Networks and Extreme Learning Machines have become prominent. The models of parameters designed in these studies have been chosen randomly and the models have been designed in this direction. The main focus of this study is to determine the ideal parameters i.e. optimum hidden layer number, optimum hidden neuron number and activation function for Deep Neural Networks and Extreme Learning Machines architectures based on growing and pruning approach and to compare the performances of the models designed. The performances of the models are evaluated on two datasets; Parkinson and Self-Care Activities Dataset. Multi experiments have verified that the Deep Neural Networks architectures present a good prediction performance and this architecture outperforms the Extreme Learning Machines.

1. Introduction

Artificial Neural Network (ANN) is computational tool inspired from biological neural network system (Parveen Kumar & Pooja Sharma, 2014). It is used in many fields such as engineering sciences, specially computer sciences like medical diagnosis (Jafari-Marandi, Davarzani, Soltanpour Gharibdousti, & Smith, 2018), feature extraction based on image classification (Aytaç Korkmaz & Binol, 2018), time series prediction (Panigrahi & Behera, 2017) etc. A considerable amount of research based on the Deep Neural Networks (DNN) and Extreme Learning Machines (ELM) models have been proposed in recent years, such as a DNN model combined with the discrete wavelet transform and principal components analysis (Mohsen, El-Dahshan, El-Horbaty, & Salem, 2018); a DNN model combined with signal processing (Sannino & De Pietro, 2018), two-hidden-layer ELM (Qu, Lang, Liang, Qin, & Crisalle, 2015), two-stage ELM (Lan, Soh, & Huang, 2010), a weighted ELM for imbalanced class distribution (Li, Kong, Lu, Wenyin, & Yin, 2014), face recognition (Mohammed, Minhas, Wu, & Sid-Ahmed, 2011), handwritten character recognition (Chacko, Vimal Krishnan, Raju, & Babu Anto, 2012), image classification (Jun, Shitong & Chung, 2011), multiclass classification (Eirola et al., 2015).

The main focus of this study is to compare and evaluate the performances of the Multi-layer ELM architectures and Multi-layer DNN architectures based on growing and pruning approach. Also, the main contribution of this study is to find optimum hidden layer number and hidden neuron number and determine the ideal activation function for ELM model and ideal couple of activation function and optimization function for DNN model. The reason why these two architectures are preferred is that these two architectures are suitable for this approach.

The rest of this study is organized as follows; Section 2 presents the related works for ELM end DNN architectures briefly. Section 3 presents some background on the learning algorithms we have used. Section 4 addresses the experiments and results of the DNN and ELM models carried out on two datasets. Finally, Section 5 draws discussion and conclusion.

2. Literature review

An overview of some of the previous studies related with DNN and ELM is presented below. Lee et al. tried to estimate LDL-cholesterol by utilizing the DNN model including three input values of total cholesterol, HDL cholesterol, and triglyceride. The model, which consists of six hidden layers with 30 nodes, was trained on the dataset collected from Korean National Health and Nutrition Examination Survey. The performance of the model was tested on another dataset collected from Wonju Severance Christian Hospital. The model presented better performance compared to other existing methods (Lee, Kim, Uh, & Lee, 2019).

Feng et al. presented a DNN regression in order to predict solidification defects on the small dataset which consists of 487 instances. According to this study, pre-trained and fine-tuned DNN outperform neural network, support vector machine, and DNN

E-mail address: kakyol@kastamonu.edu.tr

https://doi.org/10.1016/j.eswa.2019.112875
0957-4174/

trained by conventional methods (Feng, Zhou, & Dong, 2019). Kudugunta and Ferrara presented a deep neural networks based on long short-term memory architecture in order to detect bots on tweets. For this purpose, contextual features obtained from user metadata were sent to the DNN architecture. Based on results, proposed architecture presents high classification accuracy in recognizing the bots from humans (Kudugunta & Ferrara, 2018). Qi et al. presented a deep convolutional neural networks with multi-scale kernels and skip connections to diagnose breast ultrasonography images. Firstly, malignant tumors on the image were detected and then solid nodules were recognized (Qi et al., 2019). Another deep convolutional neural networks, which includes multiple-layer perceptrons and convolutional neural networks, were presented by Yang et al. They proposed a novel regulator named Structured Decorrelation Constraint in order to tackle both the generalization and optimization of deep neural networks (Yang, Xiong, Li, & Xu, 2019).

Cheng and Xiong presented an ELM model in order to improve the accuracy of dam displacement prediction (Cheng & Xiong, 2017). Yeom and Kwak focused on the prediction of the short-term electricity-load using a Takagi-Sugeno-Kang-based ELM. They achieved superior prediction performance and knowledge information with four activation functions such sigmoid, sine, radial basis function, and rectified linear unit (Yeom, Kwak, Yeom, & Kwak, 2017). Lu et al. proposed a novel adaptive weight online sequential extreme learning machine for predicting time series problems. The proposed study has good performance with respect to generalization performance, stability, and prediction ability (Lu et al., 2017). Men et al. developed a paraffin odor analysis system. The performances of the Support Vector Machine, Random Forest, and ELM algorithms on original feature set and optimized datasets are evaluated. Based on results, the ELM based model outperformed others (Men et al., 2018). Hosseinioun used the wavelet transform and adaptive ELM for prediction of the outlier occurrence in stock market time series (Hosseinioun, 2016). Lastly, Huang focused on the ELM theories such as hidden nodes and hidden neurons that need to be tuned in learning, and proved that it is good performance (Huang, Zhu, & Siew, 2006).

3. Background

3.1. Extreme Learning Machines and Deep Neural Networks

With the purpose of offering the information about the methods used in this study, the basic concepts of ELM and DNN architectures are briefly introduced in this section. Theoretical foundations of these architectures are well rooted from the classical neural networks architecture and they have been quite popular lately in the machine learning and data mining studies.

ELM proposed by Huang et al. (2006) is a new feed-forward neural network method which is presented high classification accuracy, good generalization ability rapidly (Cheng & Xiong, 2017). The basic of ELM is generalized as single hidden layer feed-forward networks where input weights and hidden biases are selected randomly. During the training process, the hidden layer of single hidden layer feed-forward networks need not to be tuned. The output weights are stated by using Moore-Penrose generalized inverse of the hidden-layer output matrix (Yeom et al., 2017). The structure of the ELM is shown in Fig. 1.

$$f_L(x) = \sum_{i=1}^{L} \beta_i G_i(w_i, b_i, x) \quad (1)$$

$$H(x) = [G_1(w_1, b_1, x), \ldots, G_L(w_L, bL_L, x)] \quad (2)$$

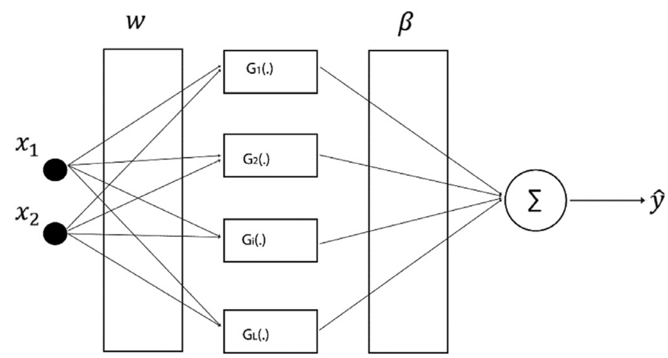

Fig. 1. An overview of conventional ELM architecture (Yeom et al., 2017).

Eq. (1) presents the output function of a generalized single hidden layer feed-forward networks.

Eq. (2) presents the output function in the hidden-layer mapping. w_i and b_i indicate weights and biases between the input layer and the hidden layer, respectively.

Huang revealed that hidden layer parameters can be assigned randomly and then the output weight can be calculated analytically (Huang, Zhu, & Siew,). Also, the execution time of an ELM model is very low and it provides more successful performance than other algorithms (Huang et al., 2006). As a known subject in the field of machine learning, of course, these parameters will vary for any dataset to be studied. Recently, there are many studies based on multi-layers ELM: a) the numbers of multi-layer and multi neurons assigned randomly b) average performance was calculated by utilizing multiple tests (i.e. the program run 1000 times). For example, Li et al. presented the number of optimum neurons (Li et al., 2014) and Xiao et al. found the best transfer function (Xiao, Li, & Mao, 2017). Deng et al. used random parameters and evaluated the averaged results in order to reduce the effect of these parameters given randomly (Deng, Zheng, & Chen, 2009).

The theoretical foundations of Deep Learning (DL) are well rooted from the classical neural networks architecture. In other words, DL is an up-to-date ANN architecture, which has been developed continuously and rapidly with different algorithms and approaches since the first day of its emergence, will continue to be popular for a long time in the computer science and other many fields. DNN which is a general deep framework covers the classification or regression analysis applications such as pattern recognition, data mining, image recognition and natural language processing etc. It is a powerful architecture and very popular in machine learning achieving successful results by making inferences from a dataset (Ravi et al., 2017).

An overview of the proposed DNN is given in Fig. 2. Input layer consists of input parameters for the input, hidden layers consist of hidden neurons and output layer consists of target class parameters.

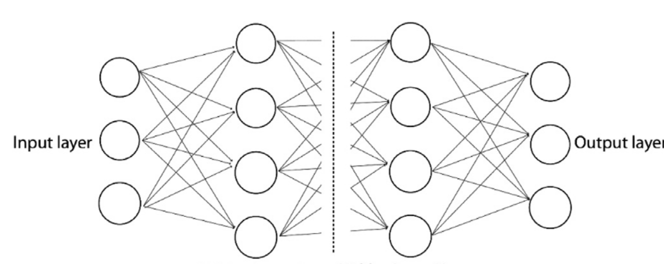

Fig. 2. An overview of the DNN architecture (Ravi et al., 2017).

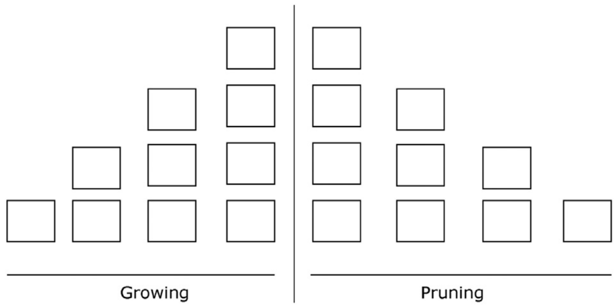

Fig. 3. An overview on growing and pruning approach for optimum neuron and layer numbers.

3.2. Growing and pruning approach

In this approach, firstly, an architecture is designed with minimum items. Minimum items indicate the number of necessary hidden neurons and layers. As can be seen in Fig. 3, by applying growing criteria, new layers and neurons are added to the architecture. In pruning approach, the architecture is designed with maximum hidden neurons assigned by growing approach and the model is started to be pruned step by step. The following operations are repeated for both architectures until reasonable performances are achieved (Thoma, 2017).

(a) Training the model
(b) Changing the weights according to a growing or pruning criteria
(c) Retraining the model

Thus, the best combination of the growing and pruning approach has been tried to seek for desired accuracy.

4. Experiments

4.1. Datasets

Table 1 addressed the information about the datasets. Self-Care Activities Dataset, hereafter SCADI having 70 instances includes self-care problem information of children with physical and motor disabilities. Outcome attribute of this dataset has seven categories as follows:

(a) Class 1 = Caring for body parts problem
(b) Class 2 = Toileting problem
(c) Class 3 = Dressing problem
(d) Class 4 = Washing oneself and Caring for body parts and Dressing problem
(e) Class 5 = Washing oneself, Caring for body parts, Toileting, and Dressing problem
(f) Class 6 = Eating, Drinking, washing oneself, caring for body parts, toileting, Dressing, looking after one's health and Looking after one's safety problem
(g) Class 7 = No Problem

SCADI dataset has 16 instances for healthy individuals and 54 instances having a problem between 1 and 6 class number.

Table 1
Information about the datasets.

Dataset	Training / Test size	Attributes	Class
SCADI	42/28	205	7
PD	453/303	753	2

Parkinson Disease (PD) dataset has 756 instances considering 564 patients with PD and 192 healthy individuals. Outcome attribute has 2 categories including 1 and 0 categorical values; having PD and not having PD respectively. It is clear that the number PD instances is more than the number non-PD instances.

4.2. Experimental procedure

All experiments are conducted with Python 3.6 programming language on Anaconda platform using 'Keras' and 'hpelm' libraries on the Anaconda environment running on a computer with Intel i7-8550U 1.99 GHz CPU and 8 GB RAM. 'Keras' is a deep learning library that contains large collections of deep learning architectures. 'Hpelm' is an extreme machine learning library that contains large collections of the ELM architectures.

Fig. 4 indicates conducted steps for reliable statistical results. The reason of comparing these two architectures is that this workflow is suitable for both architectures.

In growing approach, the related neuron or layer number is increased as long as the designed model's performance is boosted. If there is no increase in model's performance, these steps are repeated by adding a new hidden layer to the model. If the neuron number is bigger than input parameters number, it is continued with pruning approach. In pruning approach, concerning neuron is reduced by one as long as the predictive performance of a designed model is boosted. Note that the number of hidden neurons must be bigger than target class number. In other words, if the hidden neuron number is lower or equal to target class number, executing of this method is stopped automatically.

In literature studies, it has been seen that any model designed based on EML and DNN architectures consists of a specific parameter value or random neuron and layer number. For the determination of the best DNN and ELM architectures with an activation functions, hidden layers and the neuron numbers, the models are tested on two datasets. It is aimed to improve the performances of the models based on growing-pruning method and also decide the final classification models based on prediction accuracy.

The basic processes listed below are carried out for both datasets;

(a) Removing some instances including missing values.
(b) Converting the categorical information such as yes/no to categorical values such as 1/0.

The study presents 1) the finding of optimum hidden layer numbers, hidden neuron numbers, and also ideal activation function and optimization function for the best DNN model, 2) the finding of optimum hidden layer numbers, hidden neuron numbers and also ideal activation function for the best ELM model.

'tanh', 'relu', 'softmax', 'softsign' and 'sigmoid' activation functions and *'Nadam', 'RMSprop', 'Adagrad', 'Adadelta', 'Adam', 'Adamax' and 'SGD'* optimization functions are used for DNN models. These functions are collected from www.keras.com. It presents detail information about these functions. *'lin', 'sigm', 'tanh', 'rbf_l1', 'rbf_l2' and 'rbf_linf'* activation functions are used for ELM models. These functions are collected from www.hpelm.com. It presents detail information about these functions. Specifically, as a technical detail, sigmoid function is used for output layer since One-Hot Encoding is applied for target class in DNN models. Besides, in ELM architectures, the classification mode is assigned to 'c (classifier)' since the target class number is equal to 2 for Parkinson dataset. The classification mode is assigned to 'mc (multiclassifier)' since the target class number is more than 2 for SCADI dataset.

60% of the data is reserved for training and rest of the dataset is reserved for testing. Thus, each model designed will try to learn on train data and successes of the models are evaluated on test sub-datasets. The Accuracy metric is used to determine any model's

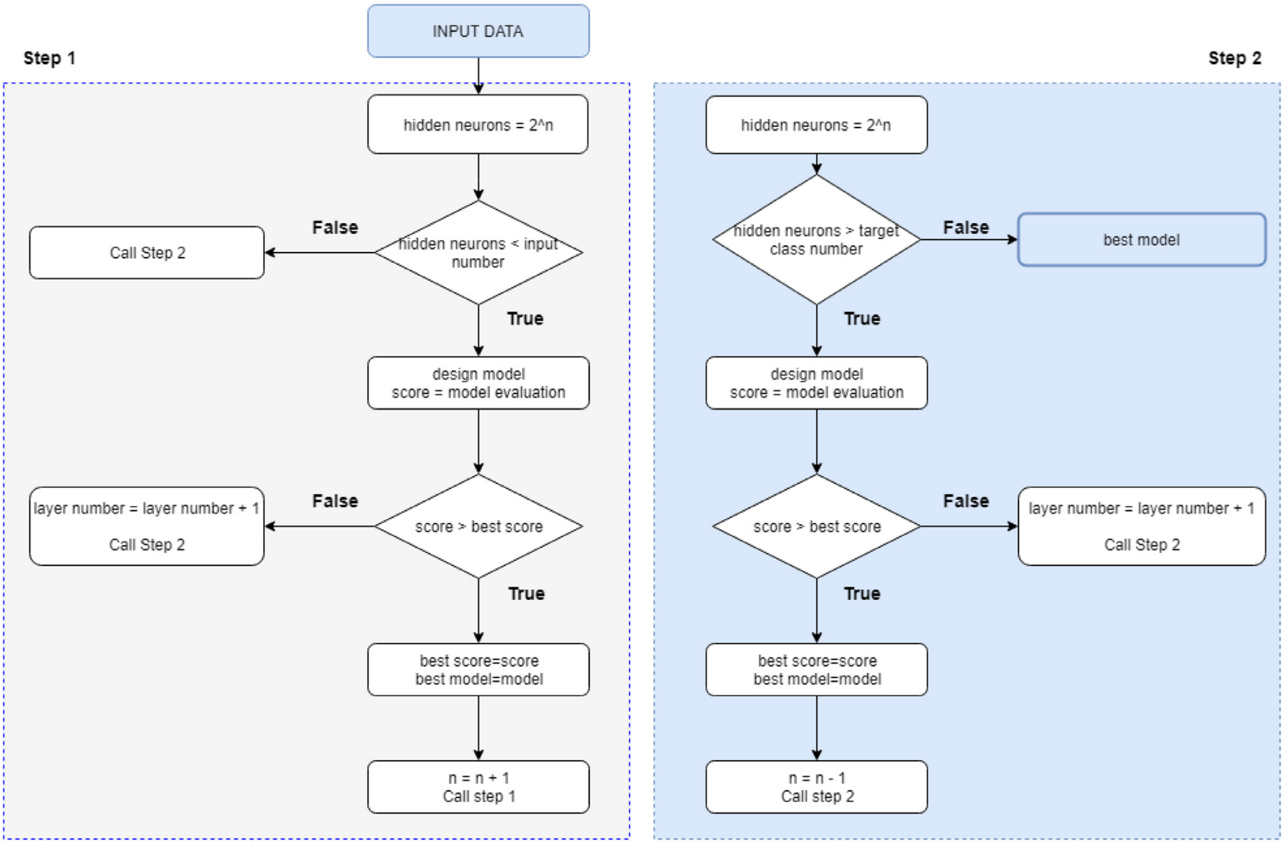

Fig. 4. The flowchart of the proposed study.

prediction performance. This metric which is shown in Eq. (3) is the ratio of the number of correctly classified instances to the number of all instances (Shaikh, 2011).

$$Accuracy = (TN + TP)/(TN + FP + TP + FN) \tag{3}$$

where True positive (TP) is the positive instance number predicted correctly. True negative (TN) is the negative instance number predicted correctly. False positive (FP) is the negative instance number predicted as positive incorrectly. False negative (FN) is the positive instance number predicted as negative incorrectly.

4.3. Results

Based on multiple experiments performed on two datasets, the maximum and minimum accuracies information for the DNN and ELM models are given in Tables 2 and 3.

When Tables 2 and 3 are examined;

(a) It can be observed that the best prediction performance is% 88.88 obtained with the ELM model with 'sigm' activation function and 2 hidden layers, each of which is 2 and 4 neurons, respectively for SCADI dataset. The best prediction performance is% 97.45 obtained with the DNN model with 'relu' activation function, 'Adagrad' optimization function and 7 hidden layers for each 2, 4, 8, 16, 32, 32 and 16 neurons, respectively. Thus, the DNN model presented the best performance for this dataset and target variable is predicted acceptable level by this model.

(b) It can be observed that the best prediction performance is% 83.70 obtained with the ELM model with 'sigm' activation function and 6 hidden layers, each of which is 2, 4, 8, 16, 32 and 64 neurons, respectively for Parkinson dataset. The best prediction performance is% 95.15 obtained with the DNN model with 'relu' activation function, SGD optimization function and 7 hidden layers for each 2, 4, 8, 16, 32, 64 and 64 neurons, respectively. Thus, the DNN model presented the best performance for this dataset and target variable is predicted acceptable level by this model.

For the sake of comparison, the performances of the best DNN models is significantly higher than the best ELM models for all datasets. In other words, the ELM models is not an effective in the prediction of target variables. So we can make a generalization that the DNN algorithm has stronger performance than ELM models. Also, Figs. 5 and 6 shows the predictive accuracies of the DNN and ELM models which presents best or worst performance on test sub-datasets. Based on these figures, the DNN model classified well the SCADI dataset with 7 hidden layers including different hidden

Table 2
Some of the models designed with DNN and their performances on the test data.

			Activation function	Optimization function	Hidden layers and each of which neurons
SCADI	Maximum accuracy	97.45	'relu'	'Adagrad'	[205, 2, 4, 8, 16, 32, 32, 16, 7]
	Minimum accuracy	39.29	'sigmoid'	'SGD'	[205, 2, 4, 7]
Parkinson disease	Maximum accuracy	95.15	'relu'	'SGD'	[753, 2, 4, 8, 16, 32,64, 2]
	Minimum accuracy	76.66	'relu'	'RMSprop'	[753, 2, 4, 2]

Table 3
Some of the models designed with ELM and their performances on the test data.

			Activation function	Hidden layers and each of which neurons
SCADI	Maximum accuracy	88.88	'sigm'	[205, 2, 4, 7]
	Minimum accuracy	40.0	'tanh'	[205, 2, 4, 8, 16, 7]
Parkinson disease	Maximum accuracy	83.70	'sigm'	[753, 2, 4, 8, 16, 32, 64, 2]
	Minimum accuracy	24.23	'rbf_l1'	[753, 2, 4, 2]

Table 4
The parameters settings for designed the DNN models presented best performance.

Parameters	SCADI dataset Value	Parkinson dataset Value
Number of input layer neurons	205	753
Number of hidden layers	7	6
Number of hidden layer-1 neurons	2	2
Number of hidden layer-2 neurons	4	4
Number of hidden layer-3 neurons	8	8
Number of hidden layer-4 neurons	16	16
Number of hidden layer-5 neurons	32	32
Number of hidden layer-6 neurons	32	64
Number of hidden layer-7 neurons	16	–
Number of output layers neuron	7	2
Activation function	relu	relu
Learning cycle	100 epochs	100 epochs
Learning algorithm	Adagrad	SGD

neurons and the DNN model classified well the Parkinson dataset with 6 hidden layers including different hidden neurons.

The parameters settings of the best models are presented in Table 4. Rectified Linear Units (relu) function which offers nonlinearity is used for activation function. It is mathematically given by Eq. (4) (Maas, Maas, Hannun, & Ng, 2013).

$$h^{(i)} = \max\left(w^{(i)T}\ x, 0\right) = \begin{cases} w^{(i)T}\ x & w^{(i)T}\ x > 0 \\ 0 & else \end{cases} \tag{4}$$

In Eq. (4), $w^{(i)T}$: i. The weight vector for hidden layer, x: input

For SCADI and Parkinson data, respectively, 'Adagrad' and 'Stochastic Gradient Descent (SGD)' optimization functions presented best performance. 'Adagrad' is a method eliminating the problem arising from the constant learning coefficient in the gradient descent method. According to this method, the learning coefficient is updated in each step (Duchi, Hazan, & Singer, 2011). SGD provides quick results by making a parameter update for each training example x and target variable y and can also be used for online learning (Ruder, 2019).

Table 5 presents the information about the previous studies and the proposed study considering the methods and approaches. The performance of the proposed study is compared with these studies.

(a) Along with many studies performed on different Parkinson's disease datasets, recently, this dataset has been introduced by Sakar et al. They presented a study on PD classification

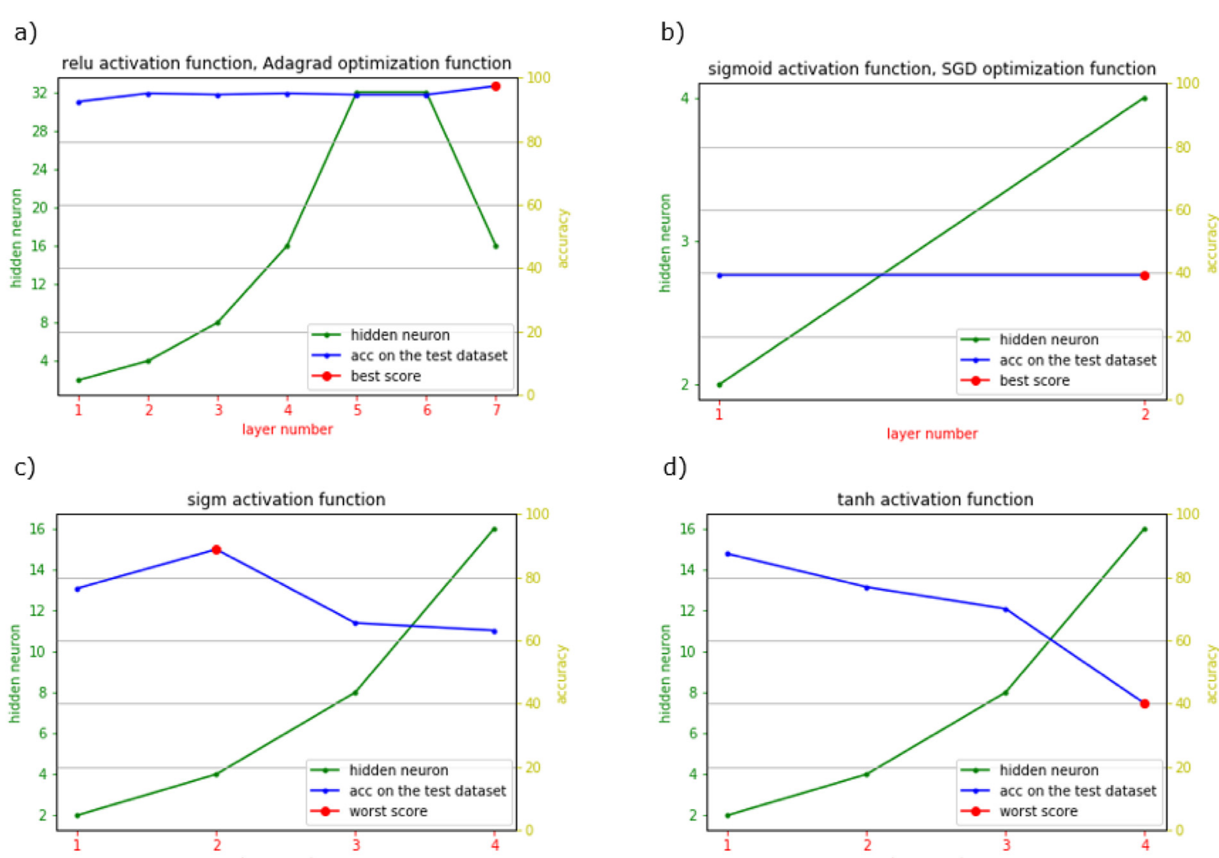

Fig. 5. The graphical representation of the performances of the DNN and ELM based models for the SCADI dataset; a) the best DNN model, b) the worst DNN model, c) the best ELM model, d) the worst ELM model.

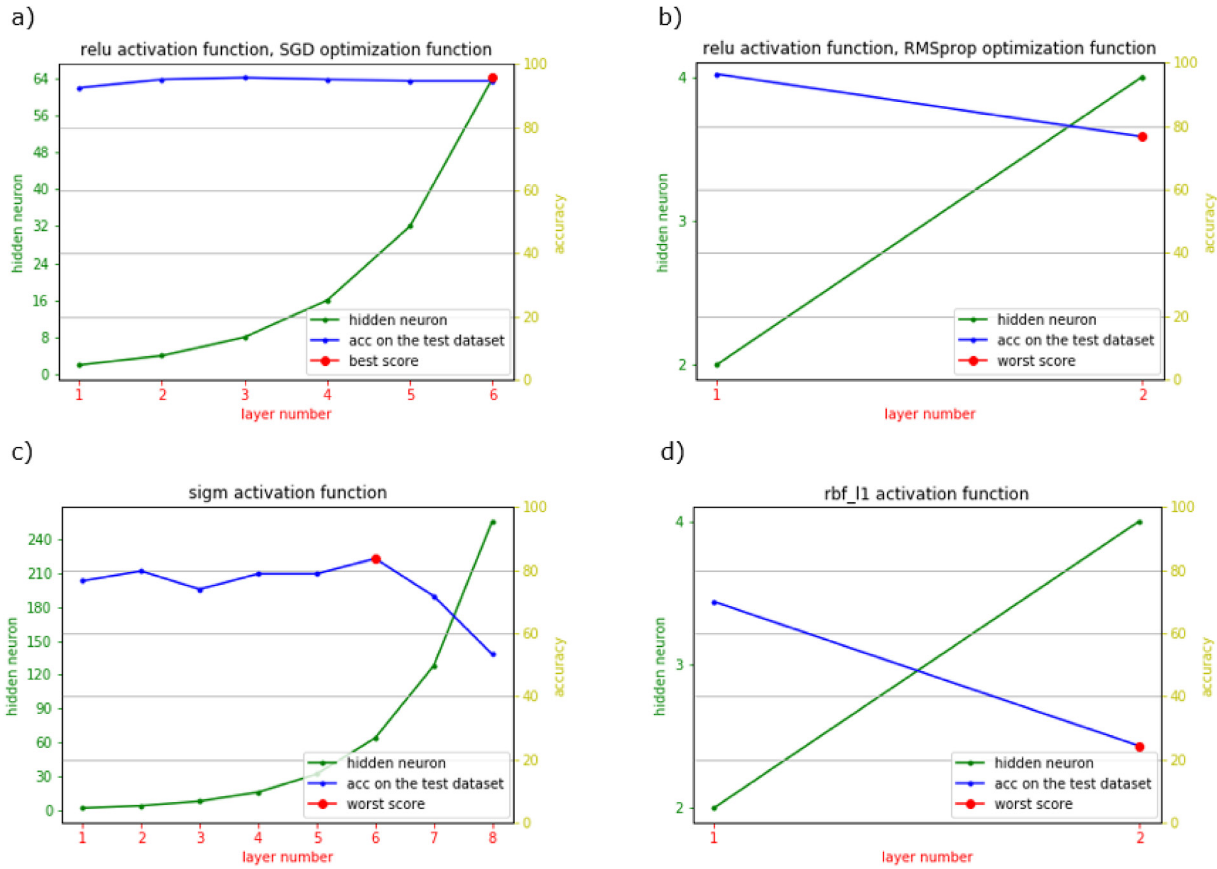

Fig. 6. The graphical representation of the performances of the DNN and ELM based models for the Parkinson dataset; a) the best DNN model, b) the worst DNN model, c) the best ELM model, d) the worst ELM model.

Table 5
Comparison of the performance of the studies in the literature.

Studies	Accuracy
Parkinson	
Sakar et al. (2013)	86%
The proposed method – DNN base	95.15%
The proposed method – ELM base	83.70%
SCADI	
Zarchi et al. (2018)	83.1%
Keles and Kilic (2018)	88.57%
Le and Baik (2019)	85.4%
Choudhury (2018)	84.75%
The proposed method – DNN base	97.45%
The proposed method – ELM base	88.88%

based on feature selection and state of arts machine learning algorithms (Sakar et al., 2013). The performance of this study is compared only with Sakar's study because another study which uses this dataset is not in literature.
The proposed approach presented 95.15% Accuracy while Sakar et al. presented % 86 Acc.

(b) Zarchi et al. introduced a new dataset namely SCADI and proposed a rule-based and Artificial Neural Network (ANN)-based two different types of expert systems for the self-care problems classification of children with physical and motor disability on this dataset. The ANN-based system has high accuracy value of 83.1% by using 10-fold cross validation technique (Zarchi, Fatemi Bushehri, & Dehghanizadeh, 2018). Keleş and Kılıç achieved 88.57% accuracy by utilizing the combination of Artificial Bee Colony algorithms and two machine learning algorithms; KNN and Naïve Bayes (Keles & Kilic, 2018). Le and Baik achieved 85.4% accuracy by utilizing the balanced dataset and feature selection methods (Le & Baik, 2019). Finally, in another study, Choudhury presented 84.75% accuracy by using the Boruta feature selection algorithm and Random Forest algorithm on SCADI dataset (Choudhury, 2018).

The proposed approach in this study presented 97.45% accuracy value. In this context, it is thought that the proposed approach is important considering literature contribution.

5. Conclusion and discussion

In this study, several DNN and ELM models based on growing-pruning method are designed and the performances of these models are examined in order to find ideal model. In this context, the study covers the finding of ideal hidden layer number, hidden neuron number, the activation function and the optimization function for the best DNN model. Also, it covers the finding of ideal hidden layer number, hidden neuron number and the activation function for the best ELM model. Another important contribution of this study is the comparison of DNN and ML-ELM. By comprehensively comparing all models;

(a) The DNN model with 'relu' activation function, 'Adagrad' optimization function and 7 hidden layers, each of which is 2, 4, 8, 16, 32, 32 and 16 neurons, respectively gave the best prediction performance with% 97.45 accuracy value on the SCADI dataset.

(b) The DNN model with ‘relu’ activation function, ‘SGD’ optimization function and 6 hidden layers, each of which is 2, 4, 8, 16, 32 and 64 neurons, respectively gave the best prediction performance with% 95.15 accuracy value on the Parkinson dataset.

Overall results show that the DNN architectures are superior to ELM architectures. However, it is worth noting that the characteristics and structure of any dataset directly affect the successes of algorithms carried out in the field of machine learning and data mining.

Declaration of Competing Interest

I wish to confirm that there are no known conflicts of interest associated with this publication and there has been no financial support for this work that could have influenced its outcome.

Credit authorship contribution statement

Kemal Akyol: Conceptualization, Formal analysis, Writing - original draft.

References

Aytaç Korkmaz, S., & Binol, H. (2018). Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection. *Journal of Molecular Structure, 1156*, 255–263. https://doi.org/10.1016/j.molstruc.2017.11.093.

Chacko, B. P., Vimal Krishnan, V. R., Raju, G., & Babu Anto, P. (2012). Handwritten character recognition using wavelet energy and extreme learning machine. *International Journal of Machine Learning and Cybernetics, 3*(2), 149–161. https://doi.org/10.1007/s13042-011-0049-5.

Cheng, J., & Xiong, Y. (2017). Application of extreme learning machine combination model for dam displacement prediction. *Procedia Computer Science, 107*, 373–378. https://doi.org/10.1016/J.PROCS.2017.03.120.

Choudhury, A. (2018). Classification of functioning, disability, and health: Icf-cy self care (SCADI dataset) using predictive analytics. Retrieved from http://arxiv.org/abs/1901.00756.

Deng, W., Zheng, Q., & Chen, L. (2009). Real-Time collaborative filtering using extreme learning machine. In *2009 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology* (pp. 466–473). IEEE. https://doi.org/10.1109/WI-IAT.2009.80.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research, 12*. https://doi.org/10.1109/CDC.2012.6426698.

Eirola, E., Gritsenko, A., Akusok, A., Björk, K.-M., Miche, Y., Sovilj, D. et al. (2015). Extreme learning machines for multiclass classification: Refining predictions with gaussian Gaussian mixture models (pp. 153–164). *Springer, Cham*. https://doi.org/10.1007/978-3-319-19222-2_13.

Feng, S., Zhou, H., & Dong, H. (2019). Using deep neural network with small dataset to predict material defects. *Materials & Design, 162*, 300–310. https://doi.org/10.1016/J.MATDES.2018.11.060.

Hosseinioun, N. (2016). Forecasting outlier occurrence in stock market time series based on wavelet transform and adaptive ELM algorithm. *Journal of Mathematical Finance, 06*(01), 127–133. https://doi.org/10.4236/jmf.2016.61013.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1–3), 489–501. https://doi.org/10.1016/J.NEUCOM.2005.12.126.

Huan, Guang-B ig, n, Zhu, Qin-Yu, & Siew, Chee-Kheong (n.d.). Extreme learning machine: A new learning scheme of feedforward neural networks. In 2004 *IEEE International international Joint joint Conference conference on Neural neural Networks networks* (IEEE Cat. No.04CH37541) (Vol. 2, pp. 985–990) IEEE. https://doi.org/10.1109/IJCNN.2004.1380068.

Jafari-Marandi, R., Davarzani, S., Soltanpour Gharibdousti, M., & Smith, B. K. (2018). An optimum ANN-based breast cancer diagnosis: Bridging gaps between ann learning and decision-making goals. *Applied Soft Computing, 72*, 108–120. https://doi.org/10.1016/j.asoc.2018.07.060.

Jun, W., Shitong, W., & Chung, F. (2011). Positive and negative fuzzy rule system, extreme learning machine and image classification. *International Journal of Machine Learning and Cybernetics, 2*(4), 261–271. https://doi.org/10.1007/s13042-011-0024-1.

Keles, M. K., & Kilic, U. (2018). Artificial bee colony algorithm for feature selection on scadi dataset. In *2018 3rd international conference on computer science and engineering (UBMK)* (pp. 463–466). IEEE. https://doi.org/10.1109/UBMK.2018.8566287.

Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. *Information Sciences, 467*, 312–322. https://doi.org/10.1016/j.ins.2018.08.019.

Lan, Y., Soh, Y. C., & Huang, G.-B. (2010). Two-stage extreme learning machine for regression. *Neurocomputing, 73*(16–18), 3028–3038. https://doi.org/10.1016/j.neucom.2010.07.012.

Le, T., & Baik, S. W. (2019). A robust framework for self-care problem identification for children with disability. *Symmetry, 11*(1), 89. https://doi.org/10.3390/sym11010089.

Lee, T., Kim, J., Uh, Y., & Lee, H. (2019). Deep neural network for estimating low density lipoprotein cholesterol. *Clinica Chimica Acta, 489*, 35–40. https://doi.org/10.1016/J.CCA.2018.11.022.

Li, K., Kong, X., Lu, Z., Wenyin, L., & Yin, J. (2014). Boosting weighted elm for imbalanced learning. *Neurocomputing, 128*, 15–21. https://doi.org/10.1016/J.NEUCOM.2013.05.051.

Lu, J., Huang, J., Lu, F., Lu, J., Huang, J., & Lu, F. (2017). Time series prediction based on adaptive weight online sequential extreme learning machine. *Applied Sciences, 7*(3), 217. https://doi.org/10.3390/app7030217.

Maas, A. L., Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *ICML workshop on deep learning for audio, speech and language processing* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.693.1422 .

Men, H., Fu, S., Yang, J., Cheng, M., Shi, Y., & Liu, J. (2018). Comparison of SVM, rf and elm on an electronic nose for the intelligent evaluation of paraffin samples. *Sensors, 18*(1), 285. https://doi.org/10.3390/s18010285.

Mohammed, A. A., Minhas, R., Wu, Q. M. J., & Sid-Ahmed, M. A. (2011). Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognition, 44*, 2588–2597. https://doi.org/10.1016/j.patcog.2011.03.013.

Mohsen, H., El-Dahshan, E.-S. A., El-Horbaty, E.-S. M., & Salem, A.-B. M. (2018). Classification using deep learning neural networks for brain tumors. *Future Computing and Informatics Journal, 3*(1), 68–71. https://doi.org/10.1016/J.FCIJ.2017.12.001.

Panigrahi, S., & Behera, H. S. (2017). A hybrid ETS–ANN model for time series forecasting. *Engineering Applications of Artificial Intelligence, 66*, 49–59. https://doi.org/10.1016/j.engappai.2017.07.007.

Parveen Kumar, E., & Pooja Sharma, E. (2014). Artificial neural networks-A study. *International Journal of Emerging Engineering Research and Technology, 2*. Retrieved from www.ijeert.org .

Qi, X., Zhang, L., Chen, Y., Pi, Y., Chen, Y., Lv, Q., et al. (2019). Automated diagnosis of breast ultrasonography images using deep neural networks. *Medical Image Analysis, 52*, 185–198. https://doi.org/10.1016/J.MEDIA.2018.12.006.

Qu, B. Y., Lang, B. F., Liang, J. J., Qin, A. K., & Crisalle, O. D. (2015). Two-hidden-layer extreme learning machine for regression and classification. *Neurocomputing*, 1–9. https://doi.org/10.1016/j.neucom.2015.11.009.

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., et al. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics, 21*(1), 4–21. https://doi.org/10.1109/JBHI.2016.2636665.

Ruder, S. (2019). An overview of gradient descent optimization algorithms *. Retrieved from http://caffe.berkeleyvision.org/tutorial/solver.html.

Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., et al. (2013). Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics, 17*(4), 828–834. https://doi.org/10.1109/JBHI.2013.2245674.

Sannino, G., & De Pietro, G. (2018). A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Future Generation Computer Systems, 86*, 446–455. https://doi.org/10.1016/j.future.2018.03.057.

Shaikh, S. A. (2011). Measures derived from a 2×2 table for an accuracy of a diagnostic test. *Journal of Biometrics & Biostatistics, 02*(05), 1–4. https://doi.org/10.4172/2155-6180.1000128.

Thoma, M. (2017). *Analysis and optimization of convolutional neural network architectures Master Thesis of* Retrieved from www.kit.edu.

Xiao, D., Li, B., & Mao, Y. (2017). A multiple hidden layers extreme learning machine method and its application. *Mathematical Problems in Engineering*, 1–10. 2017 https://doi.org/10.1155/2017/4670187 .

Yang, J., Xiong, W., Li, S., & Xu, C. (2019). Learning structured and non-redundant representations with deep neural networks. *Pattern Recognition, 86*, 224–235. https://doi.org/10.1016/J.PATCOG.2018.08.017.

Yeom, C.-U., Kwak, K.-C., Yeom, C.-U., & Kwak, K.-C. (2017). Short-Term electricity-load forecasting using a TSK-Based extreme learning machine with knowledge representation. *Energies, 10*(10), 1613. https://doi.org/10.3390/en10101613.

Zarchi, M. S., Fatemi Bushehri, S. M. M., & Dehghanizadeh, M. (2018). SCADI: A standard dataset for self-care problems classification of children with physical and motor disability. *International Journal of Medical Informatics, 114*, 81–87. https://doi.org/10.1016/j.ijmedinf.2018.03.003.