



ASSESSMENT OF PARKINSON'S DISEASE PROGRESSION USING NEURAL NETWORK AND ANFIS MODELS

*J. Hlavica**, *M. Prauzek**, *T. Peterek[†]*, *P. Musilek[‡]*

Abstract: Patients suffering from Parkinson's disease must periodically undergo a series of tests, usually performed at medical facilities, to diagnose the current state of the disease. Parkinson's disease progression assessment is an important set of procedures that supports the clinical diagnosis. A common part of the diagnostic train is analysis of speech signal to identify the disease-specific communication issues. This contribution describes two types of computational models that map speech signal measurements to clinical outputs. Speech signal samples were acquired through measurements from patients suffering from Parkinson's disease. In addition to direct mapping, the developed systems must be able of generalization so that correct clinical scale values can be predicted from future, previously unseen speech signals. Computational methods considered in this paper are artificial neural networks, particularly feedforward networks with several variants of backpropagation learning algorithm, and adaptive network-based fuzzy inference system (ANFIS). In order to speed up the learning process, some of the algorithms were parallelized. Resulting diagnostic system could be implemented in an embedded form to support individual assessment of Parkinson's disease progression from patients' homes.

Key words: *Parkinson's disease, speech signal, artificial neural networks, error back-propagation, fuzzy logic, ANFIS, UPDRS*

Received: March 25, 2015

DOI: 10.14311/NNW.2016.26.006

Revised and accepted: April 13, 2016

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's [8]. It particularly affects older people, after the age of 60. Therefore, as the population in general grows older, new methods for early diagnosis are crucial to monitor the progression of PD and allow intervention.

*Jakub Hlavica, Michal Prauzek, Department of Cybernetics and Biomedical Engineering, FEI VSB-TU Ostrava, E-mail: jakub.hlavica@vsb.cz, michal.prauzek@vsb.cz

[†]Tomas Peterek, IT4innovations, VSB-TU Ostrava, E-mail: tomas.peterek@vsb.cz

[‡]Petr Musilek, Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, E-mail: petr.musilek@ualberta.ca

At the present time, there is no cure for PD [25]. It is a chronic neurological disease, defined by tremor, slow motion and stiffness. The neurons stop producing neurotransmitter dopamine, or produce very low level of this chemical mediator that facilitates movement coordination. In principle, PD symptoms begin to manifest after the amount of dopamine drops below a certain threshold, usually 20% of the original level [11]. Only 10% of cases are genetically based.

Individuals suffering from this disorder must regularly attend medical centers to have the progression of their illness assessed. This assessment is important so that the treatment can be modified in case of emerging complications. Hence, PD assessment in the current form a priori involves increased costs related to medical visits and other expenses. The solutions introduced in this contribution and other related studies suggest to perform PD assessment from home using an embedded device, avoiding aforementioned expenses.

This paper proposes PD assessment method based on methods of computational intelligence, particularly artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS). Neural networks can be used to develop accurate computational models. They are easy to train and, after training, can perform required mappings very fast. ANFIS requires more time for proper training, but is as fast as neural networks in operation. Although it has a great potential for PD assessment applications, it may require several implementation changes to improve its training speed.

In comparison with the other recently introduced PD assessment system, models considered in this paper provide similar results in terms of error function values and other statistical parameters. However, they can offer better operational performance, especially if implemented in form of embedded devices. In the current literature, there is only one alternative model that overcomes the methods proposed in this paper: Least Squares Support Vector Machine (LS-SVM) described in [9]. However, it has several drawbacks, namely the need to reformulate the assessment problem as a binary classification and long training times. In addition, SVM-based systems require large amounts of memory to store the parameters of the trained models and relatively powerful computing devices for their operational implementation. In contrast, models proposed in this contribution are easy to implement as embedded systems. This makes them good candidates for practical realization with a great potential for widespread applicability.

Solutions applied in our contribution, especially these based on neural networks, are also characterized by simplicity of their training process. This make it easy to train a large number of candidate systems and, in turn to increase the probability of finding an near-optimal set of parameters for an effective model.

This paper is organized into six sections. Section 2 provides the background information, including an overview of current PD assessment procedures and description of relevant learning models. Section 3 then outlines other recent approaches to PD assessment. Section 4 describes the data set used for model development and detailed description of applied computational models in terms of data processing and parameter optimization. Results of computational experiments designed to evaluate the performance of the developed models are described and analyzed in Section 5. Major conclusions are presented in Section 6, along with possible direction of future work.

2. Background

2.1 Parkinson’s disease progression assessment

PD manifests itself in many forms, for instance body tremor, muscle rigidity, severe motor impairment, facial expression problems, unintelligible speech etc. Currently, the assessment of PD is performed rather subjectively by specialists on neurodegenerative disorders at medical centers. The assessment itself consists of several empirical tests [25] and the results are then mapped to the Unified Parkinson’s Disease Rating Scale (UPDRS) designed to indicate the severity of the disease.

For untreated patients, the UPDRS, also referred to as “Total UPDRS”, values range from 0 to 176, where 0 means a healthy state and 176 represents complete disability of a patient. “Total UPDRS” comprises three categories:

1. Mentation, behavior, and mood,
2. Activities of daily living,
3. Motor abilities.

Only the third category, also referred to as “Motor UPDRS”, ranging from 0 to 108, is considered in this study. In particular, we address the speech disorder that is assessed between 0 (normal) and 8 (unintelligible) [9]. This aspect of the disease is suitable for non-invasive telemonitoring, which is an emerging option for PD assessment [25].

Recently, there have been introduced several approaches to Parkinson’s disease progression assessment (PDPA) that are based on various computational methods. Some of these approaches are based on speech signal analysis. One of such approaches, developed in collaboration with Intel, takes the form of an embedded system called at-home testing device (AHTD). It is composed of loudspeakers, a high-end microphone, a display, and a number of motor impairment assessment devices [12]. Over the last few years, AHTD was distributed to a number of PD patients enrolled in a research project to develop novel PD assessment methods. The use of AHTD yielded a number of valuable data sets, including the speech signal data set used in this study. This data set is described in detail in Section 4.1.

2.2 Computational models

Computational models considered in this contribution include ANNs and ANFIS. They both implement learning methods capable of finding nonlinear mappings. This capability is important for the intended application, since the speech signal measurements used for PD progression assessment include a number of nonlinear patterns.

2.2.1 Artificial neural networks

Artificial neural networks represent a broad class of computational models inspired by operation of biological networks found in the central nervous systems and brains of animals. They can be used to approximate unknown mappings of large number of inputs, such as the speech signals considered in this study. They are usually

presented as an interconnected system of simple information processing elements called neurons. Due to their adaptive nature, neurons can learn to recognize input patterns and assign them appropriate values in the output space.

The neural networks applied in this study are feed-forward multilayer neural networks with error backpropagation learning. They consist of one or more hidden layers of nonlinear neurons along with a linear output layer [10]. In the standard *backpropagation* learning algorithm, weights and biases in the network are adjusted by following the negative direction of the gradient of the sum of squared errors (SSE) with respect to the weight variables. The update equation can be represented as

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \frac{1}{2} \frac{\partial \mathbf{e}^t \mathbf{e}}{\partial \mathbf{x}}, \quad (1)$$

where η is learning rate parameter, \mathbf{x} is the input vector, and $\mathbf{e} = \mathbf{o} - \mathbf{o}_d$ is the error calculated as the difference between network's output \mathbf{o} and desired output \mathbf{o}_d .

One of the most effective neural network learning procedures is the *Levenberg-Marquardt* (L-M) algorithm [23]. This algorithm provided the best performance for training neural networks considered in this study. It can be thought of as a combination of the steepest descent and Gauss-Newton method. When the current solution is far from a local minimum, the algorithm behaves like the steepest descent method that is slow, but has guaranteed convergence. When the solution is close to the local minimum, it becomes the Gauss-Newton method and exhibits fast convergence [17].

Using L-M algorithm, weights are updated as follows:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - (\mathbf{J}_n^T \mathbf{J}_n + \mu \mathbf{I})^{-1} \mathbf{J}_n^T \mathbf{e}_n, \quad (2)$$

where \mathbf{J} is the Jacobi matrix [23]. Marquardt parameter, μ , ensures that the (local) minimum will always be found [14]. When μ is very large, the produced direction is close to that of the steepest descent. When $\mu = 0$, search for the minimum of error function is reduced to the Gauss-Newton direction [7].

Another powerful procedure is called *resilient backpropagation* (commonly denoted as RPROP). In contrast to other gradient descent algorithms, RPROP does not use the magnitude of the gradient $\partial \mathbf{E} / \partial \mathbf{w}$, but only its sign [3]. When the gradient has the same sign for two successive iterations, the update factor is increased. If the gradient changes the sign, the update factor is decreased [10]. This algorithm ranked second (not considering cascade modification of L-M) in terms of error and correlation for the problem of mapping speech signal measurements to the motor UPDRS.

2.2.2 Adaptive neuro-fuzzy inference system (ANFIS)

ANFIS integrates the layered distributed structure of neural networks with the principles of Takagi-Sugeno fuzzy inference systems. It combines the benefits of both approaches in a single learning system capable of approximating complex nonlinear relationships [19]. ANFIS inference can be described by a set of n fuzzy IF-THEN rules

$$\text{IF } x \text{ is } A_i \text{ AND } y \text{ is } B_i, \text{ THEN } f_i = p_i x + q_i y + r_i, \quad i = 1, \dots, n,$$

where x and y are the inputs, A_i and B_i are the fuzzy sets, f_i are the outputs within the fuzzy region specified by the fuzzy rules, and p_i , q_i and r_i are adjustable parameters determined during training [1]. The training process is facilitated by algorithms similar to procedures used for training of neural networks.

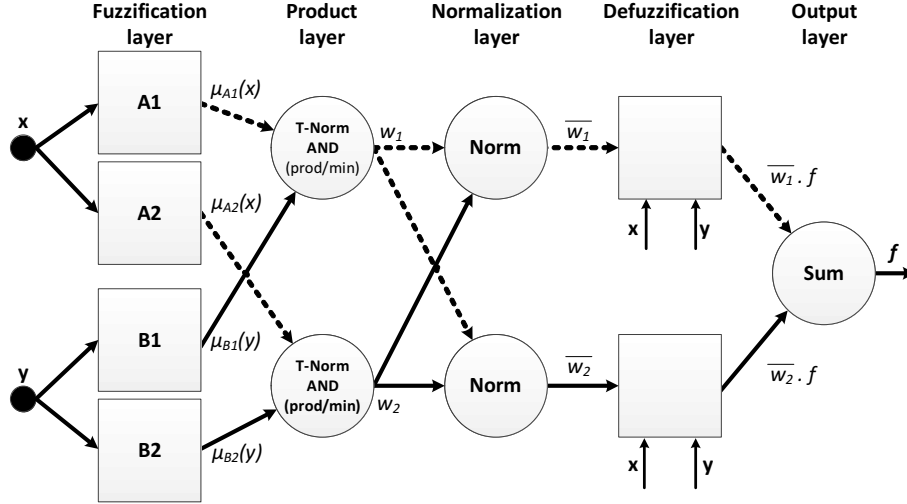


Fig. 1 ANFIS structure [21].

The architecture of the ANFIS model is illustrated in figure 1. It is composed of five layers that subsequently process inputs of the system. The *fuzzification layer* converts crisp input variables to membership values of predefined linguistic variables. Each node in this layer has adaptive parameters that determine the shape of fuzzy sets corresponding to the individual linguistic terms, e.g. parameters $\{a_i, b_i, c_i\}$ of the generalized bell membership function $\mu_A(x) = 1/(1 + |(x - c_i)/a_i|^{2b_i})$. The *product layer* contains fixed nodes that combine the incoming signals, $w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y)$. The nodes of the *normalization layer* are also fixed. They form so called normalized firing strengths $\bar{w}_i = w_i / \sum w_j$, where index j is taken over all terms corresponding to rule i . The *defuzzification layer* has adaptive nodes with consequent parameters $\{p_i, q_i, r_i\}$ that determine one crisp output value for each rule, $\bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$. Contributions of individual rules are finally combined in the *output layer* to form the net output of the model

$$f = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}, i = 1, 2.$$

ANFIS uses a hybrid learning algorithm that combines gradient descent and least squares estimation (LSE) methods to identify linear and nonlinear parameters of the system [6]. The algorithm performs two passes: a *forward pass* that considers the premise parameter fixed and adapts the consequent parameters using LSE method, and a *backward pass* that propagates the error signals backward to update the premise parameters using the gradient descent method [15].

3. Related work

One of the first studies on the use of speech signal measurements for telemonitoring of PD was published in 2009 [20]. In 2010, a study by Tsanas et al. [25] implemented several methods for mapping the results of speech signal analysis (obtained via AHTD) to the UPDRS clinical scale: least squares (LS), iteratively reweighted least squares (IRLS), least absolute shrinkage and selection operator (LASSO) and classification and regression trees (CART). The best models predict the value of Motor UPDRS with mean absolute error (MAE) of 6.80 ± 0.17 (for IRLS method) and 5.95 ± 0.19 (for CART). The values of mean squared error (MSE) have not been disclosed. The article also includes correlation analysis suggesting that it is not necessary to use all data set speech signal features to be able to map the UPDRS values accurately [25]. This work was later extended, resulting in an improvement of mapping motor UPDRS values with MAE of 6.57 ± 0.16 (for LASSO) [26].

Several new methods for PD assessment were introduced in 2012. Least Squares Support Vector Machine (LS-SVM) with the UPDRS mapping errors of 4.96 (MAE), 39.47 (MSE) and correlation coefficient 0.63. This is, at the present time, the best performing model (in terms of accuracy). Other introduced methods include General Regression Neural Network (GRNN) with MAE of 6.36 and standard Support Vector Machine (SVM) model (MAE 5.46) [9]. All these methods were later slightly improved using a logarithmic transformation [26, 9].

Application of data set restricted using feature selection resulted in UPDRS values of 5.81 and 5.66 respectively (MAE). Feature selection was based on evaluation of several SVM models where individual speech signal features were subsequently added to the training data set, in order to detect their contribution to model accuracy [2].

Parallel neural networks can be applied for PD assessment as well. In such systems, each network is presented with the full training set and as training progresses the first network propagates the data samples which it could not accurately predict to the second network. The second network, connected to the first one in parallel, is now trained with the full training data set and the additional data samples propagated from the first network; and so on. The best outputs are then selected using a rule-based system. With the increasing number of neural networks connected in parallel, the MSE value is decreasing, until a certain threshold is reached [24].

Other possible solution regarding vocal tract diagnosis is presented in [4] and is based on artificial neural network version of Linear Prediction Coder. As for the speech recognition in general, the modern attitudes tend to apply deep bidirectional Long Short-Term Memory recurrent neural networks and deep neural networks [13, 18].

4. Development of new PDPA models

4.1 Data description

Speech samples used in this study were extracted from the “Parkinsons Telemonitoring Data Set” [25] that is currently the most comprehensive database of such measurements. As the research was conducted in the US, the speech samples are in

English language. The generalization of the solution proposed in our paper would require to take into account the particularities of the national languages.

The speech signal measurements were recorded using high-end microphones connected to AHTD, sampled at 24 KHz with 16-bit resolution.

The original data set is a matrix 5875×16 containing 5875 measurements composed of 16 speech signal parameters, including 13 linear and 3 nonlinear signal characteristics. The linear characteristics include shimmer representing variance of consecutive speech signal amplitudes, and jitter describing variance of speech signal fundamental frequencies. The nonlinear speech signal parameters represent the rate of vibration of vocal chords. The data set also contains patient identification, such as ID, age, sex etc. The identification data is not explicitly used in this study and are omitted in calculations.

The original data set was divided into two groups. The first group contains 85% of the original data set and is used to train, validate and test the models of neural networks in order to adjust their structure and parameters. The second group, 15% of the original data set, is intended for an independent test used as an objective model comparison criterion to compare computational models based on various adaptation methods. The data division is performed on both inputs and targets. Targets are separate from inputs to emphasize that they were linearly interpolated before preprocessing performed in our research.

Both training and independent test data sets have very similar shapes of distribution functions and histograms. As a result, the independent test subsets capture all important features of the original data set, such as minimum and maximum values, variation etc.

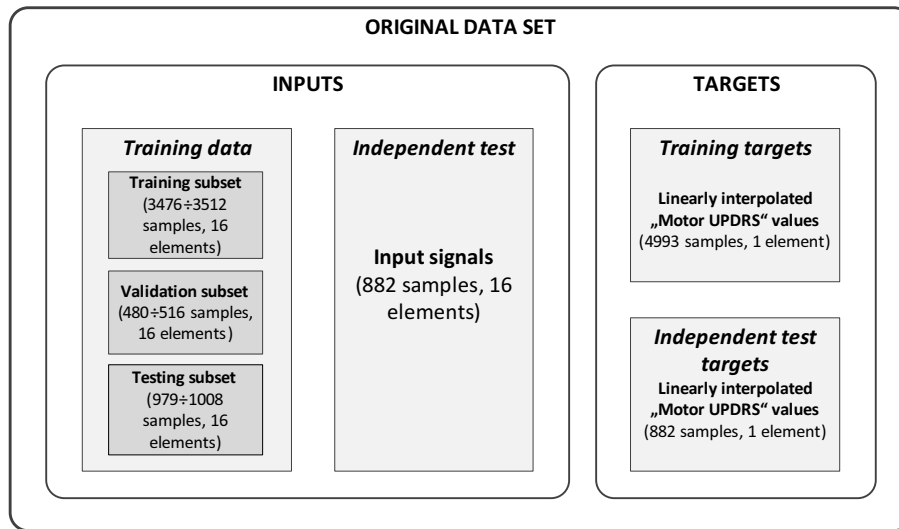


Fig. 2 Parkinson’s telemonitoring data set partition.

There is a set of static indices defined to prevent random data division for the needs of cross-validation. This assures that the computational models always have the same initial conditions. However, these indices do not divide the data equally

Parameter	Description
Jitter [%]	Average absolute difference between consecutive periods, divided by the average period.
Jitter [Abs]	Average absolute difference between consecutive periods, in seconds.
Jitter:RAP	Relative Average Perturbation, the average absolute difference between a period and the average of it and its two neighbors, divided by the average period.
Jitter (PPQ5)	Five-point Period Perturbation Quotient, the average absolute difference between a period and the average of it and its four closest neighbours.
Jitter (DPP)	Average absolute difference between consecutive differences between consecutive periods, divided by the average period.
Shimmer	Average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.
Shimmer [dB]	Average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20.
Shimmer:APQ3	Three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude.
Shimmer:APQ5	Similar to APQ3 (five-point).
Shimmer:APQ11	Similar to APQ3 (eleven-point).
Shimmer:DDA	Average absolute difference between consecutive differences between the amplitudes of consecutive periods.
NHR	Noise-to-harmonics ratio.
HNR	Harmonics-to-noise ratio.
RPDE	Recurrence Period Density Entropy - determines the repetitive signal.
DFA	Detrended Fluctuation Analysis - signal fractal scaling exponent.
PPE	A nonlinear measure of fundamental frequency variation.

Tab. I Description of speech signal parameters [5].

in terms of matrix size. Therefore, some cross-validation iterations contain slightly different number of data points than others. The relative sizes of the subsets are 70% for training, 10% for validation and 20% for testing.

4.2 PDPA models based on artificial neural networks

The topology and parameters of neural networks have significant impact on their learning and on the performance of the resulting models. However, there are no clear guidelines or rules for their selection. One approach is to execute a number of trials with different topologies and parameter values, and select the most effective

configurations based on their relative comparison. In this work, this approach is implemented through a systematic variation of selected network and learning parameters within predetermined ranges.

Learning parameters of feed-forward multilayer neural networks depend on the selected learning algorithm. For instance, in Levenberg-Marquardt algorithm, the parameter set consists of the size of the Marquardt parameter, its maximum value, and its increasing and decreasing ratio, while the learning rate parameter, typical for heuristic algorithms, is not used. Other learning algorithms may have entirely different sets of parameters.

Network topology is defined by the number of hidden layers, number of neurons in the layers and activation function(s) of the neurons. Ranges of network parameters used in this study are described in Section 5.1.

In general, neural networks are very sensitive to the initial values of weights and biases. As these values are usually assigned randomly, learning often produces disparate results even when the same settings and training data are used. To increase the chance of finding a (near) optimal solution, neural networks should be trained several times with different random initializations. To perform a large number of learning runs, it is advantageous to perform them in parallel on a powerful computing cluster. The scheduling algorithm implemented in this study is based on a modified *mpirexec* library. Majority of performed simulations were run on a 3-node cluster with 16 processors per node. This configuration allows to execute 48 concurrent training runs in parallel.

The data processing chain of the main neural network training program consists of five functional blocks shown in Fig. 3. Block ‘Data preprocessing’ defines all required and optional input arguments, checks the input vectors and matrices, and reshapes them as necessary. In addition, it removes any redundant matrix rows and normalizes the data to the range $[0, 1]$.

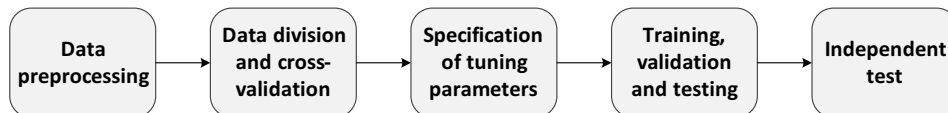


Fig. 3 *Neural network data flow.*

‘Data division and cross-validation’ block prepares data subsets for 10-fold cross-validation. Data is selected according to predefined indices to ensure that all folds have comparable initial conditions, including the training data. Block ‘Specification of tuning parameters’ provides definition of the neural network settings, such as the number of training epochs, the default Marquardt parameter and its increase/decrease ratio, or the resilient backpropagation update factor.

The next functional block performs training, validation and testing. The last block then implements independent testing that allows comparison of different network topologies and learning algorithms. Only data set aside for independent testing are used at this step, as explained in Section 4.1.

4.3 PDPA models based on ANFIS

As mentioned earlier, ANFIS model is a very powerful tool for signal processing applications. However, before applying ANFIS to speech signal measurements, they must be processed and refined in a number of functional blocks, as shown in Fig. 4.



Fig. 4 ANFIS data flow.

In block ‘Data preprocessing’, the speech signal data set enters the processing stream and training parameters are set. These parameters include the number of training epochs and the type and number of membership functions (to adjust the granularity of the system). Optionally, a set of speech signal values for an independent test can be provided as well.

The number of fuzzy rules, N_{FR} , can be represented as

$$N_{FR} = (N_{MF})^{N_{IN}}, \quad (3)$$

where N_{MF} is the number of fuzzy membership functions and N_{IN} is the number of input elements (input vectors). The default maximum number of fuzzy rules is restricted to $N_{FR} = 256$ to limit the computational burden caused by the curse of dimensionality [16]. For illustration, ANFIS model with 256 rules requires approximately 17 hours of training.

Block ‘Initialization of FIS structure’ creates the fuzzy inference system (FIS). This system is then adapted in the training process described earlier. The default logical operator is AND, implemented in the product layer of the ANFIS model. Block ‘Training and testing’ performs adaptation of system parameters using the hybrid algorithm described in Section 2.2.2. The best FIS representing the most effective ANFIS model is exported into a file for future use. Finally, block ‘System’s outputs calculations’ determines the outputs of the system for the training and testing data sets. Detailed description of application and training settings of the ANFIS model can be found in Section 5.2.

5. Results and analysis

The new computational models designed to assess Parkinson’s Disease progression were tested on the speech signal measurements data set described in Section 4.1.

5.1 Artificial neural networks testing

The developed artificial neural networks models were tested using an iterative approach and cross-validation. These two training procedures allow, respectively, to execute a large number of simulations with different architecture parameters and

to reduce the influence of initial weights and biases on network performance. The transfer functions used in the first layer of neurons were logarithmic sigmoid (ls), hyperbolic tangent (ts) and radial basis (rb). The outputs of the second (output) layer were calculated using linear transfer function.

The ratio between the numbers of neurons in the first and the second hidden layer (input layer is labeled as ‘0’) was iteratively decreased as follows: 1/0, 1/2, 1/3 and 1/4, where 0 means that the second layer of neurons was not implemented. Thus, neural network architectures having one hidden layer, two and both layers were simulated. The number of neurons in the first hidden layer was varied within the range of 8–96 in increments of 8 neurons. Therefore, each combination of neural network parameters was simulated 48 times, with 12 combinations of number of neurons for each ratio.

In total, approximately 70000 neural networks simulation runs were executed. Most configurations were evaluated using the three-fold cross-validation. The five-fold cross-validation was used only with Levenberg-Marquardt training algorithm, and ten-fold cross-validation only for the final validation of the best performing computational models (i.e. the models listed in Tab. II).

All neural networks simulations were performed in a batch mode, i.e. using cumulative weight updates after presenting all data samples. Tuning parameters were adjusted for each implemented training algorithm. As an example, for Levenberg-Marquardt numeric method, the Marquardt parameter was defined in each simulation and incremented or decremented afterwards, as well as the value of increment/decrement; this way, it was possible to find the optimal network that would reach the minimum of an error function (mean squared error, MSE). Each training algorithm uses its own parameters that are selected by the assessment software.

Algorithm	Layers	Fcn	Epch	Time	MSE	MAE	Corr
Conjugate-Beale	80/27	rb/rb	152	24	47.40	5.61	0.57
Conjugate-Fletch.	48/16	rb/rb	178	17	47.72	5.65	0.57
Conjugate-Polak	48/12	rb/ls	217	22	48.14	5.67	0.56
Scale conjugate	80/27	rb/ls	222	16	47.63	5.65	0.57
Resilient BP	96/24	ts/ls	230	11	46.55	5.55	0.58
Var. learn. rate	40/14	rb/ts	93	2	63.47	6.76	0.33
Momentum	40/14	rb/ls	241	5	58.50	6.41	0.42
Rate and mom.	80/20	rb/rb	108	4	56.56	6.23	0.45
One-step secant	80/40	rb/rb	156	38	49.83	5.83	0.54
Levenberg-Marq.	16/6	ls/ls	58	18	44.69	5.33	0.61
Cascade L.-M.	32	ts	54	29	46.23	5.49	0.59
Rapidminer BP	24/8	ls/ls	500	580	48.78	5.68	0.57

Tab. II Mapping speech signal measurements to Motor UPDRS: neural networks (Algorithm: algorithm used for training, Layers: number of neurons in the first/second layer, Fcn: type of transfer function in the first/second layer, Epch: number of training epochs, Time: training time in seconds, Results of independent tests in terms of mean square error (MSE), mean absolute error (MAE) and correlation coefficient (Corr)).

Results of the best models of neural networks trained by different training algorithms are shown in Tab. II. It can be seen that the best results were obtained by a neural network trained with Levenberg-Marquardt (L-M) algorithm. This network had 16 neurons in the first layer and 6 in the second (in total 381 adjustable parameters including neurons' weights and biases), both with logarithmic sigmoid transfer functions. L-M is not the fastest training algorithm, but in this particular application speed is not essential. The value of MAE is 5.326, better than most other published results.

An alternative to L-M is the neural network trained with Resilient backpropagation with 3985 adjustable parameters, having 96 neurons in the first layer and 24 in the second. This neural network does not require as much computational power and memory as L-M, but still provides satisfactory results, very close to the best performer. Neural network designed in data mining SW Rapidminer 5 provides only average results, although compared to L-M and Resilient BP the training itself is significantly slower and less flexible in terms of tuning parameters adjustments.

The results presented in Tab. II also confirm that heuristic training algorithms (such as the variable learning rate or momentum) and conjugate gradient algorithms work well with radial basis transfer function, while L-M and Resilient BP provide the best results for networks with sigmoidal transfer functions.

5.2 ANFIS testing

Compared to a typical neural network, ANFIS is based on a different type of information processing calculations that require more computing resources. In addition, in the current implementation, these calculations cannot be parallelized and thus the training of ANFIS models requires significantly more time. Furthermore, the processing model of ANFIS had to be simplified: while neural networks use 16 or more inputs, ANFIS models were limited to 8 inputs. Such simplification caused loss of information included in a speech signal data set and, therefore, the ANFIS models were not able to learn all patterns present in the training data. However, accepting this compromise was the only way to include ANFIS in this study while keeping training times at acceptable levels: neural networks used in this project typically required training time at the order of 1–10 seconds, but corresponding ANFIS models required up to 17 hours to train the models with required accuracy.

The following two sets of experiments were performed using ANFIS model:

- 5 inputs (elements) of speech signal:
 - 3 membership functions for each input,
 - 243 fuzzy rules,
 - adjustable parameters: 1503 (45 antecedent and 1458 consequent),
 - selected elements: *Shimmer*, *NHR*, *HNR*, *DFA* and *PPE*.
- 8 inputs (elements) of speech signal:
 - 2 membership functions for each input,
 - 256 fuzzy rules,

- adjustable parameters: 2352 (48 antecedent and 2304 consequent),
- selected elements: *Jitter(abs)*, *Shimmer*, *Shimmer:DDA*, *NHR*, *RPDE*, *DFA* and *PPE*.

The maximum number of training epochs was set to 40. Two types of membership functions were considered: Generalized bell-shaped (described in Section 2.2.2) and Gaussian. However, the second type did not provide satisfactory results.

Inputs	Rules	Oper.	MF	Epch	Time	MSE	MAE	Corr
8	256	prod	gbellmf	40	58355	53.30	5.90	0.50
8	256	min	gbellmf	40	55189	54.11	6.10	0.49
5	243	min	gbellmf	40	19956	58.66	6.15	0.44

Tab. III Mapping speech signal measurements to Motor UPDRS: ANFIS.

It can be seen from Tab. III that the best results were obtained by ANFIS models with 8 inputs and 256 fuzzy rules. In particular, these models have higher correlation coefficient and lower values of error measures. However, this configuration required approximately three times more training time compared to the models with 5 inputs and 243 rules.

It is also evident that the number of membership functions per input is not as important as the overall number of inputs. The Number of membership functions generally improves resolution of the system. However, given the high computational demands of ANFIS training, the use of larger number of inputs with lower resolution is a better choice.

5.3 Comparison of neural networks and ANFIS

Results presented in Tabs. II and III show that neural networks perform relatively better than the current implementation of ANFIS. However, due to their nature, ANFIS-based models have a considerable potential to achieve better accuracy. After training the implementation of both types of models is comparable (depending on the number of parameters in a particular case). Nevertheless, ANFIS requires only implementation of simple arithmetic and logical operations (addition, multiplication, minimum, maximum) while neural network requires additional implementation of nonlinear transfer function. In both cases, the compiled system can be implemented in form of a look-up table, further reducing the computational requirements [22].

Results of regression and correlation analysis of selected models are shown in Fig. 5. The plots represent the dependencies of target values and actual outputs of the models. The best fit (correlation coefficient $R = 0.6109$) was obtained using the neural network model trained with Levenberg-Marquardt optimization algorithm. Networks trained using Resilient backpropagation and conjugate gradient methods have correlation coefficients of 0.5840 and 0.5734, respectively.

The ANFIS model maps the speech signal measurements with less accuracy compared to the three selected neural network models. In addition to the lower value of correlation coefficient, this model produces several more significant outliers

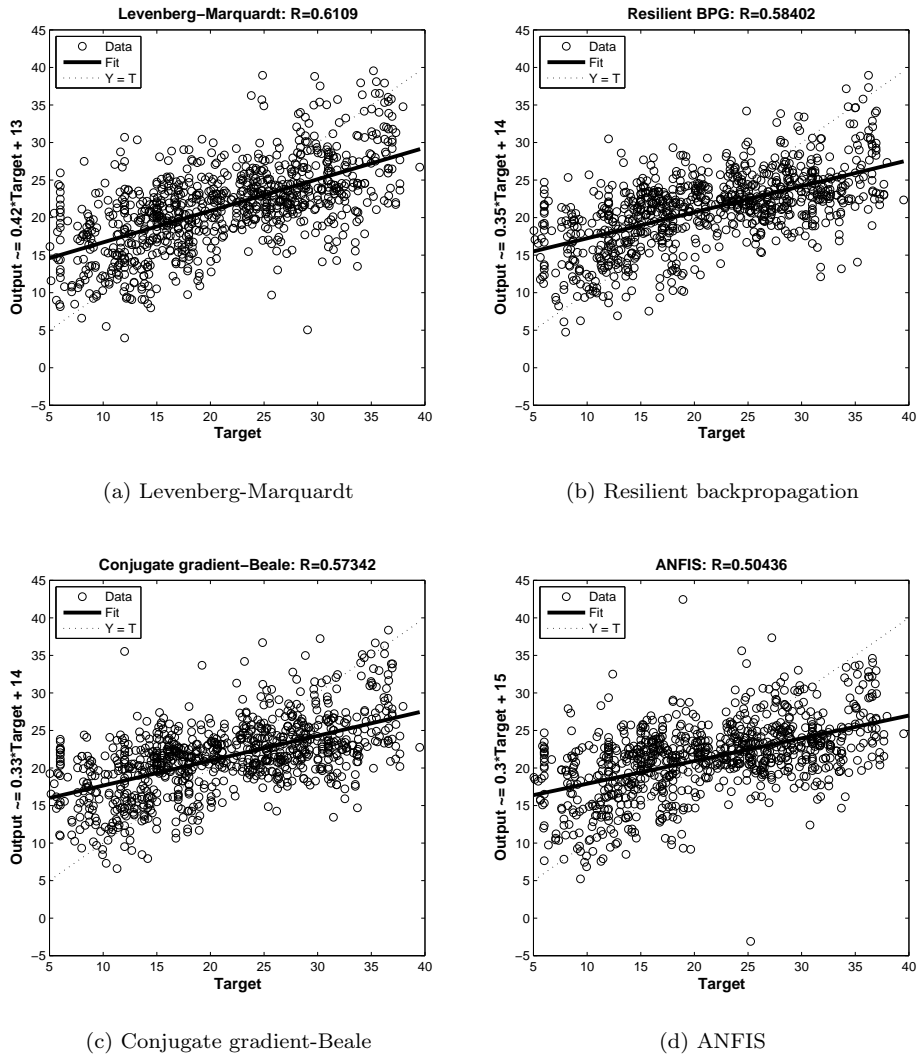


Fig. 5 Correlation analysis of the most effective models.

clearly visible in Fig. 5d. However, as mentioned earlier, the extent of training of ANFIS model was limited to avoid excessive training times. It is expected that with more comprehensive training, facilitated by future improvements of the efficiency of its implementation, ANFIS will provide results with accuracy comparable or exceeding that of the neural network models.

Fig. 6 displays box plots of absolute error values for the same computational models. Neural network with Levenberg-Marquardt algorithm performs estimations with considerably lower value of Interquartile Range (IQR) in comparison to other models. Regarding the number of its outliers, however, is the highest of all

presented models. IQR of the model trained with the resilient backpropagation reaches higher value compared to the previous model, while the number of outliers is reduced substantially. Although, the range between the upper and lower adjacent values (ends of whiskers) is significantly higher than the foregoing one. Box plot of the neural network model with conjugate gradient training shows that the median reaches the highest value of all models but generates only one outlier. ANFIS model follows a similar trend as model with Resilient backpropagation, with slightly higher value of IQR.

Overall, Figs. 5 and 6 show that although the four best-performing computational models were developed using different approaches, the results of their independent testing are fairly similar.

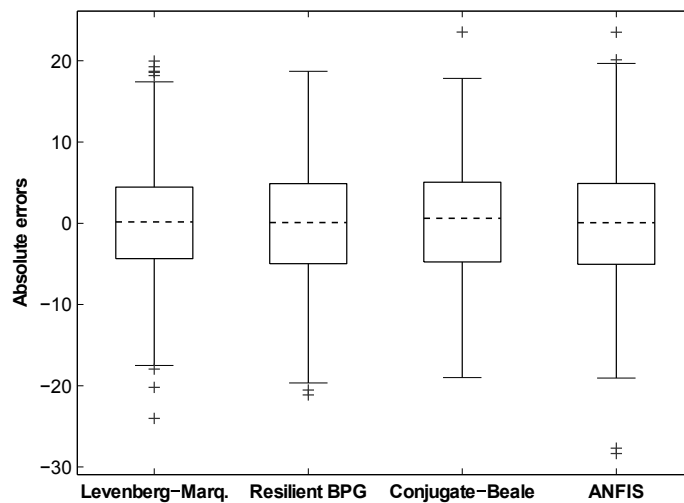


Fig. 6 Box plots depicting absolute errors.

6. Conclusions and future work

This study describes a system for mapping speech signal measurements to UPDRS values that represent the severity of the Parkinson's disease. A number of computational models were designed to carry out this task, and their performance was analyzed. In particular, models based on ANNs and ANFIS were implemented and evaluated.

All considered models perform the desired mapping with a reasonable accuracy. Presently, there is only one related model [9] that produces better results than the best model of neural network with Levenberg-Marquardt algorithm presented in this paper. The Least Squares Support Vector Machine (LS-SVM) [9] has errors MAE = 4.96, MSE = 39.47 and correlation coefficient $R = 0.63$. The best model presented in this paper produces MAE = 5.33, MSE = 44.69 and $R = 0.61$. All other mapping models presented in the current literature provide worse results.

To further improve the accuracy of models based on neural networks, the values of error function could be further lowered by prolonging the training process. However, due to overfitting, the networks could lose their ability to generalize and exhibit large errors for validation and testing data. Another alternative would be to provide a larger data set to train, validate and test the neural networks models. However, this would require expansion of current speech telemonitoring research studies by recruiting more PD patients - clearly a formidable task on its own. Finally, the available data could be preprocessed to remove outliers and thus to improve the consistency of the data set used for network training and validation.

The ANFIS model has a great potential to perform better, should all available data be used to adapt the model more effectively. However, this is not possible using the current ANFIS implementation that makes the adaptation process very time consuming. Training of a single model using all data would require hundreds of hours. Future implementation will concentrate on parallelization of ANFIS model so that individual nodes can operate simultaneously. This will speed up the adaptation process and, consequently, it will be possible to use the entire available data set for training, significantly improving the accuracy of the resulting model.

Acknowledgement

This work was supported by project SP2014/156, ‘Microprocessor based systems for control and measurement applications.’ of Student Grant System, VSB-TU Ostrava. It was also partly supported by projects IT4Innovations Centre of Excellence (CZ.1.05/1.1.00/02.0070, funded by Structural Funds of the European Union and state budget of the Czech Republic) and Bio-Inspired Methods: research, development and knowledge transfer project (CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic), and by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

- [1] AL-HMOUZ A., SHEN J., AL-HMOUZ R., YAN J. Modeling and Simulation of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for Mobile Learning. *IEEE Transactions on Learning Technologies*. 2012, 5(3), pp. 226–237, doi: [10.1109/TLT.2011.36](https://doi.org/10.1109/TLT.2011.36).
- [2] ASGARI M., SHAFRAN I. Predicting severity of Parkinson’s disease from speech. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 5201–5204, doi: [10.1109/IEMBS.2010.5626104](https://doi.org/10.1109/IEMBS.2010.5626104).
- [3] BAYKAL N., ERKMEN A. Resilient backpropagation for RBF networks. In: *Proceedings of the Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, Brighton, UK, 2000, pp. 624–627, doi: [10.1109/KES.2000.884125](https://doi.org/10.1109/KES.2000.884125).
- [4] BEDDOES M. Soft computing design for speech-based diagnosis of the human throat. In: *Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, Vancouver, CA, 2001, pp. 383–389, doi: [10.1109/NAFIPS.2001.944283](https://doi.org/10.1109/NAFIPS.2001.944283).

- [5] BOERSMA P., WEENINK D. *Praat: doing phonetics by computer* [online]. Amsterdam, Netherlands, 2011 [viewed 2016-05-05]. Available from: <http://www.fon.hum.uva.nl/praat/>.
- [6] CAI C., DU D., LIU Z. Battery state-of-charge (SOC) estimation using adaptive neuro-fuzzy inference system (ANFIS). In: *The 12th IEEE International Conference on Fuzzy Systems, FUZZ '03*, 2003, pp. 1068–1073, doi: [10.1109/FUZZ.2003.1206580](https://doi.org/10.1109/FUZZ.2003.1206580).
- [7] CHEN W., GONG Q., ZHANG L., YANG M. Parameters identification of twelve-phase synchronous generator based on Levenberg-Marquardt algorithm. In: *Proceedings of the 11th International Conference on Electrical Machines and Systems, ICEMS 2008*, Wuhan, China, 2008, pp. 3992–3995,
- [8] DE RIJK M.C., LAUNER L.J., BERGER K., BRETELER M.M.B., DARTIGUES J., BALDERESCHI M., FRATIGLIONI L., LOBO A., MARTINEZ-LAGE J., TRENK-WALDER C., HOFMAN A. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. *Neurology*. 2000, 54(11 SUPPL. 5), pp. S21–S23, doi: [10.1212/WNL.54.11.21ASourc](https://doi.org/10.1212/WNL.54.11.21ASourc).
- [9] ESKIDERE O., ERTAS F., HANILCI C. A comparison of regression methods for remote tracking of Parkinson's disease progression. *Expert Systems with Applications*. 2012, 39(5), pp. 5523–5528, doi: [10.1016/j.eswa.2011.11.067](https://doi.org/10.1016/j.eswa.2011.11.067).
- [10] FOROUZANFAR M., DAJANI H., GROZA V., BOLIC M., RAJAN S. Comparison of Feed-Forward Neural Network training algorithms for oscillometric blood pressure estimation. In: *4th International Workshop on Soft Computing Applications, SOFA 2010*, Arad, Romania, 2010, pp. 119–123, doi: [10.1109/SOFA.2010.5565614](https://doi.org/10.1109/SOFA.2010.5565614).
- [11] GEMAN O. Data processing for Parkinson's disease: Tremor, speech and gait signal analysis. In: *E-Health and Bioengineering Conference, EHB 2011*, Iasi, Romania: IEEE, 2011, pp. 1–4,
- [12] GOETZ C., STEBBINS G., WOLFF D., DELEEUW W., BRONTE-STEWART H., ELBLE R., HALLETT M., NUTT J., RAMIG L., SANGER T., WU A., KRAUS P., BLASUCCI L., SHAMIM E., SETHI K., SPIELMAN J., KUBOTA K., GROVE A., DISHMAN E., TAYLOR C. Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device. *Movement Disorders*. 2009, 24(4), pp. 551–556, doi: [10.1002/mds.22379](https://doi.org/10.1002/mds.22379).
- [13] GRAVES A., JAITLEY N., MOHAMED A.R. Hybrid speech recognition with Deep Bidirectional LSTM. In: *IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2013*, Olomouc, Czech Republic: IEEE, 2013, pp. 273–278, doi: [10.1109/ASRU.2013.6707742](https://doi.org/10.1109/ASRU.2013.6707742).
- [14] HAGAN M.T., DEMUTH H.B., BEALE M. *Neural Network Design*. 2nd ed. Stillwater, US: Martin Hagan, 2014.
- [15] JAGTAP P., PILLAI G. Comparison of extreme-ANFIS and ANFIS networks for regression problems. In: *4th IEEE International Advance Computing Conference, IACC 2014*, Gurgaon, India: IEEE Computer Society, 2014, pp. 1190–1194, doi: [10.1109/IAdCC.2014.6779496](https://doi.org/10.1109/IAdCC.2014.6779496).
- [16] JIN Y. *Advanced Fuzzy Systems Design and Applications*. 1st ed. Heidelberg, Germany: Physica-Verlag, 2003.
- [17] LE D.H., PHAM C.K., NGUYEN T., BUI T. Parameter extraction and optimization using Levenberg-Marquardt algorithm. In: *4th International Conference on Communications and Electronics, ICCE 2012*, Hue Royal City, Viet Nam, 2012, pp. 434–437, doi: [10.1109/CCE.2012.6315945](https://doi.org/10.1109/CCE.2012.6315945).

- [18] LIAO H., MCDERMOTT E., SENIOR A. Large scale deep neural network acoustic modeling with semi-supervised training data for YouTube video transcription. In: *IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2013*, Olomouc, Czech Republic: IEEE, 2013, pp. 368–373, doi: [10.1109/ASRU.2013.6707758](https://doi.org/10.1109/ASRU.2013.6707758).
- [19] LIMA C., COELHO A., VON ZUBEN F. Fuzzy systems design via ensembles of ANFIS. In: *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'02*, Honolulu, US: IEEE, 2002, pp. 506–511, doi: [10.1109/FUZZ.2002.1005042](https://doi.org/10.1109/FUZZ.2002.1005042).
- [20] LITTLE M.A., MCSHARRY P.E., HUNTER E.J., SPIELMAN J., RAMIG L.O. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*. 2009, 56(4), pp. 1015–1022, doi: [10.1109/TBME.2008.2005954](https://doi.org/10.1109/TBME.2008.2005954).
- [21] OZKAN G., INAL M. Comparison of neural network application for fuzzy and ANFIS approaches for multi-criteria decision making problems. *Applied Soft Computing Journal*. 2014, 24, pp. 232–238, doi: [10.1016/j.asoc.2014.06.032](https://doi.org/10.1016/j.asoc.2014.06.032).
- [22] PEDRYCZ W., GOMIDE F. *An Introduction to Fuzzy Sets: Analysis and Design*. Cambridge MA: MIT Press, 1998.
- [23] REYNALDI A., LUKAS S., MARGARETHA H. Backpropagation and Levenberg-Marquardt Algorithm for Training Finite Element Neural Network. In: *6th UK-Sim/AMSS European Symposium on Computer Modeling and Simulation (EMS)*, Valetta, Malta: IEEE, 2012, pp. 89–94, doi: [10.1109/EMS.2012.56](https://doi.org/10.1109/EMS.2012.56).
- [24] STROM F., KOKER R. A parallel neural network approach to prediction of Parkinson's Disease. *Expert Systems with Applications*. 2011, 38(10), pp. 12470–12474, doi: [10.1016/j.eswa.2011.04.028](https://doi.org/10.1016/j.eswa.2011.04.028).
- [25] TSANAS A., LITTLE M., MCSHARRY P., RAMIG L. Accurate telemonitoring of parkinsons disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering*. 2010, 57(4), pp. 884–893, doi: [10.1109/TBME.2009.2036000](https://doi.org/10.1109/TBME.2009.2036000).
- [26] TSANAS A., LITTLE M., MCSHARRY P., RAMIG L. Enhanced classical dysphonia measures and sparse regression for telemonitoring of Parkinson's disease progression. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, Dallas, US: IEEE, 2010, pp. 594–597, doi: [10.1109/ICASSP.2010.5495554](https://doi.org/10.1109/ICASSP.2010.5495554).