

Growing and Pruning Based Deep Neural Networks Modeling for Effective Parkinson's Disease Diagnosis

Kemal Akyol^{1,*}

Abstract: Parkinson's disease is a serious disease that causes death. Recently, a new dataset has been introduced on this disease. The aim of this study is to improve the predictive performance of the model designed for Parkinson's disease diagnosis. By and large, original DNN models were designed by using specific or random number of neurons and layers. This study analyzed the effects of parameters, i.e., neuron number and activation function on the model performance based on growing and pruning approach. In other words, this study addressed the optimum hidden layer and neuron numbers and ideal activation and optimization functions in order to find out the best Deep Neural Networks model. In this context of this study, several models were designed and evaluated. The overall results revealed that the Deep Neural Networks were significantly successful with 99.34% accuracy value on test data. Also, it presents the highest prediction performance reported so far. Therefore, this study presents a model promising with respect to more accurate Parkinson's disease diagnosis.

Keywords: Parkinson's disease, machine learning, growing and pruning, deep neural networks.

1 Introduction

A progressive, chronic neurodegenerative condition [Nolden, Tartavouille and Porche (2014)], Parkinson's Disease (PD) is a quite difficult and serious disease that causes death [Gupta, Julka, Jain et al. (2018); Sharma, Sundaram, Sharma et al. (2019)]. Field-specialists carried out an accurate diagnosis of Parkinson's disease by applying several neurological, psychological and physical examinations. They examined the symptoms and signed on the nervous system conditions of a person. Also, they used the medical background and genetic factor of the patients [Mostafa, Mustapha, Mohammed et al. (2019)]. PD is very important for the patient's life; the detection of this disease through the correct meditations in prematurity phase and later is very important [Sharma, Sundaram, Sharma et al. (2019)]. Vocal problems are one of the most common symptoms in early stage of PD. Approximately 90% of the patients suffer from vocal problems. Diagnostic systems which detect these problems have been highly recommended recently [Sakar, Isenkul, Sakar et al. (2013)].

¹ Kastamonu University, Kuzeykent Yerleşkesi, Kastamonu, 37100, Turkey.

*Corresponding Author: Kemal Akyol. Email: kakyol@kastamonu.edu.tr.

Received: 13 June 2019; Accepted: 18 September 2019.

Tab. 1 addresses the previous studies considering the methods and approaches. When examined in detail in this table, speech signal processing algorithms, feature selection and the state of arts machine learning algorithms are used in order to robust decision support system.

Table 1: Comparison of the studies in the literature considering the methods and approaches

Studies	Techniques
Sakar et al. [Sakar, Isenkul, Sakar et al. (2013)]	well-known machine learning tools
Sakar et al. [Sakar and Kursun (2010)]	features selected by the maximum-relevance-minimum-redundancy method fed as input data to Support Vector Machine
Tsanas et al. [Tsanas, Little, McSharry et al. (2011)]	feature subsets selected by the LASSO sent as input to the classification, regression trees and RF learner
Gürüler [Gürüler (2017)]	k-means clustering-based feature weighting and the selected features fed to the complex-valued artificial neural network
Little et al. [Little, McSharry, Hunter et al. (2009)]	using a kernel support vector machine
Erdogdu Sakar et al. [Erdogdu Sakar, Serbes and Sakar (2017)]	k-medoids clustering-based attribute weighting and the selected attributes fed to the support vector machine
Peker et al. [Peker, Sen and Delen (2015)]	features selected by the maximum-relevance-minimum-redundancy method fed to complex-valued artificial neural network

The main focus of this study is to carry out the diagnosis of computer-assisted PD. Also, the main contribution of this study is to analyze the effects of parameters, i.e., neuron number and activation function of Deep Neural Networks (DNN) on a model performance. That is to say, the DNN models have been designed with the parameters assigned randomly as well. This study covers the finding of optimum hidden layer and hidden neuron numbers and ideal activation function and optimization function for the best DNN model. It is aimed to improve the predictive performances of models based on growing and pruning approach. DNN has been quite popular lately in the machine learning and data mining studies, i.e., [Huang, Zhu and Siew (2004); Men, Fu, Yang et al. (2018)].

The rest of this manuscript is organized as follows. Section 2 presents the related works constructed by using DNN topologies. Section 3 introduces the proposed approach. Section 4 addresses the experiments and performance evaluations of the DNN models. Finally, Section 5 draws discussion and conclusion.

2 Literature review

There are several studies in the literature for the diagnosis of PD which is a very important disease. Gupta et al. examined optimal subset of features by utilizing the optimized cuttlefish algorithm based on the conventional cuttlefish algorithm for diagnosis of Parkinson's disease at its early stage. The study is stable with an approximately 94% accuracy [Gupta, Julka, Jain et al. (2018)]. Sharma et al. proposed a new model named Modified Grey Wolf Optimization based on the traditional Grey Wolf Optimizer for detecting the symptoms of Parkinson's disease at a premature stage. Random forest, k-nearest neighbor classifier and decision tree algorithms classified the selected features. Best prediction accuracy value of 94.83% was obtained using optimized Cuttlefish Algorithm [Sharma, Sundaram, Sharma et al. (2019)]. Mostafa et al. firstly applied the feature selection method named Multiple Feature Evaluation Approach; then, the performances of five machine learning algorithms were examined on selected features and all features within the frame of 10-fold cross-validation technique. The test results show that the proposed approach improves the performance of the models [Mostafa, Mustapha, Mohammed et al. (2019)]. Pereira et al. presented a comprehensive research about the computer-assisted PD diagnosis, and effect of recent works on PD identification, mainly from 2015 to 2016 [Pereira, Pereira, Weber et al. (2019)]. Lacy et al. performed a case study of Parkinson's disease patients. The features collected via wearable sensors were implemented to the Echo State Networks [Lacy, Smith and Lones (2018)]. Parisi et al. introduced a novel hybrid Artificial Intelligence-based classifier in order to help early diagnosis of PD. Dysphonic measures and clinical score data were examined. Feature selection is carried out with Weights derived from a Multi-Layer Perceptron; then, 20 selected features were sent as input data to the Lagrangian Support Vector Machine [Parisi, RaviChandran and Manaog (2018)]. Gupta et al. presented a study on maximizing the accuracy and minimizing a number of features selected by using optimal subset of features based on optimized version of crow search algorithm in order to predict the Parkinson's disease [Gupta, Sundaram, Khanna et al. (2018)]. Gottapu et al. applied deep learning algorithms on Parkinson's Progression Markers Initiative dataset [Gottapu and Dagli (2018)]. Banks et al. examined the effects of non-motor risk factors on freezing of gait for PD. They applied several cognitive (Executive function, visuospatial function, processing speed, learning and memory tests) and non-cognitive tests (Rapid eye movement sleep behavior disorder, depression and anxiety scales). Also, they revealed the time of freezing of gait concerned with baseline processing speed, learning and sleepiness scores [Banks, Bayram, Shan et al. (2019)]. Recently, Sakar et al. have presented a comprehensive study, which is the tunable Q-factor wavelet transform method to the voice signals of PD patients for the feature extraction. This method has a higher frequency resolution than classical discrete wavelet transform. They compared the performances of the tunable Q-factor wavelet transform and conventional feature extraction methods for diagnosis of PD using vocal disorders. For this purpose, the feature sets were sent as input data to several classifiers and ensemble learning algorithms [Sakar, Serbes, Gunduz et al. (2019)].

3 Material and methods

3.1 Data collection

The dataset has 756 instances that were collected from 188 patients with PD and 64 healthy individuals. Three voice recordings were taken from each individual. The dataset consists of 754 attributes obtained by diversified speech signal processing algorithms such as Time Frequency Features, Mel Frequency Cepstral Coefficients in order to extract clinically valuable information for PD. The patients' data, which consist of 107 men and 81 women with ages between 33 and 87, have been collected from Department of Neurology in Cerrahpaşa Faculty of Medicine, Istanbul University. The healthy individuals' data consist of 23 men and 41 women with ages between 41 and 82 [Sakar, Serbes, Gunduz et al. (2019)].

3.2 Deep neural networks (DNN)

Artificial Neural Networks (ANN) is a computational algorithm inspired from biological neural networks in human brain process in order to solve prediction problems in computer vision, data mining etc. [Kumar and Sharma (2014)]. The theoretical foundations of DNN topology are well rooted from Artificial Neural Networks. DNN, which is general deep framework for classification or regression analysis, is a very popular learning algorithm achieving successful results by making inferences from a dataset [Ravi, Wong, Deligianni et al. (2017)]. Different topologies have been conducted by designing different deep learning algorithms and approaches. It will continue to be very popular for a long time in computer science and other multi-disciplined areas. In comparison with other traditional learning methods, deep neural networks are a powerful tool in machine learning studies such as pattern recognition and natural language processing.

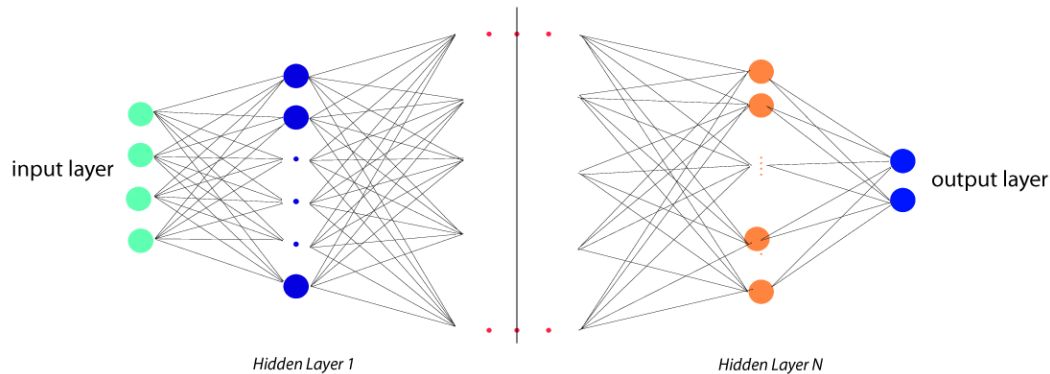


Figure 1: An overview of the DNN topology [Ravi, Wong, Deligianni et al. (2017)]

Generally, a DNN topology consists of an input layer, several hidden layers, and an output layer. An overview of typical DNNs topology is given in Fig. 1.

3.2.1 Activation function

The activation function determines whether a neuron is active for the target variable. The bias parameter adds to the sum of multiplies of the inputs and weights to help learn patterns. In order for the DNN to understand non-linear properties, an activation function is applied to the result. In this study, ‘tanh’, ‘relu’, ‘sigmoid’, ‘softmax’ and ‘elu’ activation functions were used.

3.2.2 Optimization function

The error rate of the designed DNN model is measured by the loss function defined in the last layer of the model. This function calculates how different the prediction of the model is from the actual value. An optimization function seeks to minimize a loss function. The optimal weight parameters for the nonlinear solution are calculated by the optimization function. In this study, ‘Stochastic Gradient Descent (SGD)’, ‘RMSprop’, ‘Adagrad’, ‘Adam’, ‘Adadelata’, ‘Adamax’ and ‘Nadam’ optimization functions were used.

3.3 Growing and pruning approach

In growing approach, a topology is designed with minimum hidden neuron. By applying growing criteria, new layers and neurons are added to the topology. In pruning approach, the topology is designed with maximum hidden neurons and the model is pruned. In this context, the best combination of the growing and pruning approach is tried to detect desired accuracy [Thoma (2017)]. The following steps are repeated until reasonable solutions for both approaches are achieved:

- 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 investigate for desired accuracy.

4 Modeling construction

Existing studies conducted based on original DNN topologies in literature usually present a model designed by using specific or random number of neurons and layers. The number of neurons in the hidden layers is a very important part of designing the ANN or DNN architectures. There are many studies, e.g., [Doukim, Dargham and Chekima (2010); Feng, Zhou and Dong (2019); Islam and Murase (2001); Sun (2012); Yuan, Xiong and Huai (2003)], focusing on this subject in the literature. Exactly there is no rule to decide the number of hidden neurons in each hidden layer. Using inadequate neurons in the hidden layers may result in under-fitting. Using excessive hidden neurons in the hidden layers may result in over-fitting, as well [Ke and Liu (2008)]. The number of inputs, outputs, architectures, activations, training sets, algorithms, and noises state the number of hidden neuron [Wang and Huang (2011); Zeng and Yeung (2006)]. Unlike these studies in literature, in this study the optimum the numbers of hidden layers and neurons were determined dynamically by using modified growing and pruning approach. The number of neurons in the input layer was accepted as the threshold value for the

maximum number of neurons in any layer. The number of neurons in the output layer was accepted as the threshold value for the minimum number of neurons in any layer.

This study mainly includes Step 1 (growing approach) and Step 2 (pruning approach) phases. Due to the nature of the growing and pruning approach, as long as the model performance boost in Step 1, the corresponding neuron is increased with each call of the growing method. In both phases, if the performance of the designed model is not boosting, a new layer is added. Also, when the number of hidden neurons is less than the number of input parameters, the workflow passes to Step 2. As long as the model performance boost in Step 2, the corresponding neuron number is reduced in each call of the pruning method. Finally, when the number of hidden neurons is not bigger than the target class number, the best model information is saved into the system.

Also, as known from literature, excessive use of neuron in the hidden layer(s) caused overfitting and therefore the rule given in Eq. (1) was applied.

$$\text{Hidden neurons} = \begin{cases} n^p, & n = 2, 3, \dots, p = 1, 2, \dots \\ & \text{if } n^p \leq 754 \end{cases} \quad (1)$$

As seen in Tab. 3, starting with ‘ p ’ value=1 and ‘ n ’ value=2, the ‘ p ’ value is increased by 1 in each step and the number of hidden neurons is calculated as n^p . If the number of hidden neurons is greater than the number of input parameters, a new layer with n^p is added to the neuron topology. From this stage, the ‘ p ’ value is reduced in each model design and the number of hidden neurons is calculated. This process is continued recursively until the hidden neuron count is smaller than the output parameter. Besides, in order to detect most appropriate ‘ n ’ value within 2 and 28, the number of hidden neurons for each layer was calculated according to Eq. (1) and topologies were designed accordingly. The criterion here is that the number of hidden neurons is smaller than the number of input parameters according to the growing and pruning approach. As long as this condition is met, new neurons or layers will be added in the topology.

Table 2: Confusion matrix

		Actual values	
		No	Yes
Predicted values	No	TN	FN
	Yes	FP	TP

Table 3: ‘*n*’ values range considering hidden neurons number

<i>p</i>									
n	1	2	3	4	5	6	7	..	10
2	2	4	8	16	32	64	128	..	*1024>754
3	3	9	27	81	243	729	*2187>754		
4	4	16	64	256	*1024>754				
..					
..					
9	9	81	729	*6561>754					
10	10	100	*1000>754						
..						
..						
27	27	729	*19683>754						
28	28	*784>754							

Note: 754 indicates input parameters number. ‘*’ indicates limit exceeding values for ‘*n*’ and ‘*p*’ since ‘*n*^{*p*}’ is bigger than 754.

As a technical detail, “One-Hot Encoding” was applied for target class in DNN models. The sigmoid function is used for the binary classification task. Predictive accuracy is one of the common measures employed in the machine learning and data mining studies. As can be seen in Eq. (2), the total number of true negative and true positive instances to the total number of instances is computed using this metric. True positive rate and false positive rate metrics were used for the receiver operating characteristic (ROC) curve results. These metrics given in Eqs. (3) and (4) are calculated by utilizing the confusion matrix presented in Tab. 2.

$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{FP} + \text{FN} + \text{TP}) \quad (2)$$

$$\text{True Positive Rate} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

$$\text{False Positive Rate} = \text{FP} / (\text{FP} + \text{TN}) \quad (4)$$

Herein, TP denotes the number of correctly predicted positive samples, TN denotes the number of negative predicted samples accurately, FP denotes the number of positive samples incorrectly predicted, FN denotes the number of negative samples incorrectly predicted.

DNN models were designed by using ‘Keras’ library with ‘Tensorflow’ backend. ‘Keras’ is a deep learning library that contains large collections of deep learning topologies. All coding was performed with Python 3.6 programming language.

Typical parameters defined for all DNN models were given in Tab. 4.

Table 4: Typical parameters of DNN models

Parameter	Value
epochs value	100
callbacks	True
batch_size	100

5 Results

The dataset was randomly split into train, validation and test sets by 60%, 20% and 20% respectively. In order to design the best DNN model, the growing and pruning approach was used for the determining of the numbers of neurons and layers. For fair comparison, all experiments were carried out on the same train, validation and test datasets. Based on multiple experiments, the study focused on the answers of the questions listed below to reveal the best performances dealing with the PD classification problem:

- How are the model parameters determined?
- Which activation and optimization functions must be used for the DNN model?

In this context, a number of experiments were conducted with DNN including different parameters such as activation function, hidden layer and hidden neurons. Modeling results having maximum accuracy obtained by applying different activation functions in hidden layers of DNN and applying different optimization methods on training of DNN were presented in Tab. 5. Also, Fig. 2 presents the charts of the best results by considering activation function and optimization functions. Based on the results, it can be observed that the best DNN model with ‘tanh’ activation function and ‘Adam’ optimization function was constructed by using 5 hidden layers, each of which consists of 3, 9, 27, 81 and 243 hidden neurons respectively. The parameters of the best topology are presented in Tab. 6 and this topology illustrated in Fig. 3.

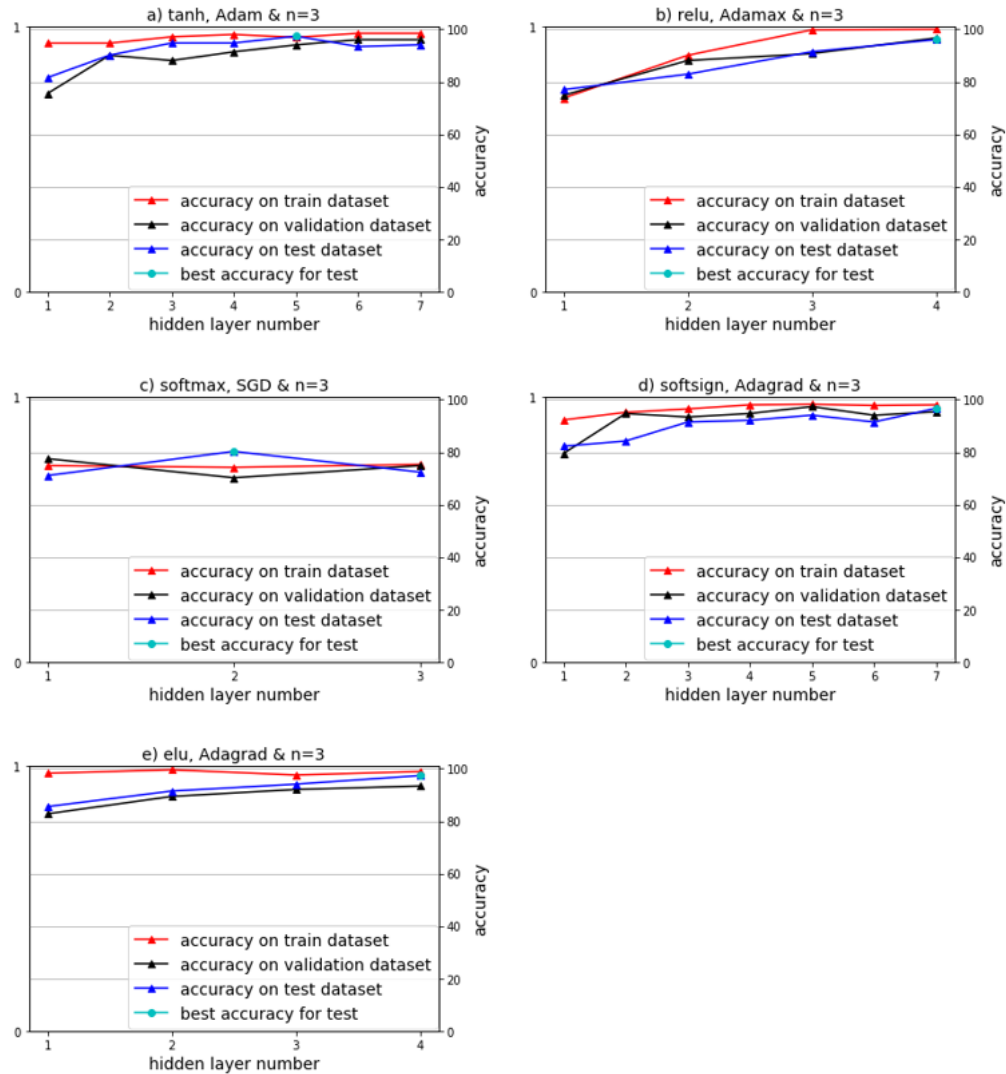


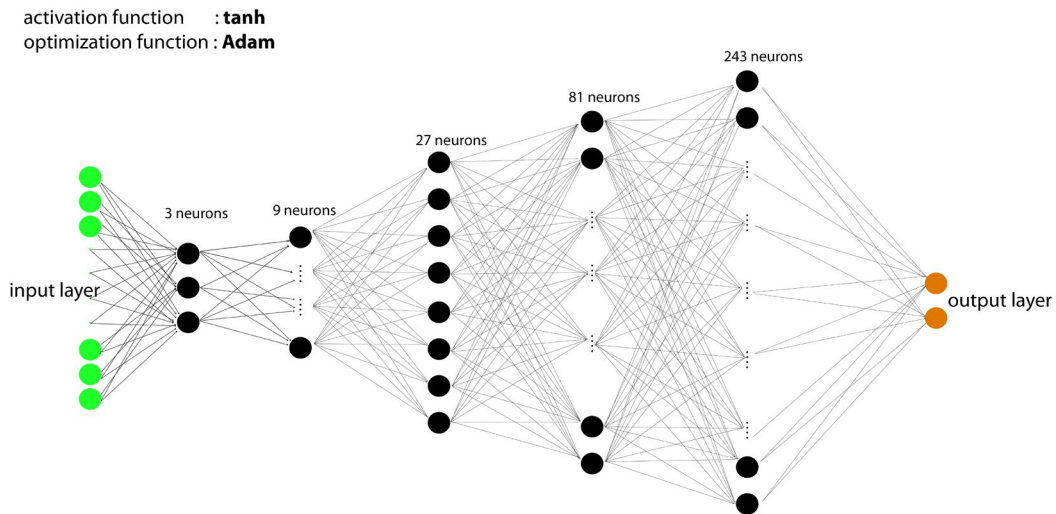
Figure 2: The graphical results for best topologies

Table 5: Hyper parameters information considered best topologies

Activation function	Optimization function	n value	Layer number	Train Accuracy	Validation Accuracy	Test Accuracy
tanh	Adam	3	5	98.45	92.72	99.34
relu	Adamax	3	4	100.0	96.69	96.05
softmax	SGD	3	2	74.17	70.2	80.26
softsign	Adagrad	3	7	98.01	95.36	96.71
elu	Adagrad	3	4	98.9	93.38	97.37

Table 6: The parameters information for the DNN model presented best performance

Parameters	Value
Number of input layer neurons	753
Number of hidden layers	5
<i>Ideal n value</i>	3
Number of hidden layer-1 neurons	3
Number of hidden layer-2 neurons	9
Number of hidden layer-3 neurons	27
Number of hidden layer-4 neurons	81
Number of hidden layer-5 neurons	243
Number of output layer neurons	2
Activation Function	'tanh'
Learning cycle	100 epochs
Learning Algorithm	Adam

**Figure 3:** The proposed DNN topology

The proposed DNN model presents 98.45%, 92.72 and 99.34% accuracy values on train, validation and test sub-datasets respectively. Hence, the accuracy results show us that the DNN model with 'tanh' activation function and 'Adam' optimization function is the best model for PD diagnosis.

ROC curve is used for visualizing the performance of a model based on true positive rate in machine learning and data mining approaches. Also, Area Under the Curve (AUC) indicates how well the positive instances are separated from the negative instances. AUC is a value between 0 and 1. A model whose predictions are 100% correct has 1.0 value of AUC. The one whose predictions are 100% wrong has 0.0 value of AUC [Fawcett

(2006)]. In this study, the ROC curve given in Fig. 4 was used to measure the quality and the effectiveness of the proposed DNN model. Fig. 4 illustrates the trade-off between true positive rate and false positive rate. The left corner depicts the percentages of the samples which were correctly detected as PD. The plots in this figure also show that the experiments have achieved high accuracy rates with minimal false rate on train, validation and test sub-datasets, respectively.

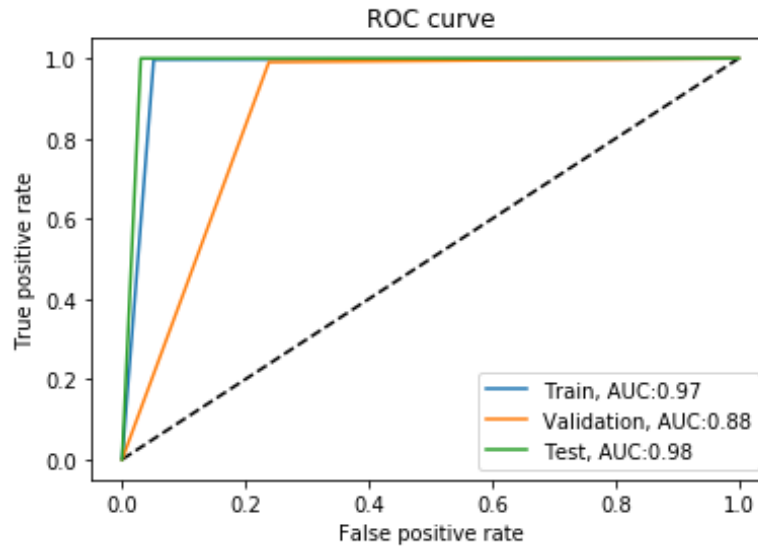


Figure 4: Performance of ROC curve of the proposed DNN model

6 Discussion

According to the results obtained from this study, Deep Neural Networks architecture, which offers the best performance with optimal activation and optimization function pair as well as optimal hidden layer and neuron number, the 5-layer DNN model, which uses 'tanh' activation function and Adam optimization function, offers the best performance with 99.34% accuracy on test dataset. There is no another study performed on the Parkinson disease dataset which is newly introduced to literature by Sakar et al. [Sakar, Serbes, Gunduz et al. (2019)]. For this reason, as shown in Tab. 7, the performance of proposed study is compared with Sakar et al. only [Sakar, Serbes, Gunduz et al. (2019)].

Table 7: Comparison of the studies

Studies	Techniques	Accuracy
Sakar et al. [Sakar, Serbes, Gunduz et al. (2019)]	tunable Q-factor wavelet transform based feature extraction and ensemble learning	86.0%
The proposed method-DNN base	DNN topology based on growing and pruning method	99.34%

7 Conclusion

Diagnosing of PD is very important because it is a serious disease that causes death. This study proposed an approach based on the DNN, which is a new machine learning and computational intelligence algorithm, in order to improve the efficiency of the computer-aid tools for PD diagnosis. In depth experimentations, the growing and pruning approach based DNN models were constructed in order to reveal the answers of the questions “Whether do hidden layers and neurons need to be tuned for improving learning?” and “What is the best model for PD diagnosis”. This approach provides the optimum activation function, neuron number and layer numbers for both topologies. Also, it gives ideal optimization function for DNN topology. By comprehensively comparing all models, the DNN model with ‘tanh’ activation function and Adam optimization function presented the best prediction performance on test dataset.

For further research, the answers of the questions such as ‘what is the ideal epoch value?’ or ‘what is the base value for growing pruning approach?’ will be investigated. Also, an intelligent predictor will be designed by integrating the DNN model presenting the best performance to solve real data which will be collected from a local hospital.

Conflicts of Interest: Kemal Akyol declares that he has no conflicts of interest to report regarding the present study.

References

- Banks, S. J.; Bayram, E.; Shan, G.; LaBelle, D. R.; Bluett, B.** (2019): Non-motor predictors of freezing of gait in Parkinson’s disease. *Gait & Posture*, vol. 68, pp. 311-316.
- Doukim, C. A.; Dargham, J. A.; Chekima, A.** (2010): Finding the number of hidden neurons for an MLP neural network using coarse to fine search technique. *10th International Conference on Information Science, Signal Processing and their Applications*, pp. 606-609.
- Erdogdu Sakar, B.; Serbes, G.; Sakar, C. O.** (2017): Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson’s disease. *PLoS One*, vol. 12, no. 8, e0182428.
- Fawcett, T.** (2006): An introduction to ROC analysis. *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874.
- Feng, S.; Zhou, H.; Dong, H.** (2019): Using deep neural network with small dataset to predict material defects. *Materials & Design*, vol. 162, pp. 300-310.
- Gottapu, R. D.; Dagli, C. H.** (2018): Analysis of Parkinson’s disease data. *Procedia Computer Science*, vol. 140, pp. 334-341.
- Gupta, D.; Julka, A.; Jain, S.; Aggarwal, T.; Khanna, A. et al.** (2018): Optimized cuttlefish algorithm for diagnosis of Parkinson’s disease. *Cognitive Systems Research*, vol. 52, pp. 36-48.
- Gupta, D.; Sundaram, S.; Khanna, A.; Ella Hassanien, A.; de Albuquerque, V. H. C.** (2018): Improved diagnosis of Parkinson’s disease using optimized crow search algorithm. *Computers & Electrical Engineering*, vol. 68, pp. 412-424.

Gürüler, H. (2017): A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural Computing and Applications*, vol. 28, no. 7, pp. 1657-1666.

Huang, G. B.; Zhu, Q. Y.; Siew, C. K. (2004): Extreme learning machine: a new learning scheme of feedforward neural networks. *2004 IEEE International Joint Conference on Neural Networks*, vol. 2, pp. 985-990.

Islam, M. M.; Murase, K. (2001): A new algorithm to design compact two-hidden-layer artificial neural networks. *Neural Networks*, vol. 14, no. 9, pp. 1265-1278.

Ke, J; Liu, X. (2008): Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction. *IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, pp. 828-832.

Kumar, P.; Sharma, P. (2014): Artificial neural networks-a study. *International Journal of Emerging Engineering Research and Technology*, vol. 2, no. 2, pp. 143-148.

Lacy, S. E.; Smith, S. L.; Lones, M. A. (2018): Using echo state networks for classification: a case study in Parkinson's disease diagnosis. *Artificial Intelligence in Medicine*, vol. 86, pp. 53-59.

Little, M. A.; McSharry, P. E.; Hunter, E. J.; Spielman, J.; Ramig, L. O. (2009): Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015-1022.

Men, H.; Fu, S.; Yang, J.; Cheng, M.; Shi, Y. et al. (2018): Comparison of SVM, RF and ELM on an electronic nose for the intelligent evaluation of paraffin samples. *Sensors*, vol. 18, no. 1, pp. 1-17.

Mostafa, S. A.; Mustapha, A.; Mohammed, M. A.; Hamed, R. I.; Arunkumar, N. et al. (2019): Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. *Cognitive Systems Research*, vol. 54, pp. 90-99.

Nolden, L. F.; Tartavouille, T.; Porche, D. J. (2014): Parkinson's disease: assessment, diagnosis, and management. *Journal for Nurse Practitioners*, vol. 10, no. 7, pp. 500-506.

Parisi, L.; RaviChandran, N.; Manaog, M. L. (2018): Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Systems with Applications*, vol. 110, pp. 182-190.

Peker, M.; Sen, B.; Delen, D. (2015): Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. *Journal of Healthcare Engineering*, vol. 6, no. 3, pp. 281-302.

Pereira, C. R.; Pereira, D. R.; Weber, S. A. T.; Hook, C.; de Albuquerque, V. H. C. et al. (2019): A survey on computer-assisted Parkinson's disease diagnosis. *Artificial Intelligence in Medicine*, vol. 95, pp. 48-63.

Ravi, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J. et al. (2017): Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4-21.

Sakar, B. E.; Isenkul, M. E.; Sakar, C. O.; Sertbas, A.; Gurgen, F. et al. (2013): Collection and analysis of a parkinson speech dataset with multiple types of sound

recordings. *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828-834.

Sakar, C. O.; Kursun, O. (2010): Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of Medical Systems*, vol. 34, no. 4, pp. 591-599.

Sakar, C. O.; Serbes, G.; Gunduz, A.; Tunc, H. C.; Nizam, H. et al. (2019): A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. *Applied Soft Computing*, vol. 74, pp. 255-263.

Sharma, P.; Sundaram, S.; Sharma, M.; Sharma, A.; Gupta, D. (2019): Diagnosis of Parkinson's disease using modified grey wolf optimization. *Cognitive Systems Research*, vol. 54, pp. 100-115.

Sun, J. (2012): Learning algorithm and hidden node selection scheme for local coupled feedforward neural network classifier. *Neurocomputing*, vol. 79, pp. 158-163.

Thoma, M. (2017): *Analysis and Optimization of Convolutional Neural Network Architectures (Master's Thesis)*. Karlsruhe Institute of Technology, Karlsruhe, Germany.

Tsanas, A.; Little, M. A.; McSharry, P. E.; Ramig, L. O. (2011): Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of the Royal Society Interface*, vol. 8, no. 59, pp. 842-855.

Wang, X.; Huang, Y. (2011): Convergence study in extended kalman filter-based training of recurrent neural networks. *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 588-600.

Yuan, H. C.; Xiong, F. L.; Huai, X. Y. (2003): A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy. *Computers and Electronics in Agriculture*, vol. 40, no. 1-3, pp. 57-64.

Zeng, X.; Yeung, D. S. (2006): Hidden neuron pruning of multilayer perceptrons using a quantified sensitivity measure. *Neurocomputing*, vol. 69, no. 7-9, pp. 825-837.