Bayesian-driven Multi-layer Perceptron Applied to Liver Fibrosis Stadialization

Smaranda Belciug[1], Florin Gorunescu[2]
[1]Department of Computer Science, University of Craiova, Romania
[2]Department of Biostatistics and Informatics,
University of Medicine and Pharmacy of Craiova, Romania,
smaranda.belciug@inf.ucv.ro, florin.gorunescu@webmail.umfcv.ro

Abstract

This paper proposes the application to the liver fibrosis stadialization of a novel training technique of feed-forward neural networks based on the Bayesian paradigm. Using the Pearson's *r* correlation coefficient instead of the standard backpropagation algorithm to update the synaptic weights of a multi-layer perceptron, the proposed model is compared with traditional machine learning algorithms (standard MLP, RBF, PNN, SVM) using a real-life liver fibrosis dataset. The statistical comparison results indicated that the Bayesian-trained MLP proved to be at least as efficient as its classic competitors.

Keywords: *Multi-layer Perceptron, Bayesian Paradigm; Liver Fibrosis Stadialization.*

1. Introduction

Hepatic fibrosis is the major indicator of progressive liver diseases and, therefore, also in the cases of patients with chronic hepatitis C. Liver fibrosis refers to the accumulation of tough, fibrous scar tissue in the liver that occurs in most types of chronic liver diseases. Advanced liver fibrosis results in cirrhosis, liver failure and portal hypertension, and often requires liver transplant. The precise stage is the most important predictor of disease progression and determines the need for antiviral therapy. From a medical point of view, liver fibrosis is evaluated semi-quantitatively according to the METAVIR F scoring system. Concretely, the five stages are the following: F0 – no fibrosis, F1 – portal fibrosis without septa, F2 –portal fibrosis and few septa, F3 – numerous septa without cirrhosis and F4 – cirrhosis.

A new trend in clinical practice is based on finding a correct method for the evaluation of the liver fibrosis (i.e. liver fibrosis stadialization) in a non-invasive way, using all together biochemical tests, imaging methods, and intelligent systems provided by Computer Science. A modern technological approach within the computer-aided medical diagnosis (CAMD) process in the evaluation of liver fibrosis is the Fibroscan® (Echosens, Paris, France), a specially adapted ultrasound device using the principle of the one-dimension transient elastography (TE) for the assessment of liver stiffness [1]. The practice of Fibroscan is based on establishing some cut-off values of the liver stiffness for each stage of fibrosis.

Neural networks (NNs) have become a popular tool in CAMD. The ability of NNs to learn from input data with or without a teacher makes them very flexible and powerful in medical diagnosis. Recent studies focus on the applications of different NNs algorithms to automatically diagnose a wide range of diseases [2-7].

A decision-making process combines, from a Bayesian point of view, prior knowledge with information extracted from observations. The idea behind the Bayesian paradigm is that one can predict the class label of an object, given its attributes values, by the use of the Bayes' rule. The Bayesian approach can be used to learn the weights in NNs. Some studies used Bayesian NNs to solve biomedical problems [8-10].

Different from other approaches dealing with the Bayesian paradigm in conjunction with network models and based on previous work [11], this paper proposes a new technique to update the synaptic weights in a multi-layer perceptron (MLP) using the correlation coefficient to update the synaptic weights. Technically, the associations between object's attributes and the network output, or the error function, respectively, are expressed through the Pearson's *r* correlation coefficient. The statistical comparison results indicated that the Bayesian-trained MLP proved to be at least as efficient as its classic machine learning competitors.

The remainder of this paper is organized in five sections. Section 2 is devoted to the presentation of both the design and implementation of the novel model, and the real-world liver fibrosis dataset for the benchmark process. Section 3 presents the experimental results of applying the model in terms of performance analysis and performance assessment. Section 4 deals with discussions and conclusions.

2. Materials and methods

2.1. Discovering knowledge in data. Bayesian learning paradigm

The training dataset $\{\mathbf{x}_1, \mathbf{x}_2,..., \mathbf{x}_N\}$ contains N objects, where each object is coded as a vector $\mathbf{x}_k = (x_1^k,...,x_i^k,...,x_p^k; y_j)$. x_i^k, i = 1, 2,..., p, represents the i-th feature of the k object, k = 1, 2,…,N, and y_j, j = 1, 2,..., q, represents the label of the decision class C_j the object $\mathbf{x}_k$ belongs to. In order to use the Bayesian approach to train the neural network, the connection between attributes and the network error is important to be quantified. From a probabilistic/statistical point of view, the set representing the attributes $\{ x_i^1, x_i^2,..., x_i^N \}$ can be seen as a random sample of length N corresponding to the random variable (r.v.) X_i. In the same way, one can interpret the set of labels $\{ y_j^1, y_j^2,..., y_j^N \}$ as a random sample of length N corresponding to the categorical r.v. *Y*.

Let $E(n)$ be the error of the network in iteration n, which can be viewed as well as a r.v.

A natural way to discover the potential information within data is to assess the statistical dependence between the r.v.'s X_i, i =1, 2,..., p, and the network error $E(n)$, using measures of association [12]. The most simple and direct manner to estimate a (linear) relationship between variables is represented by the Pearson's *r* correlation coefficient, given by:

$$r = r(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2 \cdot}}, \quad (1)$$

where $\{x_i\},\{y_i\},1 \le i \le n$, represents samples corresponding to the r.v.'s *X* and *Y*,

The Bayes' theorem, stating that:

$$P\{A_i \mid B\} = \frac{P\{B \mid A_i\}P\{A_i\}}{\sum_{i=1}^{n} P\{B \mid A_i\}P\{A_i\}},\ P\{B\} > 0,\ P\{A_i\} > 0,\ i = 1,\ 2,...,\ n, \quad (2)$$

where B is an arbitrary event, and $\{A_1, A_2, ..., A_n\}$ is a partition of the sample space Ω, can represent an alternative learning technique used to train NNs; $P\{A_i|B\}$ is known as *posterior probability*, $P\{A_i\}$ as *prior probability*, $P\{B|A_i\}$ as *likelihood*, and $P\{B\}$ as *evidence*. Note that, in classification/decision-making problems, given an object with attributes $\{A_1, A_2,..., A_n\}$ belonging to class C, one often assumes the so-called *naïve Bayes* hypothesis stating the independence of attributes for a given class C, namely:

$$P\{A_1, A_2, ..., A_n \mid C\} = P\{A_1 \mid C\} \cdot P\{A_2 \mid C\} \cdot P\{A_n \mid C\}. \quad (3)$$

The Bayesian paradigm may be used to learn the MLP weights by considering the concept of subjective probability instead of objective probability. As in [11], from a *subjective Bayesianism*, the probability is interpreted as a measure/degree of belief [13]. Accordingly, assuming that the real-valued synaptic weights belong to the interval [0, 1], the synaptic weights might be interpreted as probability-like measure encoding the strength of a connection. On the other hand, since the Pearson's r correlation coefficient is a measure of the strength of association, with values ranging from -1 to +1, it might be interpreted as probability-like measure encoding the strength of the relationship between two r.v.'s.

Let w_{ij} be the synaptic weights with corresponding r.v.'s W_{ij}. Suppose that the events A_{ij} corresponding to W_{ij} provide a partition of the "weight space" W. Under these circumstances, the MLP synaptic weights at iteration n are adjusted according to the formula:

$$w_{ij}(n+1) = P\{A_{ij} \mid E(n)\} = \frac{P\{E(n) \mid A_{ij}\}P\{A_{ij}\}}{\sum_{i,j} P\{E(n) \mid A_{ij}\}P\{A_{ij}\}},\ i = 1,\ 2,...,\ p,\ j = 1, 2,..., q, \quad (4)$$

where $P\{A_{ij}\}$ represents the *prior* (*probability*), $P\{E(n) \mid A_{ij}\}$ the *likelihood*, and $P\{E(n)\}$ the *evidence*.

2.2. The multi-layer perceptron model

Recall the elements of a feed-forward NN (or MLP):

- input vector $\mathbf{x} = (x_1, x_2,..., x_p)$ formed by p feature components x_i;
- related output/response (multivariate) variable y_j, $j = 1, 2,..., q$;
- (synaptic) weights w_{ij}, connecting the output of neuron i to the input of neuron j, where neuron j lies in a layer to the right of neuron i;
- activation non-linear function f.

A very popular example of continuously differentiable non-linear activation function commonly used in MLP is the logistic sigmoid:

$$f(x) = \frac{1}{1+e^{-x}}, \quad (5)$$

Practitioners recommend the use of normalized inputs instead of the original ones in order to increase the convergence speed, while an appropriate synaptic weights initialization allows the training algorithm to produce a good set of weights and may improve the training speed [14]. Therefore, normalized inputs and a standard synaptic weights initialization were used. A key observation in its practical use, based on the *universal approximation theorem* applied to MLP, is that a network with a single hidden layer is sufficient to uniformly approximate any continuous function [14], [15].

The following features characterize the proposed Bayesian-driven MLP (BMLP) model:

- One hidden layer with the number of hidden units equaling the number of decision classes;
- The standard logistic sigmoid as activation function;
- Standard initialization, normalized inputs, shuffled examples and batch training mode [14-15];
- Network output computed using the *winner-takes-all* paradigm (the neuron with the largest output value gives the decision class);
- Testing/generalization performance adequate to the problem at hand as stopping criterion.

BMLP algorithm

1. For each decision class C_j, j = 1, 2,..., q, and for each attribute A_i, i = 1, 2,..., p, compute the corresponding mean attribute value m_i^j.
2. For each hidden neuron HN_j, j = 1, 2,..., q, compute the linear discriminant u_j, given by:

$$u_j = \sum_{i=1}^{p} \left(x_i^k \cdot w_{ij} \cdot \frac{1}{(x_i^k - m_i^j)^2} \right), j = 1, 2,..., q, k = 1, 2,\ldots, \mathrm{N}. \quad (6)$$

3. For each hidden neuron HN_j, j = 1, 2,..., q, consider the non-linear activation function given by the sigmoid:

$$f(u_j) = \frac{1}{1+e^{-u_j}}, j = 1, 2,..., q. \quad (7)$$

4. For each decision class C_j, j = 1, 2,..., q, encode the corresponding label y_j using the "*1-of-q*" rule for nominal/categorical data, i.e., $y_1 \sim (0, 0,\ldots, 1)$, $y_2 \sim (0, 0,\ldots,1, 0),\ldots, y_q \sim (1, 0,\ldots, 0)$.
5. The hidden layer is seen as a discrete random variable, whose distribution is characterized by a probability mass function g, which values are given by:

$$g_j = \frac{\exp\left(f(u_j)\right) - \max_i\{f(u_i)\}}{\sum_{i=1}^{q}\left[f(u_i) - \max_i\{f(u_i)\}\right]}, j = 1, 2,..., q. \quad (8)$$

6. For each input item $\mathbf{x}_k$ of the training set TS, compute the corresponding error function as follows:

$$error_k = \sum_{j=1}^{q} (y_j - g_j),\ k = 1, 2, \ldots, \mathrm{N}. \tag{9}$$

7. Build the error array $E = (error_1, error_2, \ldots, error_N)$, using the error at each step.
8. Update the synaptic weights according to the formula:

$$w_{ij}(n+1) = w_{ij}^{*} = \frac{r\left((x_i^k - m_i^j), E\right) \cdot r\left((x_i^k - m_i^j), y_j\right)}{\sum_i r\left((x_i^k - m_i^j), E\right) \cdot r\left((x_i^k - m_i^j), y_j\right)},\ i = 1, 2, \ldots, p,\ j = 1, 2, \ldots, q, \tag{10}$$

where w_{ij}^{*} denotes the updated synaptic weight.

Repeat steps 2-8 for a certain number of epochs until the stopping criterion is satisfied. Return the synaptic weights w_{ij}^{*}.

2.3. Liver fibrosis dataset

The proposed BMLP model has been applied on a real-world medical dataset related to liver fibrosis described below. The dataset consists of 743 consecutive patients with chronic HCV infection, examined at the 3rd Medical Clinic, University of Medicine and Pharmacy "Iuliu Hatieganu", Cluj-Napoca, Romania, between May 2007 and August 2008. All of them had positive HCV-RNA in their serum and underwent percutaneous liver biopsy (LB) for grading and staging the diseases. All patients were referred to liver stiffness measurement (LSM) 1 day prior to LB. Besides the epidemiological, anthropometric and clinical parameters, the biological parameters were determined for all patients on the same day as LSM. Table 1 presents an example of the parameters (25 classification attributes) and the diagnosis classes for liver fibrosis (Metavir F), taken into consideration in our study for five patients (P_1-P_5), one for each decision class.

Table 1. Example of medical features for five patients

Features	**P_1**	**P_2**	**P_3**	**P_4**	**P_5**
Metavir F (stages)	*0*	*1*	*2*	*3*	*4*
Stiffness (Fibroscan specific output)	3.2	5.3	6	15.5	27
Sex (M/F)	1	2	1	2	1
Age (years)	33	55	31	50	46
BMI (body mass index) (kg/cm)	26.2	24.3	33.95	27.34	30.12
Glycemia (mg/dl)	89.3	110	96	118	84
Triglycerides (mg/dl)	208	54	154	54	93
Cholesterol (mg/dl)	263	133	198	143	152
High density lipoprotein cholesterol (mg/dl)	56	69	36	40	21
Aspartate aminotransferase (U/L)	20	62	46	62	105
Alanin aminotransferase (U/L)	59	61	117	76	167
Gama glutamyl transpeptidase (U/L)	75	19	41	40	187
Total bilirubin (mg/dl)	0.51	1.19	0.93	0.54	0.97
Alkaline phosphatase (U/L)	174	283	248	258	245
Prothrombin index (%)	108	102	107.3	97.1	42.7
Tqs (tocopheryl quinones) (mg/dl)	14.5	15.2	15	15.7	25.4

Follow Table 1. Example of medical features for five patients

Features	P_1	P_2	P_3	P_4	P_5
INR (prothrombin time ratio) (seconds)	0.87	0.98	0.95	1.02	1.85
Prolonged activated partial thromboplastin time (seconds)	27.9	30.9	28.4	29	30.9
Haematids (erytthrocytes) (mii/µl)	5.11	4.91	4.81	4.53	5.11
Hemoglobin (g/dl)	16.4	14.5	14.9	14.6	15.2
Hematocrit (%)	46.1	40.9	42.6	41.3	45.2
Medium eritrocity volume (fL)	91.1	83.2	88.6	91.1	88.5
Avg. eritrocitary hemoglobin (pg/cell)	32	29.5	31	32.4	29.7
Avg. concetration of hemoglobinin a red blood cell (g/dl)	35.2	35.5	35	35.5	33.5
Thrombocytes (mii/µl)	208	198	174	161	106
Sideraemia (mg/dl)	134	70	175.7	79.4	103

2.4. Statistical assessment

The effectiveness of the Bayesian-driven MLP algorithm in comparison with standard NN models, such as the classical MLP trained with the BP algorithm (MLPBP), the radial-basis function (RBF), the probabilistic neural network (PNN), and the support vector machine (SVM), was assessed by the 10-fold cross-validation. Technically, the testing classification accuracy for each model has been computed 10 times, each time leaving out one of the sub-samples and using that sub-sample as a test sample for cross-validation. In this way, each sub-sample is being used nine times as training sample and just once as testing sample. The model's correct classification rates, computed as mentioned above, were averaged to give the 10-fold estimate of the classification accuracy. This procedure has been repeated 10 times to complete a cross-validation cycle, consisting of the 10 runs of each model. This statistical procedure used for relative small databases is detailed in [16]. In the concrete case of the liver fibrosis dataset, around 669 cases are considered as learning samples, while the rest of 74 as testing samples in each of the ten cross-validation cycles.

The following rule to compare the algorithms was used: each algorithm has been executed in 100 independent computer runs (each model has been independently run 100 times in a complete cross-validation cycle), providing a statistical power greater than 95%, with type I error = 0.05 for the statistical comparison tests. The average accuracy obtained over the 100 complete cross-validation cycles represented the average decision performance of each competitor, subsequently used for the statistical comparison.

The statistical analysis assessing the effectiveness of BMLP in comparison with the other neural computing techniques involved two statistical tests applied to the independent samples of computer runs:

- The parametric *t*-test for independent samples;
- The difference between two proportions (DBTP) - two-sided test (*z*-value).

Note. Authors have made a Java implementation of the BMLP algorithm. It is worth mentioning that an important characteristic of the Java implementation is that all data about patients collected by physicians can, at any time, be added, modified or deleted, with no change in the source of the program whatsoever, because of JDBC (*Java Database Connectivity*).

3. Results

The results of the experiments on the liver fibrosis dataset, aiming to provide the effectiveness of this novel approach in liver fibrosis stadialization, are presented in Table 2 in terms of diagnosis accuracy and the corresponding standard deviation (SD), averaged over 100 computer runs of a complete cross-validation cycle.

Table 2. BMLP diagnosis performance averaged over 100 computer runs (mean/SD)

Mean accuracy (%)	SD (%)
61.97	4.10

Both the overall and per-class classification statistics for a sample corresponding to an average computer running are displayed in Table 3. Thus, we can establish: (*a*) the total number of cases in each class, (*b*) cases of each class that were correctly (and incorrectly) classified, and (*c*) cases of that class which could not be classified at all (unknown cases). The observed class is displayed at the top of the table, and the predicted class down the side. Each cell contains a number showing how many cases that were actually of the given observed class were assigned by the model to the given predicted class.

Table 3. Overall and per-class statistics

Class (Metavir F)	0	1	2	3	4
0	0	0	0	0	0
1	27	180	118	30	16
2	2	46	47	22	12
3	0	1	1	3	1
4	0	5	11	31	190

Table 3 provides a good insight regarding the difficulties that must face a classifier when applying on such a medical problem. As we can easily see, there is a quite real trend of misclassification regarding the intermediate stages (F2, F3), but concerning the early and the final stages (F1-beginning of the disease, F4-cirrhosis), which are the most important, the diagnosis is very promising.

Several standard neural networks, such as: MLPBP, RBF, PNN, and SVM [16], had been tested against the liver fibrosis dataset. To evaluate the performance of the novel approach, we compared it with the performance of these advanced models applied on the same dataset, described and reported in literature [17-19]. Their performance, displayed in Table 4, cannot be directly compared for all cases with the ones obtained by BMLP since the 10-fold cross-validation has not always been used.

Table 4. Statistical comparison of the diagnosis accuracy

Classifier	Mean/SD performance (%)
BMLP	61.97/4.10
MLPBP	60.81/2.26
RBF	55.35/3.45
PNN	53.64/3.11
SVM	61.33/NS

According to this result, the novel algorithm provided at least the same performance compared with other NNs approaches. Moreover, it is worth to mention that the computational effort was lower compared to the standard MLP trained with the BP algorithm. During the multiple computer runs, one can estimate that the average computing speed has increased by about 25% due to the simplified learning technique.

To quantify statistically the magnitude of the contrast between the performance of the novel approach and the performance of its competitors, both the *t*-test for independent samples and the difference between two proportions (DBTP) - two-sided test (*z*-value) have been considered. The results of the two statistical comparison tests using 100 random trials are displayed in Table 5.

Table 5. Benchmark results: *t*-test and DBTP test

Competitors	*t*-test (*p*-level)	DBTP (*p*-level)
BMLP vs. MLPBP	2.47/0.014	0.86
BMLP vs. RBF	12.31/0.00	0.34
B-MLP vs. PNN	16.32/0.00	0.23
B-MLP vs. SVM	-	0.92

The results of the two tests, though different, are consistent with the raw information from Table 4, confirming statistically the contrast. Basically, depending on the paradigm of each test, while the *t*-test for independent samples highlights a significant statistical difference between BMLP and the other NNs, the test concerning the difference between two proportions reveals no statistical significant difference.

4. Discussion and conclusions

This paper firstly deals with a novel learning technique for MLP, based on the Bayesian paradigm and replacing the standard BP algorithm. Secondly, the novel approach has been assessed in real-world application regarding the liver fibrosis stadialization.

Different from other approaches dealing with the Bayesian paradigm, the paper proposes a novel technique in the updating process, the synaptic weights being considered, from a subjective Bayesianism, as probabilities expressing the association between attributes and the error function through the Pearson's *r* correlation coefficient.

For the performance assessment, both the *t*-test for independent samples and the difference between two proportions (DBTP) - two-sided test (*z*-value) have been considered.

The idea to use the Bayesian paradigm to train a MLP is straightforward and advantageous in several aspects:

- The Bayesian approach for updating the synaptic weights is transparently presented.
- The corresponding algorithm is easy to understand and implement.
- The model is potentially adaptable to a wide variety of medical decision problems.

Finally, note that the performance on BMLP used for liver fibrosis stadialization equaled or exceeded the results reported in literature.

Future research may lie in:

- The use of alternative approaches to the Pearson's *r* correlation coefficient.
- The use of alternative non-linear activation functions.

Acknowledgment

This work was supported by the strategic grant POSDRU/159/1.5/S/133255, Project ID 133255 (2014), co-financed by the European Social Fund within the Sectorial Operational Program Human Resources Development 2007-2013. The liver fibrosis database was obtained from the 3rd Medical Clinic, University of Medicine and Pharmacy Cluj-Napoca, Romania. Thanks go to Prof. Radu Badea, Monica Lupsor, and Horia Stefanescu.

References

[1] 1.L. Sandrin, B. Fourquet, J.M. Hasquenoph, *et al*, "Transient elastography: a new noninvasive method for assessment of hepatic fibrosis", Ultrasound Med Biol. vol. 29, 2003, pp.1705-1713.

[2] 2. S.N. Ogulata, C. Sahin, and R. Erol, "Neural network-based computer-aided diagnosis in classification of primary generalized epilepsy by EEG signals," J. Med. Syst., vol. 33(2), 2009, pp. 107-112.

[3] 3. M. Kitajima, T. Hirai, S. Katsuragawa, *et al*., "Differentiation of common large sellar-suprasellar masses effect of artificial neural network on radiologists' diagnosis performance," Acad. Radiol., vol. 16(3), 2009, pp. 313-320.

[4] 4. K.C. Lai, H.C. Chiang, W.C. *et al*., "Artificial neural network-based study can predict gastric cancer staging," *Hepatogastroenterology*, vol. 55(86-87), 2008, pp. 1859-1863.

[5] 5. J.S. Chiu, Y.F. Wang, Y.C. Su, *et al*., "Artificial neural network to predict skeletal metastasis in patients with prostate cancer," *J. Med. Syst*., vol. 33(2), 2009, pp. 91-100.

[6] 6. J.G. Hincapie, and R.F. Kirsch, "Feasibility of EMG-based neural network controller for an upper extremity neuroprosthesis," IEEE Trans. Neural. Syst. Rehabil. Eng., vol. 17(1), 2009, pp. 80-90.

[7] 7. T. Ince, S. Kiranyaz, and M. Gabbouj, "Automated patient-specific classification of premature ventricular contractions," In Proc. Conf. IEEE Eng. Med. Biol. Soc., 2008, pp. 5474-5477.

[8] P.J.G. Lisboa, H. Wong, P. Harris, *et al*., "A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer", Artif Intell Med. vol. 28(1), 2003, pp. 1-25.

[9] D. Gao, M. Madden, D. Chambers, *et al*., "Bayesian ANN Classifier for ECG Arrhythmia Diagnostic System: A Comparison Study" In: Proc. IEEE Intl. Conference on Neural Networks, 2005, July 31 2005-Aug. 4, Montreal, Canada, vol. 4, 2005, pp. 2383- 2388.

[10] R. Donald, T. Howells, I. Piper, *et al*., "Early warning of EUSIG-defined hypotensive events using a Bayesian Artificial Neural Network", Acta Neurochir Suppl. vol. 114, 2012, pp. 39-44.

[11] S. Belciug, and F.Gorunescu, "Error-correction learning for artificial neural networks using the Bayesian paradigm. Application to automated medical diagnosis" Journal of Biomedical Informatics, 2014, DOI: 10.1016/j.jbi.2014.07.013. (*in press*)

[12] D.J. Sheskin, "Handbook of Parametric and Nonparametric Statistical Procedures" (3rd ed.): Chapman and Hall/CRC, 2004.

[13] J. Press, "Subjective and Objective Bayesian Statistics. Principles, Models, and Applications" (2nd ed.): Wiley, 2003 (http://onlinelibrary.wiley.com/doi/10.1002/9780470317105.fmatter/pdf; 17 Mar 2010).

[14] S. Haykin, "Neural networks. A comprehensive foundation" Prentice-Hall, 1999.

[15] C. Bishop, "Neural networks for pattern recognition" Oxford University Press, 1995.

[16] F. Gorunescu, "Data Mining. Concepts, Models and Techniques" Berlin Heidelberg: Springer-Verlag, 2011/2013.

[17] R. Stoean, C. Stoean, M. Lupsor, *et al*., "Evolutionary-driven support vector machines for determining the degree of liver fibrosis in chronic hepatitis C", Artificial Intelligence in Medicine vol. 51, 2011, pp. 53-65.

[18] F. Gorunescu, S. Belciug, M. Gorunescu, *et al.,* "Intelligent decision-making for liver fibrosis stadialization based on tandem feature selection and evolutionary-driven neural network" Expert Systems with Applications vol. 39(17), 2012, pp. 12824-12832.

[19] F. Gorunescu, and S. Belciug, "Evolutionary strategy to develop learning-based decision systems. Application to breast cancer and liver fibrosis stadialization", Journal of Biomedical Informatics vol. 49, 2014, pp. 112-118.