

ANN for prognosis of abdominal pain in childhood: use of fuzzy modelling for convergence estimation

George C. Anastassopoulos, Lazaros S. Iliadis

Abstract. This paper focuses in two parallel objectives. First it aims in presenting a series of Artificial Neural Network models that are capable of performing prognosis of abdominal pain in childhood. Clinical medical data records have been gathered and used towards this direction. Its second target is the presentation and application of an innovative fuzzy algebraic model capable of evaluating Artificial Neural Networks' performance [1]. This model offers a flexible approach that uses fuzzy numbers, fuzzy sets and various fuzzy intensification and dilution techniques to perform assessment of neural models under different perspectives. It also produces partial and overall evaluation indices. The produced ANN models have proven to perform the classification with significant success in the testing phase with first time seen data.

1 INTRODUCTION

The wide range of problems in which Artificial Neural Networks can be used with promising results, is the reason of their growth [2, 3]. Some of the fields that ANNs are used are: medical systems [4-6], robotics [7], industry [8 – 11], image processing [12], applied mathematics [13], financial analysis [14], environmental risk modelling [15] and others.

Prognosis is a medical term denoting an attempt of physician to accurately estimate how a patient's disease will progress, and whether there is chance of recovery, based on an objective set of factors that represent that situation. The inference about prognosis of a patient when presented with complex clinical and prognostic information is a common problem, in clinical medicine. The diagnosis of a disease is the outcome of combination of clinical and laboratorial examinations through medical techniques.

In this paper various ANN architectures using different learning rules, transfer functions and optimization algorithms have been tried. This research effort was motivated from the fact that reliable and seasonable detection of abdomen pain constitute attainments in effective treatment of disease and avoidance of relapses. That is why the development of such an intelligent model that can collaborate with the doctors will be very useful towards successful treatment of potential patients.

2 DIAGNOSTIC FACTORS OF ABDOMINAL PAIN

Several reports have described clinical scoring systems incorporating specific elements of the history, physical examination, and laboratory studies designed to improve diagnostic accuracy of abdominal pain [16]. Nothing is guaranteed, but

Democritus University of Thrace, Hellenic Open University
anasta@med.duth.gr, liliadis@fmenr.duth.gr

decision rules can predict which children are at risk for appendicitis (appendicitis is the most common surgical condition of the abdomen). One such numerically based system is based on a 6-part scoring system: nausea (6 point), history of local RLQ pain (2 point), migration of pain (1 point), difficulty walking (1 point), rebound tenderness / pain with percussion (2 point), and absolute neutrophil count of $>6.75 \times 10^3/\mu\text{L}$ (6 point). A score <5 had a sensitivity of 96.3% with a negative predictive value of 95.6% for AA.

To date, all efforts to find clinical features or laboratory tests, either alone or in combination, that are able to diagnose appendicitis with 100% sensitivity or specificity have proven futile. Also, there is only one research work [4] in bibliography based on ANN that deals with the abdominal pain prognosis in childhood.

The incidence of Acute Appendicitis (AA) is 4 cases per 1000 children. However appendicitis despite pediatric surgeons' best efforts remains the most commonly misdiagnosed surgical condition. Although diagnosis and treatment have improved, appendicitis continues to cause significant morbidity and still remains, although rarely, a cause of death. Appendicitis has a male-to-female ratio of 3:2 with a peak incidence between ages 12 and 18 years. The mean age in the pediatric population is 6-10 years. The lifetime risk is 8.6% for boys and 6.7% for girls.

The 15 factors that are used in the routine clinical practice for the assessment of AA in childhood are: Sex, Age, Religion, Demographic data, Duration of Pain, Vomitus, Diarrhea, Anorexia, Tenderness, Rebound, Leucocytosis, Neutrophilia, Urinalysis, Temperature, Constipation. The sex (males), the age (peak of appearance of A.A in children aged 9 to 13 years), and the religion (hygiene condition, feeding attitudes, genetic predisposition) were in relation with a higher frequency for AA. Anorexia, vomitus, diarrhea or constipation and a slight elevation of the temperature ($37^0\text{C} - 38^0\text{C}$) were common manifestation of AA. Additionally, abdominal tenderness principally in the RLQ of the abdomen and the existence of the rebound sign, are strongly related with AA. Leucocytosis ($>10.800\text{ K}/\mu\text{L}$) with neutrophilia (neutrophil count $>75\%$) is considered to be a significant clue for AA. Urinalysis is useful for detecting urinary tract disease, normal findings on urinalysis are of limited diagnostic value for appendicitis.

The role of race, ethnicity, health insurance, education, access to healthcare, and economic status on the development and treatment of appendicitis are widely debated. Cogent arguments have been made on both sides for and against the significance of each socioeconomic or racial condition. A genetic predisposition appears operative in some cases, particularly in children in whom appendicitis develops before age 6 years. Although the disorder is uncommon in infants and elderly, these groups have a disproportionate number of complications because of delays in diagnosis and the presence of comorbid conditions.

As diagnosis, there are four stages of appendicitis, including acute focal appendicitis, acute suppurative appendicitis, gangrenous appendicitis and perforated appendicitis. These distinctions are vague, and only the clinically relevant distinction of perforated (gangrenous appendicitis includes into this entity as dead intestine functionally acts as a perforation) versus non-perforated appendicitis (acute focal and suppurative appendicitis) should be made.

The present study is based on data set that is obtained from the Pediatric Surgery Clinical Information System of the University Hospital of Alexandroupolis, Greece. It consisted of 516 children's medical records. Some of these children had different stages of appendicitis and, therefore, underwent operative treatment. This data set was divided into a set of 422 records and another set of 94 records. The former was used for training of the ANN, while the latter for testing. A small number of data records were used as a validation set during training to avoid overfitting. Table 1 represents the stages of appendicitis as well as the corresponding cases for each one. The 3rd column of Table 1 depicts the coding of possible diagnosis, as they used for ANN training and testing stages.

Table 1. Possible diagnosis and corresponding cases.

	Diagnosis	Coding	Cases
Normal	Discharge	-2	236
	Observation	-1	186
Operative treatment	No findings	0	15
	Focal appendicitis	1	34
	Phlegmonous or Suppurative appendicitis	2	29
	Gangrenous appendicitis	3	8
	Peritonitis	4	8

3 NEURAL NETWORK DESIGN

Data were divided into two groups, the training cases (TRAC) and the testing cases (TESC). The TRAC consisted of 417 concrete medical data records and the TESC consisted of 101. Each input record was organised in a format of fifteen fields, namely sex, age, religion, area of residence, pain time period, vomit symptoms, diarrhoea, anorexia, located sensitivity, rebound, wbc, poly, general analysis of urine, body temperature, constipation. The output record contained a single field which corresponded to the potential outcome of each case.

The determination if the TRAC and TESC data sets was performed in a rather random manner. The training and testing sample size which would be sufficient for a good generalization was determined by using the Widrow's rule of thumb for the LMS algorithm which is a distribution free, worst case formula [2] and it is shown in the following equation 1. W is the total number of free parameters in the network (synaptic weights and biases) and ϵ denotes the fraction of the classification errors permitted during testing. The O notation shows the order of quantity enclosed within [2].

$$N = O\left(\frac{W}{\epsilon}\right) \quad (1)$$

In the case examined here with 417 training examples used, the classification error that could be tolerated would be about 4%.

3.1 Description of the experiments performed

During experimentations, numerous ANN architectures, learning algorithms and transfer functions were combined in an effort to obtain the optimal network. For the Tangent Hyperbolic

(TanH) transfer function the input data were normalized (divided properly) in order to be included in the acceptable range of $[-3, 3]$ to avoid problems such as saturation, where an element's summation value (the sum of the inputs times the weights) exceeds the acceptable network range [17]. Standard back-propagation optimization algorithms using TanH, or Sigmoid or Digital Neural Network Architecture (DNNA) transfer functions, combined with the Extended Delta Bar Delta (ExtDBD) or with the Quick Prop learning rules [18, 19] were employed. The ExtDBD is a heuristic technique reinforcing good general trends and damping oscillations [20].

Modular and radial basis function (RBF) ANN applying the ExtDBD learning rule and the TanH transfer function were also used in an effort to determine the optimal networks. RBFs have an internal representation of hidden neurons which are radially symmetric, and the hidden layer consists of pattern units fully connected to a linear output layer [21, 22].

3.2 ANN evaluation metrics applied

Traditional ANN evaluation measures like the Root Mean Square Error (RMS error), R^2 and the confusion matrix were used to validate the ensuing neural network models. It is well known that the RMS error adds up the squares of the errors for each neuron in the output layer, divides by the number of neurons in the output layer to obtain an average, and then takes the square root of that average. The confusion matrix is a graphical way of measuring the network's performance during the "training" and "testing" phases. It also facilitates the correlation of the network output to the actual observed values that belong to the testing set in a visual display [17], and therefore provides a visual indication of the network's performance. A network with the optimal configuration should have the "bins" (the cells in each matrix) on the diagonal from the lower left to the upper right of the output. An important aspect of the matrix is that the value of the vertical axis in the generated histogram is the Common Mean Correlation (CMC) coefficient of the desired (d), and the actual (predicted) output (y) across the Epoch.

Finally, the FUSETRESYS (Fuzzy Set Transformer Evaluation System) that constitutes an innovative ANN evaluation system has been applied offering a more flexible approach [1].

3.3 Technical description of the FUSETRESYS ANN evaluation model

Fuzzy logic enables the performance of calculations with mathematically defined words called "Linguistics" [1, 23-25]. FUSETRESYS faces each training/testing example as a Fuzzy Set. It applies triangular or trapezoidal membership functions in order to determine the partial degree of convergence (PADECOV) of the ANN for each training/testing example separately. The following equations 2 and 3 represent a triangular and a trapezoidal membership functions respectively [1].

$$\mu_s(x;a,b,c)=\max\left\{\min\left\{\frac{x-a}{b-a},\frac{c-x}{c-b}\right\},0\right\}; a<b<c \quad (2)$$

$$\mu_s(x;a,b,c,d)=\max\left\{\min\left\{\frac{x-a}{b-a},1,\frac{d-x}{d-c}\right\},0\right\}; a<b<c<d \quad (3)$$

The model can produce various overall degrees of convergence (OVDECOV) for all of the training examples by applying either fuzzy T-Norm or fuzzy S-Norm conjunction operations, depending on the optimistic or pessimistic point of view of the developer. T-

Norms tend to produce lower aggregation indices so in the case of ANN evaluation they can be considered as a pessimistic approach, whereas the opposite happens with S-Norms [26]. In fact, each distinct Norm evaluates the performance of an ANN under a different perspective. For example the drastic product assigns the ANN a high OVDECOV only if it does not have extreme deviations between the desired and the produced classifications during the training/testing process [1] whereas the Einstein T-Norm acts in a more average mode. The following equations 4 and 5 present the drastic product and the Einstein product T-Norms. More details on fuzzy conjunction operators can be found in [26-28].

$$\mu_{\left(\tilde{A} \cap \tilde{B}\right)} = \text{Min} \left\{ \mu_{\tilde{A}}(X), \mu_{\tilde{B}}(X) \right\} \text{ if } \text{Max} \left\{ \mu_{\tilde{A}}(X), \mu_{\tilde{B}}(X) \right\} = 1 \text{ else}$$

$$\mu_{\left(\tilde{A} \cap \tilde{B}\right)} = 0 \quad (4) \quad \mu_{\left(\tilde{A} \cap \tilde{B}\right)} = \frac{\mu_{\tilde{A}}(X) \mu_{\tilde{B}}(X)}{2 - [\mu_{\tilde{A}}(X) + \mu_{\tilde{B}}(X) - \mu_{\tilde{A}}(X) \mu_{\tilde{B}}(X)]} \quad (5)$$

The fact that the FUSESTRESYS evaluates each training/testing example separately, offers a more clear view of the ANN's performance. In this way the developers know if the network operates extremely bad or well in specific cases.

Also when there are several neurons in the output layer, the traditional approaches produce separate evaluation results for each one whereas the FUSESTRESYS can produce an additive performance index (ADPERI) of the ANN. This could be done under different perspectives and under different degrees of optimism [1].

Finally the application of *fuzzy set hedges* offers the “dilution” and the “intensification” options. In this way by using the dilution approach the developer softens the membership function over the fuzzy set and weakens the membership constraints so that a point of the Universe of discourse is “truer” than it would be before [1, 27]. On the contrary the intensification hardens the MF over the FS and strengthens the membership constraints so that a point on the domain is “less true” than it used to be [1, 27]. The following equations 6 and 7 correspond to the intensification and dilution functions respectively.

$$\mu_{\text{intensify}(A)}(X_i) = \mu_A^n(X_i) \quad (6) \quad \mu_{\text{dilute}(A)}(X_i) = \mu_A^{\frac{1}{n}}(X_i) \quad (7)$$

In this way the ANN can be evaluated strictly by using a “very well fit” evaluation option, or in a more relaxed way by using the “somewhat fit” option. Of course it is in the developer's hand to decide the potential type of the ANN's evaluation and the degree of dilution or intensification. For a more detailed description of FUSESTRESYS please see [1].

4 RESULTS AND DISCUSSION

4.1 ANN analysis

Several experiments were performed. The following table 2 presents the structure of the four most effective Back Propagation (BP) multilayer (ML) neural networks. In all cases of ANN models, the classical approach for overcoming the overfitting problem has been followed. More specifically, a set of validation data have been provided to the algorithm in addition to the training data. The algorithm has monitored the error with respect to this validation set, while using the training set to drive the gradient descent search. The number of weight tuning iterations performed by the system, were determined in each case based on the criterion of lowest error over the validation set. Two copies of the best

performing weights are kept: one copy for training and another one of the best performing weights thus far.

Table 2. Structure of the four most successful ML ANN

Optimization algorithm	Input Layer	Hidden sub-layer neurons	Second Hidden sub-layer	Learning Rule/Transfer Function
ANN ML#1. Reinforcement ANN using BackPropagation	15	7	7	Genetic Algorithm /TanH
ANN ML#2. Multilayer Backpropagation	15	9	0	Norm-Cum_Delta/ TanH
ANN ML#3. Multilayer Backpropagation	15	9	9	Norm-Cum_Delta/ TanH
ANN ML#4. Multilayer Backpropagation	15	7	0	ExtDBD/ TanH

Table 3 shows the architecture and structure of the four most successful *radial basis function* (RBF) ANN.

Table 3. Structure of the two most successful RBF ANN

Optimization algorithm	Input Layer	Proto	Number of neurons Hidden layer	Output Layer	Learning Rule/Transfer function
ANN 5. ANN6 Radial Basis Function	15	30	2	1	Norm Cum_Delta / Sigmoid or TanH

The following Table 4 presents the training and testing results for the most successful ML and RBF ANN using R^2 . Also for the three most successful networks, namely 2,3,4 the FUSESTRESYS model was applied to determine the average degree of convergence. According to the results, the most suitable network was ML#3. The structure and the architecture of this successful network have been described in the above Table 2.

Table 4. Evaluation of the most successful ML and RBF ANN

ANN code	R^2 in Training	R^2 in Testing	Average Degree of success in Testing using FUSESTRESYS for the three best ANN
1	0.8258	0.8247	
2	0.9615	0.9471	0.9699
3	0.9721	0.9489	0.9716
4	0.9352	0.9588	0.9799
5	0.9114	0.9000	
6	0.9346	0.9400	

The following figure 1 is a graphical representation of the all the PADECOV of the ML#2, for each testing example. The absolute degree of convergence has the value of one. A serious effort was

made towards the development of modular ANN (MODANN) for the classification problem solution. The term MODANN refers to the “adaptive” mixtures of *local experts* (LOCEXP) as proposed by [29].

They consist of a group of BP ANN referred to as local experts competing to learn different aspects of a problem. A “gating ANN” controls the competition and learns to assign different parts of the data space to different networks.

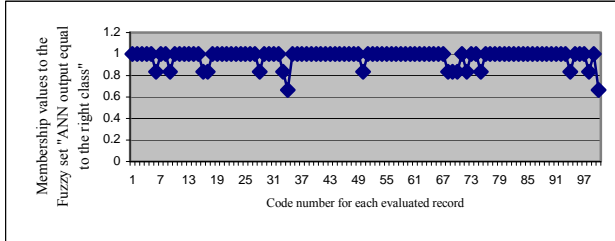


Figure 1. Representation of all the PADECOV of the network ML# 2

The LOCEXP have the same architecture but they can apply distinct learning rules or transfer functions. Also the number of the output processing elements of the gating network is equal to the number of LOCEXP used. The number of the neurons in the hidden layer of the gating network should be larger than the number of the output processing elements [17].

Table 5. Refereed Back Propagation using Gating Networks with two Competing local experts

Refereed #1 REF ANN using Gating ANN with 2 Local Experts				
Learning rule	Transfer	Error	Output	Noise
<i>Local Expert's #1 functions</i>				
ExtDBD	TanH	standard	Direct	Uniform
<i>Local Expert's #2 functions</i>				
Norm-Cum Delta	TanH	standard	Direct	Uniform
<i>Local Expert's Architecture</i>				
Input neurons 15		Hidden neurons 6		Output neurons 1
<i>Gating ANN functions</i>				
Learning rule	Transfer	Error	Output	Noise
ExtDBD	Linear	Standard	SoftMax	Uniform

The above table 5 presents the structure and the architecture of the optimal MODANN that was developed for the medical classification problem examined here. The performance of the developed modular network is very satisfying, having an R^2 value of 0.9434 and a FUSETRESYS produced average PADECOV equal to 0.9733 (using the Triangular membership function) in the testing process using the first time seen testing data set.

The following figure 2 depicts the *gating probabilities* for the optimal MODANN..

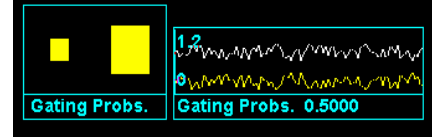


Figure 2. Gating Probabilities of the MODANN with code #1Ref.

Table 6. Small sample of the PADECOV indices

PADECOV by FUSETRESYS		
ML#2	ML#3	#1REF
0.83333	0.83333	1
0.83333	0.83333	0.833
1	1	1
0.83333	0.83333	1
1	1	1
0.83333	0.83333	1
0.833333	0.833333	1
OVDECOV Einstein		
0.98299	0.97784	0.971

The above Table 6 presents a small sample of the 101 distinct PADECOV values produced by the FUSTRESYS.

Also the Einstein T-Norm was applied for the determination of the overall degree of convergence of the ANN. The ML#2 ANN had a very high OVEDECOV index with a value of 0.98299 whereas the other ML#3 ANN and the MODANN #REF1 had OVEDECOV indices as high as 0.97. The Drastic Product T-Norm was not applied in this research effort because it was proven unnecessary from the data in table 5 where there were no serious indications of extreme bad ANN performance in any of the testing examples.

Table 7. OVDECOV values when intensification and dilution of first order was applied using Einstein product and Triangular membership function

	OVDECOV of ML #2	OVDECOV of #1REF	OVDECOV of ML #3
Dilution “Partly fit”	0.99972	0.99934	0.99957
Intensification “Very well fit”	0.75932	0.64887	0.71033

5 CONCLUSIONS

The above research has obtained six ANNs with good level of convergence and it has proven that there exist at least four ANNs that have high performance indices, in the case of abdominal pain classification. Namely the best ANNs are two ML BP ANN, a RBF ANN and a MODANN using a referee gating network and two local experts. All of them have been described in the previous sections.

A very interesting part of the whole research effort is the application of an innovative ANN evaluation model called FUSETRESYS that uses fuzzy logic and fuzzy algebra proposed in [11].

The new evaluation scheme has performed individual convergence indices namely PADECOV, for the output of each single data record used in the testing phase. The worst PADECOV value equals to 0.6666 which actually is the degree of membership of each data record to the FS “Actual output value equal to the desired value”. This worst case appears three times exactly in the

same cases of data records, for the ML#2, ML#3, #1REF ANN and it shows that the classification capacity of the developed networks is not bad even in the worst cases. This conclusion becomes stronger by considering the fact that the second worst PADECOV index has a value of 0.833.

If an overall ANN validation is performed the traditional evaluation instruments agree with the FUSESTRESYS that the most suitable ANN is the ML BP with code# 4 whereas all of the other developed ANN have almost an equally good performance. The Einstein T-Norm produces a higher “good performance index” for the MODANN than the traditional methods.

As it can be seen in table 7, the OVDECOV indices have very high values for ML#2 and for REF#1 and ML#3 networks when a “Partly fit” validation is performed. There is significant differentiation when a very strict evaluation is done under the linguistic “Very well fit”. The OVDECOV indices fall from 0.99 to 0.75 for ML#2, from 0.99 to 0.65 for #REF and from 0.99 to 0.71 for ML#3 respectively. This is a very useful approach and it shows the actual power of FUSESTRESYS due to the fact that it shows the differentiation of the average convergence degree of the three ANN when more strict validation methods are applied. So ANN fed with the same data records in testing and appearing to have more or less the same performance, they are very seriously differentiated when more strict convergence validation methods are performed.

The proposed ANN architecture faces the appendicitis prediction quite satisfactory, based on both the above presented results, and the pediatric surgeon’s opinion that used these ANNs in their everyday routine clinical practice.

The innovative ANN evaluation model that was applied successfully in this research effort will be used extensively in the future, in an integrated effort to check its validity under various perspectives.

ACKNOWLEDGEMENTS

We would like to thank the pediatric surgeons of the Pediatric Surgeon Department of Medical School of Democritus University of Thrace, for their contribution in the concession of the medical records.

REFERENCES

- [1] L. Iliadis, ‘An intelligent Artificial Neural Network evaluation system using Fuzzy Set Hedges: Application in wood industry’, *Proceedings of the 19th IEEE ICTAI The Annual IEEE International Conference on Tools with Artificial Intelligence*. IEEE Volume II, 366-370, Los Alamitos California, 2007.
- [2] S. Haykin, *Neural Networks: A comprehensive foundation*,. McMillan College Publishing Company, New York, 1999.
- [3] P. Picton, *Neural Networks*, (2nd edition) Palgrave, New York, USA, 2000.
- [4] D. Mantzaris, G. Anastassopoulos, A. Adamopoulos, I. Stephanakis, K. Kambouri and S. Gardikis, ‘Abdominal Pain Estimation in Childhood based on Artificial Neural Network Classification’, *Proc. of the 10th International Conference on Engineering Applications of Neural Networks (EANN 2007)*, 129-134, August, 2007.
- [5] G.P.K. Economou, D. Lymberopoulos, E. Karavatselou and C. Chassomeris, ‘A new concept toward computer-aided medical diagnosis - A prototype implementation addressing pulmonary diseases’, *IEEE Transactions on Information Technology in Biomedicine*, **5** (1): 55-66, (2001).
- [6] J. Shieh, S. Fan and W. Shi, ‘The intelligent model of a patient using artificial neural networks for inhalational anaesthesia’, *J. Chin. Inst. Chem. Engrs.*, **33**, No. 6, 609-620, (2002).
- [7] V. Rankovic and I. Nikolic, ‘Control of industrial Robot using neural network compensator’, *Theoretical Applications of Mech.*, **32**, No. 2, 147-163, (2005).
- [8] S. Avramidis and L. Iliadis, ‘Wood-water sorption isotherm prediction with artificial neural networks: a preliminary study’, *Holzforchung*, **59**, 336 – 341, (2005).
- [9] S. Mansfield, L. Iliadis, S. Avramidis, ‘Neural Network Prediction of Bending Strength and Stiffness in Western Hemlock’, *HOLZFORSCHUNG*, **61**, Issue 6, 707-716 Walter De Gruyter & Co Berlin, New York, 2007.
- [10] B. Cannas, A. Fanni, A. Montisci, G. Murgia and P. Sonato, ‘Dynamic Neural Networks for Prediction of Disruptions in Tokamaks’, *Proceedings of the 10th International Conference on Engineering Applications of Neural Networks (EANN)*, Thessaloniki, Greece, 2007.
- [11] L. Iliadis, S. Spartalis and S. Tachos, ‘A Fuzzy intelligent Artificial Neural Network evaluation System: Application in Industry’, *Proceedings of the 10th International Conference Engineering Applications of Neural Networks*, 320-326, Thessaloniki, Greece, 2007.
- [12] R. Kuhn, R. Bordas, B. Wunderlich, B. Michaelis and D. Thevenin, ‘Colour class identification of tracers using artificial neural networks’, *Proceedings of the 10th International Conference on Engineering Applications of Neural Networks (EANN)*, Thessaloniki, Greece, 2007.
- [13] L. Iliadis and S. Spartalis, ‘Artificial Neural Networks equivalent to Fuzzy Algebra T_Norm conjunction operators’, *Proceedings (Book of extended abstracts) of the ICCMSE 2007* Published by the AIP (American Institute of Physics), USA, 2007.
- [14] P. Hajek and V. Olej, ‘Municipal creditworthiness Modelling by clustering methods’, *Proceedings of the 10th International Conference on Engineering Applications of Neural Networks (EANN)*, Thessaloniki, Greece, 2007.
- [15] L. Iliadis, ‘A decision support system applying an integrated Fuzzy model for long - term forest fire risk estimation’, *Environmental Modelling and Software*, **20**, Issue 5, 613-621, (2005).
- [16] M. Blazadonakis, V. Moustakis, and G. Charissis, ‘Deep Assessment of Machine Learning Techniques Using Patient Treatment in Acute Abdominal Pain in Children’, *Artificial Intelligence in Medicine*, **8**, 527- 542, (1996).
- [17] Neuralware, *Getting started. A tutorial for Neuralworks Professional II/PLUS*, Carnegie, PA, USA, 2001.
- [18] R.A. Jacobs, ‘Increased rates of convergence through learning rate adaption’, *Neural Networks* **1**, 295-307, (1988).
- [19] A.A. Minai, R.D. Williams, ‘Acceleration of back-propagation through learning rate and momentum adaption’, *International Joint Conference on Neural Networks*, **1**, 676-679, (1990).
- [20] D.E. Rummelhart, G.E. Hinton and R.J. Williams, Learning internal representations by error propagation. Institute for Cognitive Science Report 8506. San Diego, University of California (1985).
- [21] J. Platt, “A resource allocating network for function interpolation”, *Neural Computation*, **3**, 213-225, (1991).
- [22] J. Moody, C.J. Darken, ‘Fast learning in networks of locally tuned processing units’, *Neural Computation*, **1**, 281-294, (1989).
- [23] R. Callan, *The Essence of Neural Networks*. Prentice Hall, UK, 1999.
- [24] W. Pedrycz, ‘Structural interpolation and approximation with fuzzy relations: A study in knowledge reuse’, *Journal Studies in Fuzziness and Soft Computing*, **215**, 65-77, (2007).
- [25] B. J. Park, W. Pedrycz, S.K. Oh, ‘Fuzzy polynomial neural network: hybrid architectures of fuzzy modelling’, *IEEE Trans. Fuzzy Systems*, **10** (5), 607-621, (2002).
- [26] V. Kecman, *Learning and Soft Computing*, MIT Press. London England, 2001.
- [27] E. Cox, *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration*, Elsevier Science, USA, 2005.
- [28] T. Calvo, G. Mayor and R. Mesira, *Aggregation Operators: New Trends and Applications*, (Studies in Fuzziness and Soft Computing). Physica-Verlag, Heildeberg, 2002.
- [29] R.A. Jacobs, M.I. Jordan, S.J. Nowlan and G.E. Hinton, ‘Adaptive mixtures of local experts’, *Neural computation*, **3**, 79-87, 1991.

