



# On the Use of Neuroevolutionary Methods as Support Tools for Diagnosing Appendicitis and Tuberculosis

Alvaro David Orjuela-Cañón<sup>1</sup>(✉), Hugo Fernando Posada-Quintero<sup>2</sup>, Cesar Hernando Valencia<sup>3</sup>, and Leonardo Mendoza<sup>4</sup>

<sup>1</sup> Universidad Antonio Nariño, Bogotá D.C., Colombia  
alvorjuela@uan.edu.co

<sup>2</sup> University of Connecticut, Storrs, CT, USA

<sup>3</sup> Universidad Santo Tomás, Bucaramanga, Colombia

<sup>4</sup> Universidade Estadual do Rio de Janeiro, Rio de Janeiro, Brazil

**Abstract.** Artificial neural networks are being used in diagnosis support systems to detect different kind of diseases. As the design of multilayer perceptron is an open question, the present work shows a comparison between a traditional empirical way and neuroevolution method to find the best architecture to solve the disease detection problem. Tuberculosis and appendicitis databases were employed to test both proposals. Results show that neuroevolution offers a good alternative for the tuberculosis problem but there is lacks of performance in the appendicitis one.

**Keywords:** Neuroevolution · Artificial neural networks  
Diagnosis support systems · Tuberculosis · Appendicitis

## 1 Introduction

Nowadays with the increasing volume in data, different fields have been advantage beneficiaries with the advances of the computational intelligence. More techniques and methods are employed in big data, knowledge extraction and support of decision making in diverse applications. One example of this can be encountered in medical purposes, where a better management of data, faster performance and improvement of the level of accuracy detection in diseases have made evident [1].

Different proposals from the computational intelligence have been employed to support problems in medical and biomedical applications. Learning systems as artificial neural networks (ANN) are commonly used in different applications such as image analysis and pattern recognition, biochemical analysis, drug design and diagnostic systems mainly [2–4]. In this last field, ANN have been used as support assistance for diagnosis of appendicitis in patients presenting with acute right iliac fossa [5], and comparisons with traditional medical models as the decision trees also have been implemented [6, 7]. Exploring different proposals of these learning systems, a database from 801 patients was applied to distinct proposals of ANN as radial basis function, multilayer network and probabilistic networks with results that reached rates of 99% in

accuracy [8]. Other study to regard is related with analysis of data from a rural location, based on information collected in a period of 12 months from 156 patients, which also employed an ANN [9]. For the tuberculosis case, ANN have been widely used in diagnosis support systems. Proposals with different input information have been exposed, having results with sensitivity values upper than 80%, reaching rates of 100% in some cases. For specificity, results have been less satisfactory, with registered values that have dropped to 40%, in the worst case [10–13]. All those results show differences according to available information used in each study. Some studies just used a couple of variables, and a very few cases used all medical data of patients. It also depends on the quality of the information system used for such studies.

Neuroevolution of augmenting topologies (NEAT) is a technique to evolve ANN based on genetic algorithms. It was proposed in the beginning of the present millennium and is based on modifications of the weights and structure of the network synapses. Meanwhile the balance between the fitness of the evolved solutions and their diversity is maintained, crossover among topologies is developed, applying speciation [14–16].

Present work establishes a comparison between two different techniques to obtain the best architecture of an ANN to solve the disease detection problem in diagnosis support systems for tuberculosis and appendicitis. First technique is based on a traditional empirical mode to find the number of nodes in the hidden layer of a multilayer perceptron (MLP), which consists in to modify the number of nodes and test the network performance. Second method is based on NEAT technique to find the nodes and connections to solve the mentioned problems.

## 2 Methodology

Two databases were used for our comparison due to its similarities in number of examples, proportion of positive cases and number of variables to be considered. Both databases are detailed in this section. Then, aspects about the implementation of the ANN for disease detection will be explained.

### 2.1 Databases

First database was obtained from the TB Program at Hospital Santa Clara (HSC) in Bogotá D.C. - Colombia. Information from people under suspicion of pulmonary tuberculosis in the period from January 2008 to March 2011 was considered. The Ethics and Research Committee of the HSC approved this study. An informed consent was not needed because all data were obtained in a retrospective and anonymous mode. Only data from subjects with confirmed diagnostic were considered (using culture and individuals that finished the anti-TB therapy). At the end, information of 105 subjects was used: 83 subjects (79%) with TB confirmed and 22 subjects (21%) that were determined without the disease using diagnosis of exclusion. Confirmation of the TB cases was achieved using a culture test. For TB negative cases, tests did not have a positive culture test, other disease was found meanwhile the treatment, and as mentioned, a diagnosis of exclusion was used.

Features were extracted from different information. A first examination of signs and symptoms was performed by medical personnel, and a clinical suspicion diagnosis was determined. This variable was represented in an input variable named “Clinical information”, which takes a “1” value when just the medical report was considered, and “0” when other test result or additional information lead the subject to start the treatment. Other included variables were extracted from sex, age, homeless, diabetes status, and HIV (human immunodeficiency virus)/AIDS (acquired immunodeficiency syndrome) status. This last was determined using the study of clinical suggestion and confirming the status with exams, but without complementary information as CD4 cell or viral load. All variables were coded with zeros and ones according to negative or positive presence, respectively (Table 1). Age variable was maintained as numeric, with its original information, and a normalization given by the maximum value was achieved. This procedure was developed to avoid saturation of values in the synaptic weights of network and to avoid a wrong representation of the information in training.

**Table 1.** Variables used in the Tuberculosis diagnostic problem.

Variable		Quantity
Sex	Male	72 (69%)
	Female	33 (31%)
Age	Mean + std	40.8 + 17.7
Clinical information	Yes	30 (29%)
	No	75 (71%)
HIV/AIDS	Yes	28 (27%)
	No	77 (73%)
Homeless	Yes	33 (27%)
	No	72 (72%)
Diabetes	Yes	2 (2%)
	No	103 (98%)

Second database represents seven medical measures with information from laboratory tests to confirm the diagnosis of appendicitis [17]. A total of 106 patients admitted to the emergency room during a three-month in 1980 and a six-month period of 1981–1982 were included with 85 with confirmed diagnostic by biopsy analysis with a histologic examination of the removed appendix. The age range was from two to 81 years with a mean age of 25. Fifty-five patients were males and 51 were females [18].

Variables were extracted from information from temperature and an admission blood sample. Then, total white count, manual differential count, cytochemical differential, and C-reactive protein quantity were taken into account. All values were real and then when normalized to be in the interval [0–1], according with the previous works reported in [17, 18]. Information about these variables can be seen in Table 2, where ‘Yes’ means subjects with confirmed disease.

**Table 2.** Variables used in the appendicitis diagnostic problem.

Variable		Mean + std
Temperature (°F)	Yes	100.2 + 1.4
	No	99.2 + 1.1
Total White Blood Count (WBC)	Yes	13.8 + 3.7
	No	9.2 + 4.8
Neutrophils count (%)	Yes	79.9 + 8.0
	No	68.3 + 11.4
Neutrophils count ( $\times 10^9/L$ )	Yes	11.0 + 3.3
	No	6.6 + 4.5
Manual differential count (%)	Yes	16.7 + 14.3
	No	8.3 + 10.6
Manual differential count ( $\times 10^9/L$ )	Yes	2.2 + 2.1
	No	0.7 + 0.8
C-reactive proteins	Yes	4.0 + 4.3
	No	2.2 + 4.4

## 2.2 Models Determination

For estimating the statistical error and generalization of the models, using the explained dataset, cross-validation technique was employed [19]. In this case, the dataset was divided into three sets ensuring that data from people without and with the disease are equitable distributed. This is performed to assess the generalization of the trained model, preserving a portion of data that was not used in training. Table 3 shows these sets and the number of its members.

Commonly, one hidden layer and one output layer are enough to solve classification problems, regardless of the type of input variable [20]. The ANN used in this work had an input layer composed of seven units, each one for each variable, and one output layer composed just of one neuron. Values of +1 and -1 were used to represent if input data corresponds to a patient with TB or not, respectively. Neurons in the hidden layer were established in an experimental way, testing from two to ten neurons. All neurons had a hyperbolic tangent function as activation function.

Between different algorithms to train the ANN, resilient backpropagation was used because its speed and low computational cost [21]. In training, also a cross-validation strategy was considered. In each case, training was performed with two sets (see Table 3) and results for validation were computed with the left out set, maximizing the classification rate between TB and no disease. To avoid overfitting, an early stopping procedure was implemented. Performances of the obtained models were evaluated using sensitivity and specificity. The different trainings were performed employing MATLAB 2017a (The MathWorks, Inc, Natick, MA) through its Neural Networks Toolbox [22].

**Table 3.** Dataset division for training and validation.

Fold	Tuberculosis			Appendicitis		
	Positive	Negative	Total	Positive	Negative	Total
1	28	7	35	29	7	36
2	28	7	35	28	7	35
3	27	8	35	28	7	35
Total	83	22	105	85	21	106

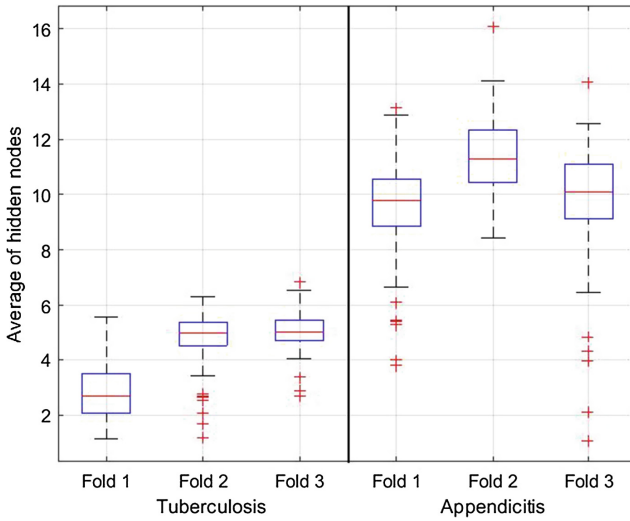
In the development of the NEAT approach, a proposal begins with a network in a similar way that a MLP feed-forward network of only input neurons and output neurons. As evolution progresses through discrete steps, the complexity of the network grows, inserting more nodes into a connection between input-output path and creating new connections within the actual nodes. For this, parameters of the network are represented into a phenotype of genetic algorithms, encoding the schemes that means every connection and neuron in an explicit representation. In this way, the NEAT method attempt simultaneously the learned weights and an appropriate topology for the MLP.

For this case, a cross-validation was used and 100 initializations were implemented to compute statistical measures of the results and to see the coherence between them. Fitness computation was based in the error given by the distance of the output from the correct answer summed for all patterns included in the training set. The resulting number was squared to give more proportionally more fitness the closer a network was to a solution [14]. First generation was formed by networks without hidden nodes with one output and a number of inputs according with the number of variables in each database. Also, a node with bias information was settled as one. There were connections corresponding with these initial nodes, representing the genes in each genome, and each connection with a random weight.

One hundred runs were developed by each fold considered in the database for training. Each run was composed by 200 as maximum number of generation for a generational loop, and the population size was of 150. The population with best results was save as more representative of each run. Finally, the results were collected to study the generalization of the network in front of new patterns taken from the fold left out. The NEAT tool for Matlab © was employed to developed the experiments [16].

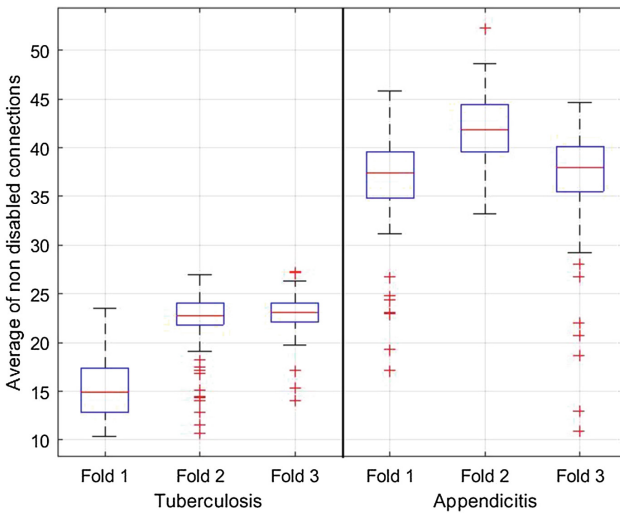
### 3 Results

First, results with comparison for both databases when the NEAT method was applied are presented. Then, results for each disease are shown in terms of sensitivity and specificity. Figures 1 and 2 show the results for the three used folds, visualizing the average of nodes in the hidden layer found by the NEAT method. It is possible to see that the tuberculosis detection, architectures between two and six nodes were obtained to do the classification. Meanwhile, the Appendicitis problem shows that between ten and twelve nodes, in mean, are necessary to perform the classification.



**Fig. 1.** Results for average of number of hidden nodes found for the NEAT method in both databases.

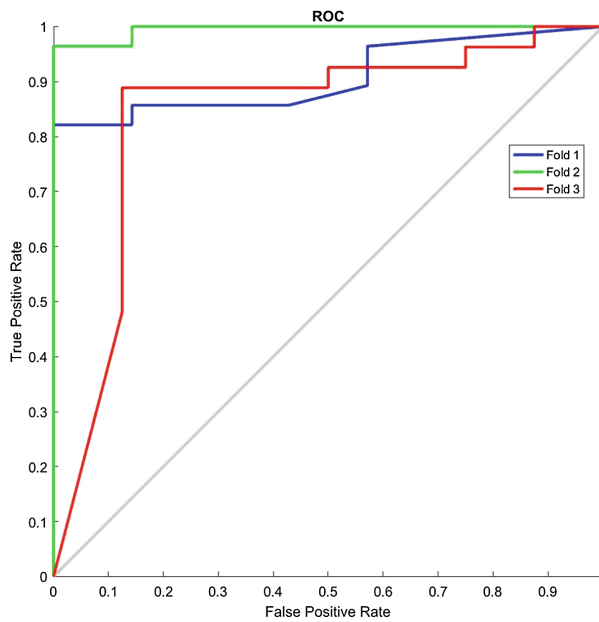
Other comparison can be visualized in the number of connections in the hidden layer (Fig. 2), where the appendicitis problem seems to be more difficult to classify. This can be indicated by the higher numbers for nodes and connections, compared with the tuberculosis case.



**Fig. 2.** Results for average of disabled connections found for the NEAT method in both databases.

Neuroevolution results were compared with the empirical method to find the number of nodes in the hidden layer in a MLP between two and ten, according as mentioned before. The best values of each fold were taken into account to compute the comparative results. Table 4 resumes these corresponding results. For this, the best value in each fold was taken into account and mean and standard deviation were computed for these measures. In empirical cases, the model to detect appendicitis had three nodes in the hidden layer for all folds. The model to detect tuberculosis had two, five and nine nodes in the hidden layer, respectively.

Figures 3 and 4 show the ROC curve for the used folds. It is possible to see that best results were reached with the empirical method (Fig. 4). Therefore, the NEAT method reached comparable results for the detection (Fig. 3). In a similar way, Figs. 5 and 6 show the results for appendicitis detection, allowing to observe similarities in both methods, but with best results for the empirical one (Table 4).



**Fig. 3.** ROC curve for three folds in the tuberculosis problem using NEAT method.

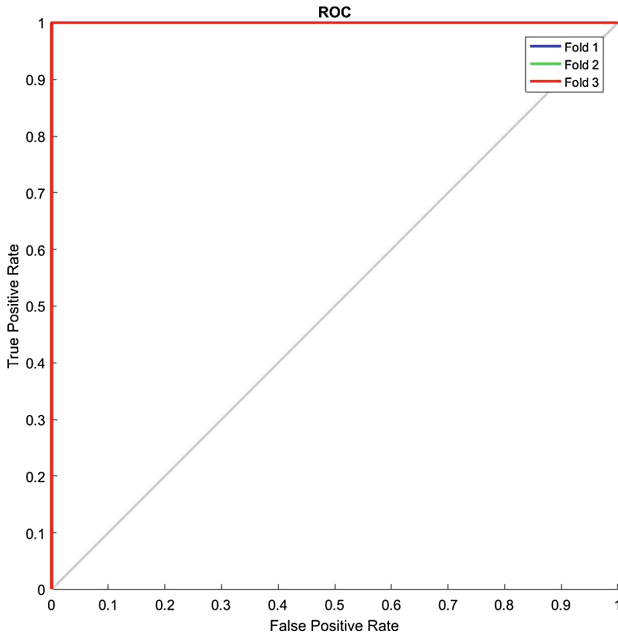


Fig. 4. ROC curve for three folds in the tuberculosis problem using the empirical method.

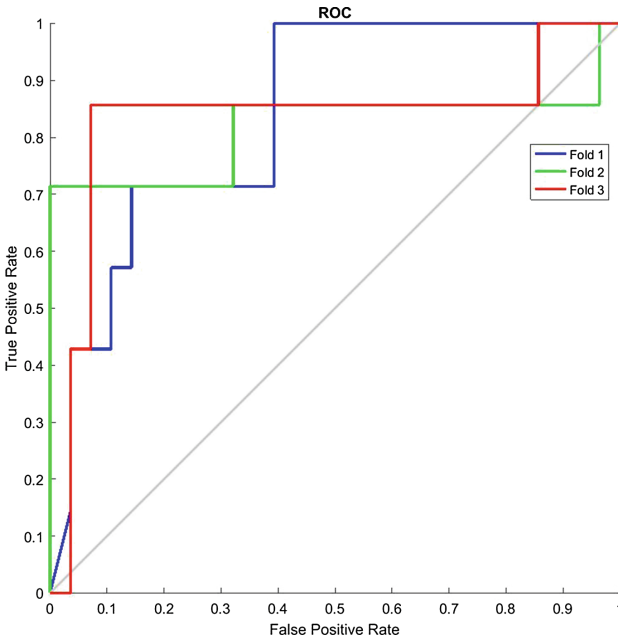
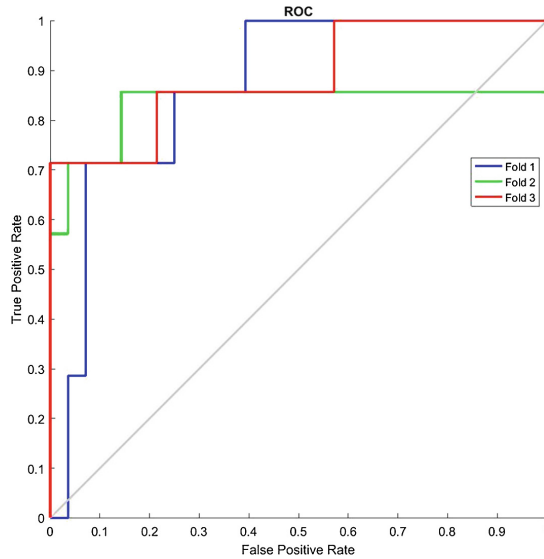


Fig. 5. ROC curve for three folds in the appendicitis problem using the NEAT method.





**Fig. 6.** ROC curve for three folds in the appendicitis problem using the empirical method.

**Table 4.** Comparison results for both employed techniques.

Measure (%)	Tuberculosis		Appendicitis	
	Neuro-evolution	Empirical	Neuro-evolution	Empirical
Sensitivity	88.09 $\pm$ 7.43	96.42 $\pm$ 3.57	91.86 $\pm$ 11.03	96.51 $\pm$ 3.44
Specificity	90.47 $\pm$ 8.24	100 $\pm$ 0	68.45 $\pm$ 5.15	71.43 $\pm$ 0

## 4 Discussion

First observation is given by the complexity of models for both diseases. Tuberculosis diagnosis problem come from an easier problem compared with the appendicitis problem, this can be seen according with the number of hidden nodes and the connections in the network when NEAT method was used (Figs. 1 and 2). There is possible to note that tuberculosis problem can be solve with between three and six nodes, while appendicitis problem needs between ten and twelve nodes in mean to detect the disease.

After the comparison with the empirical technique, it is possible to observe that the results are equivalent for the tuberculosis detection (Table 4). NEAT offers best results for this particular problem, finding a better sensitivity and specificity. At the same time a smaller architecture also is found. This because the results are for architectures between three and six nodes in the NEAT method compared with three, five or nine nodes for the empirical way. This, in terms of number of connections is more efficient because for NEAT method this number can be in the interval 15 to 25 in mean. Instead

the model with six nodes in the hidden layer for the empirical method produces around 50 connections.

In the case of appendicitis detection, the values for sensitivity and specificity were not so good. It is notable the results for the specificity when the empirical way was implemented, but sensitivity was not well succeeded as in reported works, where reached rates of 89% [18] and 97% [9]. However, the results are comparable and the scope of this work was to compare two different methods to obtain ANN architectures. In this case, a NEAT method offered comparable results as regards to measures of sensitivity and specificity, and size of the architecture related with nodes and connections.

Differences in the results can be explained by the type of used variables. For the tuberculosis case, most of variables are binary inputs, manifesting existence or absence of the variable. Meanwhile the used variables in the appendicitis problem were all real, making more difficult the classification. In spite of ANN can be deal with any type of input variable, for the present case, this could manifest a problem. This supposes a future work to do by the light of these results.

## 5 Conclusions

Two methods to obtain networks architectures were compared for the disease detection problem. A neuroevolution of augmenting topologies and traditional empirical way were employed to detect diseases based on different input variables. Results show that for tuberculosis case, the NEAT method offered better results and for appendicitis case the results are comparable in terms of sensitivity and specificity, but for architecture size represent a bigger network.

**Acknowledgment.** Authors thank the Universidad Antonio Nariño under project 2016207, University of Connecticut, Universidad Santo Tomas and Universidade Estadual do Rio de Janeiro for the support and financial assistance in this work. Also, the Hospital Santa Clara and Carlos Awad for making available the database related with tuberculosis.

## References

1. Kalantari, A., Kamsin, A., Shamsirband, S., Gani, A., Alinejad-Rokny, H., Chronopoulos, A.T.: Computational intelligence approaches for classification of medical data: state-of-the-art, future challenges and research directions. *Neurocomputing* **276**, 2–22 (2017)
2. Amato, F., López, A., Peña-Méndez, E.M., Vavnghara, P., Hampl, A., Havel, J.: Artificial neural networks in medical diagnosis (2013)
3. Amato, F., González-Hernández, J.L., Havel, J.: Artificial neural networks combined with experimental design: a soft approach for chemical kinetics. *Talanta* **93**, 72–78 (2012)
4. Seera, M., Lim, C.P.: A hybrid intelligent system for medical data classification. *Expert Syst. Appl.* **41**, 2239–2249 (2014)
5. Prabhudesai, S.G., Gould, S., Rekhraj, S., Tekkis, P.P., Glazer, G., Ziprin, P.: Artificial neural networks: useful aid in diagnosing acute appendicitis. *World J. Surg.* **32**, 305–309 (2008)

6. Huang, R., Lee, S., Lai, C., Hsiao, Y., Ting, H.: Acute effects of obstructive sleep apnea on autonomic nervous system, arterial stiffness and heart rate in newly diagnosed untreated patients. *Sleep Med.* **14**, e27 (2013)
7. Ting, H.-W., Wu, J.-T., Chan, C.-L., Lin, S.-L., Chen, M.-H.: Decision model for acute appendicitis treatment with decision tree technology - a modification of the alvarado scoring system. *J. Chin. Med. Assoc.* **73**, 401–406 (2010)
8. Park, S.Y., Kim, S.M.: Acute appendicitis diagnosis using artificial neural networks. *Technol. Health Care* **23**, S559–S565 (2015)
9. Yoldaş, Ö., Tez, M., Karaca, T.: Artificial neural networks in the diagnosis of acute appendicitis. *Am. J. Emerg. Med.* **30**, 1245–1247 (2012)
10. Santos, A.M., Pereira, B.B., Seixas, J.M., Mello, F.C.Q., Kritski, A.L.: Neural networks: an application for predicting smear negative pulmonary tuberculosis. In: Auget, J.L., Balakrishnan, N., Mesbah, M., Molenberghs, G. (eds.) *Advances in Statistical Methods for the Health Sciences*, pp. 275–287. Springer, Boston (2007)
11. El-Solh, A.A., Hsiao, C.-B., Goodnough, S., Serghani, J., Grant, B.J.B.: Predicting active pulmonary tuberculosis using an artificial neural network. *Chest J.* **116**, 968–973 (1999)
12. Elveren, E., Yumuvak, N.: Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm. *J. Med. Syst.* **35**, 329–332 (2011)
13. Er, O., Temurtas, F., Tanrikulu, A.Ç.: Tuberculosis disease diagnosis using artificial neural networks. *J. Med. Syst.* **34**, 299–302 (2010)
14. Stanley, K.O., Miikkulainen, R.: Evolving neural networks through augmenting topologies. *Evol. Comput.* **10**, 99–127 (2002)
15. Floreano, D., Dürr, P., Mattiussi, C.: Neuroevolution: from architectures to learning. *Evol. Intell.* **1**, 47–62 (2008)
16. Miikkulainen, R.: Neuroevolution. In: *Encyclopedia of Machine Learning*. Springer, New York (2010)
17. Marchand, A., Van Lente, F., Galen, R.S.: The assessment of laboratory tests in the diagnosis of acute appendicitis. *Am. J. Clin. Pathol.* **80**, 369–374 (1983)
18. Weiss, S.M., Kapouleas, I.: An empirical comparison of pattern recognition, neural nets and machine learning classification methods. *Read. Mach. Learn.* 177–183 (1990)
19. Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Presented at the (1995)
20. Haykin, S.: *Neural Networks and Learning Machines*. Prentice Hall, Upper Saddle River (2009)
21. Riedmiller, M., Braun, H.: A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: 1993 IEEE International Conference on Neural Networks, pp. 586–591 (1993)
22. Beale, M.H., Hagan, M.T., Demuth, H.B.: *Neural network toolbox getting started guide R2011b* (2011)