IMPROVING PERFORMANCE IN NEURAL NETWORKS USING FEATURE SELECTION BASED ON RANDOM FORESTS: APPLICATION TO AUTOMATED MEDICAL DIAGNOSIS

Smaranda Belciug, Florin Gorunescu

Abstract: *Neural networks are well-known for their effectiveness as classifiers. In this context, feature selection has become a usefulness technique to make the classifier faster, cost-effective, and more accurate. The effect of a random forests-based feature selection on the classification accuracy of a multi-layer perceptron has been explored in this paper. Up to 6% improvement in classification accuracy and 40% improvement in computation speed has been observed when using the tandem model on five real-world publically available medical datasets regarding colon cancer, breast cancer, diabetes, thyroid and fetal heartbeat databases.*

Keywords: *multi-layer perceptron, feature selection, random forests, automated medical diagnosis*

ACM Classification Keywords: *F.1 Computation by abstract devices – Self-modifying machines, H.2.8 Database Applications – Data mining*

Introduction

The use of machine learning (ML) tools in processing medical data represents nowadays an important trend [Yang, 2010]. Intelligent medical decision systems, built using state-of-the-art ML algorithms, are particularly conceived for automatic diagnosis, and this approach became a *fruitful niche* within the vast Bioinformatics field.

Within the most powerful ML algorithms one can mention, in particular, the neural networks (NNs). NNs represent “massively parallel distributed processors” [Haykin, 1999], and are used to model arbitrary non-linear data. The computational metaphor behind them is their unique property of mimicking the human brain knowledge-processing paradigm. The most used NN architecture is represented by the multi-layer perceptron (MLP), which stores knowledge in its synaptic weights. The back-propagation (BP) algorithm [Bishop, 2004] is the most popular learning algorithm for MLP. It is based on the minimization of the sum-of-squares error (SSE) using the gradient descent technique.

NNs have been successfully used in automated medical diagnosis [Amato et al., 2013], [Belciug and Gorunescu, 2014], breast cancer detection and recurrence [Kalteh et al., 2013], [Belciug and

Gorunescu, 2013], cardiovascular diseases [Nakajima et al., 2015], diabetes [Nguyen et al., 2014], colon cancer [Spelt et al., 2013], Alzheimer's disease [Coppedè et al., 2013], etc.

Different approaches to improve the NNs performance have been proposed. Various methods underlie them, such as nearest neighbor [Alejo et al., 2006], clustering [Madasamy and Tamilselvi, 2013], hybridization with genetic algorithms [Belciug and Gorunescu, 2013], Bayesian learning [Belciug and Gorunescu, 2014], statistical tools (discriminant function analysis, regression analysis, correlation) along with NN sensitivity analysis [Gorunescu et al., 2012], swarm optimization [Dheeba and Selvi, 2012], etc.

An effective way to improve the NN classification accuracy, regardless of the use of different algorithms to strengthen their performance, resides in using the feature selection (FS). The direct use of medical databases without previous analysis and filtering is often counterproductive, even when using efficient algorithms such as MLP. FS represents an efficient way to solve this issue by choosing the most relevant attributes from data. There are different approaches regarding FS [Liu and Motoda, 2007]. One can mention association rules [Karabatak and Ince, 2009], particle swarm optimization [Shen et al., 2009], genetic algorithms [Marinakis et al., 2009], hill climbing [Stoean et al., 2011], statistical tools [Gorunescu et al., 2012], etc.

Different from other approaches trying to improve the MLP classification accuracy, the current work proposes a FS-based technique obtained using Random Forests (RFs). RFs are among the most popular ML algorithms due to their relatively good accuracy and robustness [Breiman, 2001]. They are used for classification and regression, and it is noteworthy that they provide an efficient method for FS. In this paper, RFs are applied in the beginning on the whole dataset using the method of mean decrease impurity. The feature (predictor) importance is then computed, and all the features are ranked accordingly. Next, the features with the rank higher than a default threshold are chosen as input variables (independent predictors) for MLP. In this way, the most important features will decide the class (label) for a certain item. The statistical comparison indicates that the novel tandem FS-MLP outperforms the conventional MLP algorithm regarding both the decision accuracy and the computation speed.

The remainder of this paper is organized in four sections. Section 2 briefly presents the MLP, FS, and RF techniques, focusing on the detailed presentation of the tandem RF-MLP model, along with the five publically available datasets. Section 3 presents the experimental results in terms of performance analysis and performance assessment, while Section 4 deals with the conclusions and future work.

Methodology

This section briefly outline the principles of MLP, usefulness of using FS along with MLP, and the RF-based FS procedure used in this paper. A detailed description of the novel tandem FS-MLP model is provided at the end of the section.

Multi-layer perceptron

The key paradigm underlying NNs is the information processing system consisting of a large number of highly interconnected processing units called *neurons*, organized in a layered parallel structure called *network*. The *learning* process is typically achieved through progressive adjustments of synaptic weights in order to attain a desired design objective. A neuron consists of a set of inputs x_i, weighted by the corresponding synaptic weights w_i, added together and passed onto an activation function which bounds the acceptable amplitude range of the neuron output to some finite value. The MLP model, which is the most commonly used NN architecture, is based on computational units, which compute a non-linear function of the scalar product of the input vector and the synaptic weight vector. The MLP architecture involves some critical hyper-parameters in its design, such as the network depth, the number of hidden neurons per layer, the initial learning rate, and the momentum. A key observation in the practical use of MLP, due to a theorem by Kolmogorov, is that MLP with only one hidden layers, i.e. a *three-layer perceptron* (3-MLP), is theoretically sufficient to model almost any real-life problem [Bishop, 2004], [Haykin, 1999].

There is a two-fold information process consisting of the *training phase* and the *testing phase*. In the training phase, a training dataset is used to determine the weight parameters w_i that define the neural model. Next, the trained neural model is used to process (unknown) test patterns, yielding the true classification performance. To summarize, MLP is firstly trained to associate outputs with input patterns, and, secondly, it provides the output that corresponds to a taught input pattern that is least different from the given pattern.

Feature selection

Based on some filtering criteria, such as statistical tools, ML algorithms, etc., FS techniques choose the most relevant attributes from the original dataset and remove unimportant or redundant features. FS may be considered a key factor for pattern recognition, since even the best classifiers will perform poorly if the features are not chosen appropriately.

The feature selection methods may be broadly categorized as filters and wrappers [Yuan et al., 1999]. Thus, a filter method selects features based only on the intrinsic characteristic of the data, while a wrapper method uses a search model as part of evaluation of the relevance of a feature. Since RFs are a popular method for feature ranking, we took advantage of this ability to select good features, and used the Data mining module (Random Forests for Regression and Classification) within Statistica 7 package, StatSoft. Inc. to choose the most appropriate features for the classification process.

Random Forests

RFs are “a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest” [Breiman, 2001]. Each node in the tree predictors is a condition on a single feature, splitting the dataset into two subsets, such that similar response values reach eventually the same set. The usual impurity measures used for classification are either Gini impurity or information gain/entropy. For classification problems, the tree response takes the form of a class membership, which associates a set of independent (predictor) variables with one of the categories present in the dependent variable.

In the tree training process, it can be computed how much each feature decreases the weighted impurity. The feature (predictor) importance is computed by summing the drop in node impurity over all nodes in the tree(s), and expressing these sums relative to the largest sum found over all predictors. It is noteworthy that this is different from the notion of predictor or variable importance [Breiman et al., 1984].

In the FS process, we have considered the *predictor importance* representing the importance of each feature in the classification process, so that we can distinguish the features that make the major contributions to the prediction of the decision class. The predictor importance is ranking on a 0-100 scale basis for each predictor variable in the analysis.

Tandem RF-MLP model

The novel decision model consisting of the tandem RF-MLP is based on the following algorithm.

Algorithm RF-MLP

1. Run the RF algorithm on the dataset using the Statistica 7 module “Random Forests for Regression and Classification” with the following settings:

- Equal misclassification costs;

- Estimated prior probabilities (the likelihood that a case will fall into one of the classes is proportional to the dependent variable class sizes);

- Hold out cross-validation (2/3 was considered as training set, and 1/3 as testing set);

2. Rank features on a 0-100 scale through the “Predictor importance” RF feature, with the default threshold 50.

3. Keep the most influential features (i.e., rank higher than 50) for the subsequent MLP use.

4. Apply the MLP classifier to the selected features as network inputs.

Remark. It is worth noting that the issue of establishing the optimal threshold, keeping the most influential attributes, remains an open issue, depending on the concrete decision problem to solve. The number of MLP hidden units has been chosen heuristically for each dataset, in order to obtain optimal classification performance at the same time with the network simplicity. Concretely, we have 8 for CC, 10 for BCWD, 7 for PID, 13 for THY, and 15 for CARD.

Datasets

Five different real-world datasets have been used to evaluate the usefulness of the proposed approach. They are publically available, enabling the evaluation of the novel methodology and the comparison against other state-of-the-art techniques. Their sizes range from 368 to 7200, the number of input variables, both numerical and categorical, ranges from 7 to 21, and the number of output variables ranges from 2 to 10. Table 1 provides the details of each dataset.

Table 1. Datasets

Dataset	No. of instances	Input variables	Output variables (classes)
Colon Cancer (CC)	368	8 (numerical, categorical)	3 (short stay, medium stay, long stay)
Breast Cancer Wisconsin Diagnostic (BCWD)	569	10 (numerical)	2 (benign, malign)
Pima Indian Diabetes (PID)	768	7 (numerical, categorical)	2 (negative, positive)
Thyroid (THY)	7200	21 (numerical, categorical)	3 (degree: 1, 2, and 3)
Cardiotocography (CARD)	2126	21 (numerical, categorical)	10 (calm sleep, REM sleep, calm vigilance, active vigilance, shift pattern, accelerative/decelerative pattern, largely decelerative pattern, flat-sinusoidal pattern, suspect pattern)

The datasets originate from the University of Medicine and Pharmacy of Craiova, Romania, and from the UCI Machine Learning Repository. Details of the envisaged datasets: https://sites.google.com/site/imediatreat/data-sets, http://archive.ics.uci.edu/ml/datasets.html

Remark. The above datasets are not balanced. It is well-known that the success of ML classifiers is quite limited when they are applied on imbalanced datasets. To overcome this situation, we have considered the prior probabilities of each decision class, reflecting the degree of belief in class memberships of instances (i.e., whether an instance (input variable) belongs to a certain class (output variable)). In this regard, we have considered the prior probability as the percentage of each class (category) in the dataset, as a reasonable approach.

Results

Experimental results on real-world medical datasets are presented in this section. They are directly compared with results obtained by using six well-known state-of-the-art ML algorithms on the same datasets, in order to highlight the effectiveness of the proposed model.

Experimental results

The performance of the proposed tandem RF-MLP model has been assessed by using two important statistical measures: the average (testing) accuracy (ACC), and the corresponding standard deviation (SD) obtained by (independently) running the model 100 times in a complete hold out cross-validation cycle. The purpose of using these two statistical parameters was to highlight both the level of classification performance achieved by the model (ACC), and the 'stability' of the classification accuracy obtained in multiple independent computer runs (SD), because the proposed algorithm is of stochastic nature and ACC differs from a computer run to another. Thus, smaller SD indicates evidence of a more stable model with respect to different computer runs. For binary classification, we have also considered the area under the ROC curve (AUC).

Taking advantage of the RF ability to select the most important features for the classification process, we have considered for each dataset the features with the rank higher than a default threshold equaling 50 to a 0-100 scale, although any other choice being possible (user-defined). The experimental results regarding RF-MLP are displayed in Table 2 in terms of ACC and SD, along with the corresponding selected features and the running time (RF-MLP vs. 3-MLP) obtained in 100 complete hold out cross-validation cycles.

Table 2. Experimental results

Datasets	CC	BCWD	PID	THY	CARD
No. of selected features	5 out of 8	6 out of 10	6 out of 8	2 out of 21	12 out of 21
Performance measures (%)	ACC/SD	ACC/SD	ACC/SD	ACC/SD	ACC/SD
	71.17/3.72	91.14/1.54	75.06/3.81	95.69/0.30	84.08/0.85
Running time RF-MLP/3-MLP	2’04”/2’14”	2’57”/3’12”	3’42”/3’47”	16’28”/27’19”	9’47”/7’12”

As expected, the classification performance (i.e., testing accuracy) strongly depends on each dataset, ranging from 71% (CC dataset) to 96% (THY dataset). On the other hand, the 'stability' of the algorithm, measured through SD, also depends on each dataset, ranging from 0.30% (THY dataset) to 3.81%

(PID dataset). It is noteworthy that the best classification accuracy corresponds to the greatest stability, obtained in the case of the THY dataset (95.69%, 0.30%), although this is the largest dataset (7200 instances). Surprisingly in this case is that FS kept only 2 attributes (i.e., thyroid stimulating hormone, and free thyroxin index) out of 21. The use only of the two attributes provided about 96% classification accuracy, proving once again the significant role of FS in classification issues.

To synthetically illustrate the 'stability' of the model during different computer runs, a “box-and-whisker” plot displaying the ACC variability over the 100 independent computer runs is depicted in Fig. 1.

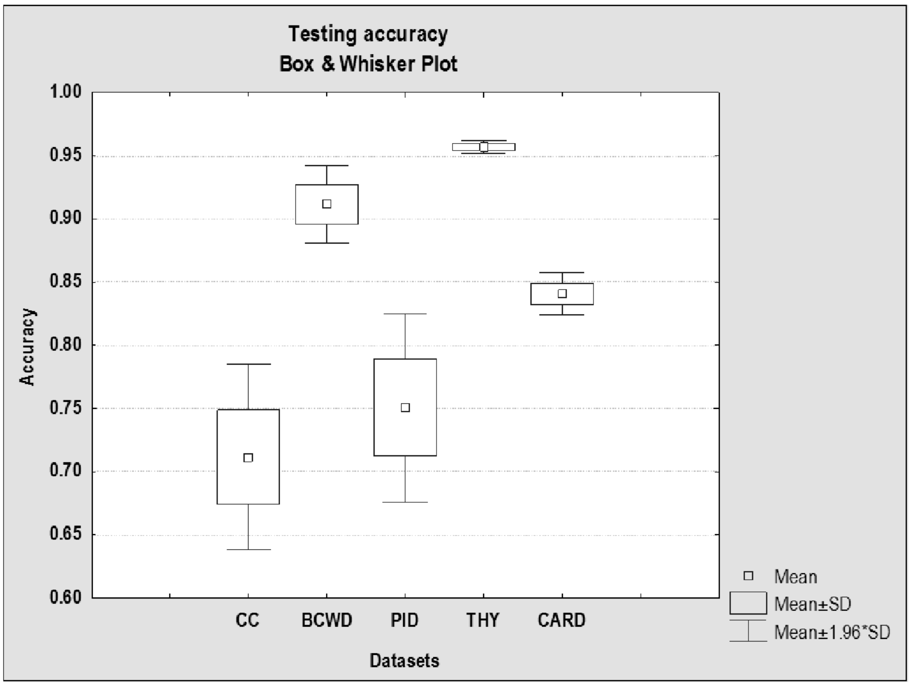

Fig. 1. Box and whisker plot (performance variability illustration)

The “box-and-whisker” plot summarizes the central tendency (central line), the variability around the central tendency (box), and the range of the variable (whiskers around the box). Low SD values (narrow box and close whiskers) mean high model 'stability, in other words, the classification accuracy does not depend significantly on the computer run. Thus, the best 'stability' along with the highest accuracy have been obtained for THY dataset, while the lowest performance has been obtained for CC and PID datasets.

Benchmark: RF-MLP performance evaluation

Table 3 presents the experimental results, in terms of ACC and corresponding AUC, obtained by each ML algorithm during 100 independent computer runs in complete hold out cross-validation cycle.

Table 3. Experimental results: Acc (%), AUC

Model		Classification performance ACC/AUC)			
	CC	BCWD	PID	THY	CARD
	Acc -	Acc AUC	Acc AUC	Acc -	Acc -
RF-MLP	71.17 -	91.14 0.987	75.06 0.859	95.69 -	84.08 -
3-MLP	68.34 -	90.39 0.984	70.73 0.816	95.67 -	81.74 -
RBF	63.57 -	87.03 0.921	71.44 0.819	93.75 -	75.10
PNN	62.63 -	70.92 0.769	65.75 0.796	93.12 -	72.46
SVM	58.69	94.55	76.56	92.58	72.93
n-B	63.04	92.30	75.52	95.27	57.33
k-NN	63.14	93.78	68.75	94.57	75.37

From Table 3 one can see that:

- Overall, the performance of all classifiers strongly depends on the specific dataset used, as expected;

- The use of FS through RF improved the accuracy of the conventional MLP;

- As compared with other state-of-the-art ML algorithms, the tandem RF-MLP model provided a better performance.

In conjunction with Table 3, the visual comparison of the classification performance of the five ML competitors, illustrated in Fig. 2., synthetically reveals the difference between them also depending on dataset.

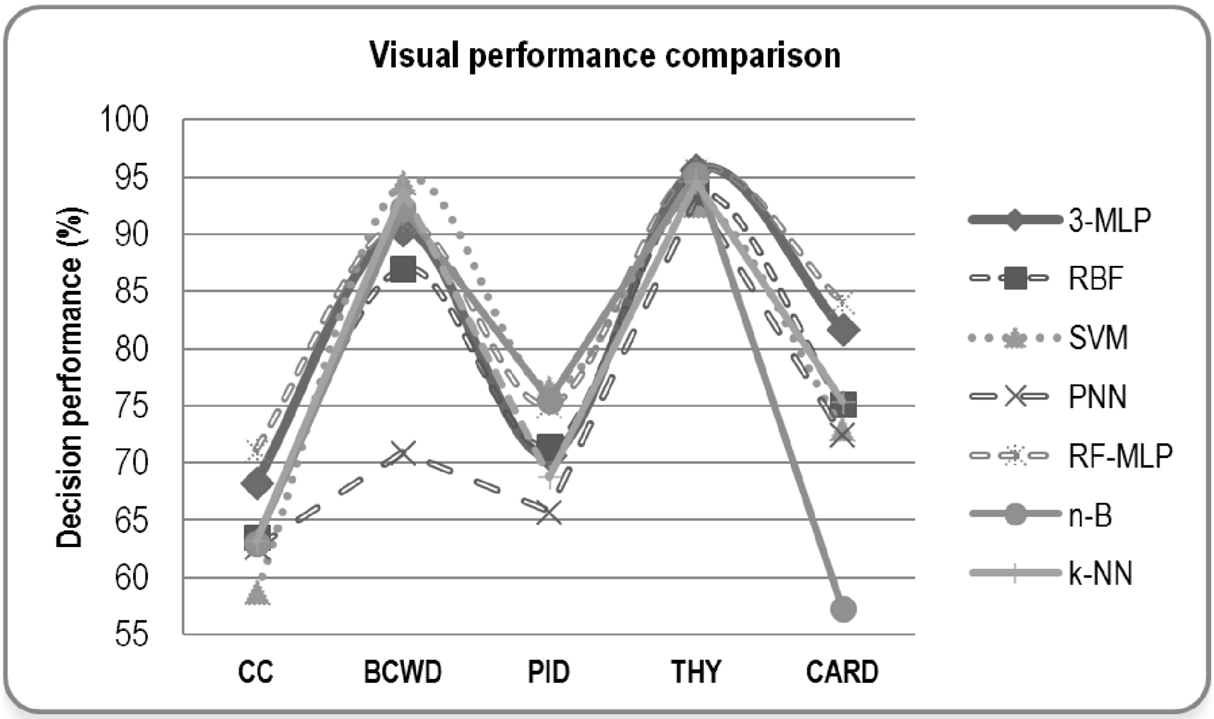

Fig. 2. Visual comparison of the ML algorithms performance

Next, the one-way ANOVA technique along with the Tukey's honestly significant difference (Tukey HSD) *post-hoc* test have been used to statistically address the difference in classification accuracy of the ML algorithms involved in the benchmarking process. The one-way ANOVA along with the Tukey HSD test were performed using the IBM SPSS 21.0 package. The ANOVA output, consisting of (combined) sums of squares (SS), degrees of freedom (df), mean squares (MS), F-value, and *p*-level (contrasts: quadratic polynomial), is presented in Table 4.

Table 4. One-way ANOVA comparison for mean testing accuracy

Dataset	SS	df	MS	F-value	*p*-level
CC	4866.08	3	1622.03	53.63	0.00
BCWD	26895.24	3	8965.08	115.87	0.00
PID	4397.19	3	1465.73	132.97	0.00
THY	522.52	3	174.17	35.19	0.00
CARD	8939.06	3	2979.69	55.99	0.00

The *post-hoc* Tukey HSD test has revealed statistically significant differences in classification performance (*p*-level < 0.05) regarding the following ML algorithms/datasets:

- RF-MLP vs. 3-MLP (mean diff. = 2.83, std. err. = 0.78), RF-MLP vs.RBF (mean diff. = 7.59, std. err. = 0.78),
RF-MLP vs. PNN (mean diff. = 8.54, std. err. = 0.78) on CC dataset;

- RF-MLP vs.RBF (mean diff. = 4.11, std. err. = 1.24), RF-MLP vs. PNN (mean diff. = 20.22, std. err. = 1.24) on BCWD dataset;

- RF-MLP vs. 3-MLP (mean diff. = 4.32, std. err. = 0.47), RF-MLP vs.RBF (mean diff. = 3.61, std. err. = 0.47),
RF-MLP vs. PNN (mean diff. = 9.30, std. err. = 0.47) on PID dataset;

- RF-MLP vs.RBF (mean diff. = 1.93, std. err. = 0.31), RF-MLP vs. PNN (mean diff. = 2.57, std. err. = 0.31) on THY dataset;

- RF-MLP vs.RBF (mean diff. = 8.97, std. err. = 1.03), RF-MLP vs. PNN (mean diff. = 11.61, std. err. = 1.03) on CARD dataset.

The comparison regarding classification performance between RF-MLP and the non-neural models (SVM, n-B, *k*-NN) has been assessed using the follow-up two-sided *z*-test. Statistically significant differences have been disclosed on CARD dataset, i.e., RF-MLP vs. SVM (*p*-level = 0.05), and RF-MLP vs. *k*-NN (*p*-level = 0.0001).

To complete the benchmarking process, we have presented in Table 5, the comparison between the novel tandem model RF-MLP and other well-established ML classifiers, i.e., optimal discriminant plane (ODP), regularized discriminant analysis (RDA), hierarchical pyramid neural network (HPNN), Cox regression (CR), partially connected neural network (PCNN), hybrid MLP/genetic algorithm (MLP/GA), decision trees (DT), logit (LOG), and random forest (RF), with results reported in literature [Belciug and Gorunescu, 2013], [Stoean et al. 2015], [Gorunescu and Belciug, 2014], [Aeberhard et al., 1994], [Wilson and Martinez, 1997], [Kinney, 1988], [Sahin and Subasi, 2015].

Remark. It should be noted that a direct comparison with RF-MLP is not very suitable since in these studies the ML algorithms taken into considerations were not applied to all the datasets, and the results have not been obtained using the same experimental settings. Let us also notice more or less important differences between the performances obtained by simulation using Statistica 7 package and displayed in Table 3, and those reported in literature, obtained by ML algorithms specially designed by authors.

Table 5. BPSS-MLP performance compared to other ML algorithms

Model	Dataset-Acc (%)				
	CC	BCWD	PID	THY	CARD
RF-MLP	71.17	91.14	75.06	95.69	84.08
RBF	-	87.42	70.83	-	-
PNN	-	71.08	65.62	-	-
MLP	71.2	81.39	74.47	92.85	97.78
k-NN	-	94.12	71.90	-	98.40
ODP	-	-	-	-	-
RDA	-	-	-	-	82.10
HPNN	-	-	-	-	-
CR	-	-	-	-	-
SVM	73	96.92	76.30	-	-
PCNN	-	81.07	-	-	-
MLP/GA	-	93.58	-	-	-
DT	72.32	-	-	-	86.36
LOG	60	-	-	-	-
RF	-	-	-	-	99.18

Conclusion

In the automated medical diagnosis, hybridizing in a collaborative manner state-of-the-art ML algorithms and latest medical approaches has become a very important interdisciplinary technology. The effectiveness of improving the performance of NNs, based on FS in tandem with the traditional MLP, was investigated on the task of providing a reliable intelligent decision support system for the automated diagnosis of various diseases. The performance of the novel model equaled or exceeded the results reported in literature on publicly available colon cancer, breast cancer, diabetes, thyroid, and cardiotocography.

Future research has to investigate at least the use of FS provided by RFs in tandem with other state-of-the-art ML algorithms, or hybrid algorithms.

Acknowledgements

The authors acknowledge the support of the research grant no. 26/2014, code PN-II-PT-PCCA-2013-4-1153, entitled IMEDIATREAT - Intelligent Medical Information System for the Diagnosis and Monitoring of the Treatment of Patients with Colorectal Neoplasm - financed by the Romanian Ministry of National Education (MEN) - Research and the Executive Agency for Higher Education Research Development and Innovation Funding (UEFISCDI). The CC database was part of this support. The BCWD database was obtained from the University of Wisconsin Hospitals, Madison. Thanks go to Dr. William H. Wolberg, W. Nick Street and Olvi L. Mangasarian for providing the data. For the PID database thanks go to Vincent Sigillito, John Hopkins University, the donor of the data. For the THY database thanks go to the Garavan Institute, Sydney, Australia, and R. Quinlan (documentation). For the CARD database thanks go to Marques de Sa, J. Bernades, and A. de Campos, Portugal.

Bibliography

[Aeberhard et al., 1994] S.Aeberhard, D.Coomans, O.de Vel. The performance of statistical pattern recognition methods in high dimensional settings. IEEE Signal Processing Workshop on Higher Order Statistics. Ceasarea. 1994;4:14-16.

[Alejo et al., 2006] R.Alejo, V.Garcia, J.M.Sotoca, R.A.Mollineda, J.S.Sanchez. **Improving the classification accuracy of RBF and MLP neural networks trained with imbalanced samples. Proc.** Proceedings of the 7th international conference on Intelligent Data Engineering and Automated Learning -IDEAL'06, 464-471, Springer, 2006.

[Amato et al., 2013] F.Amato, A.Lopez, E.M.Pena-Mendez, et al. Artificial neural networks in medical diagnosis. J Appl Biomed 2013;11:47–58.

[Belciug and Gorunescu, 2013] S.Belciug, F.Gorunescu. A hybrid neural network/genetic algorithm system applied to the breast cancer detection and recurrence. Expert Syst 2013;30(3):243-254.

[Belciug and Gorunescu, 2014] Belciug S, Gorunescu F. Error-correction learning for artificial neural networks using the Bayesian paradigm. Application to automated medical diagnosis. J Biomed Inform. 2014 Dec;52:329-337.

[Bishop, 2004] C.Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 2004.

[Breiman et al., 1984] L.Breiman, J.H.Friedman, R.A.Olshen, C.J.Stone. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

[Breiman, 2001] L.Breiman. Random Forests. Machine Learning. 2001;45(1):5-32.

[Coppedè et al., 2013] F.Coppedè, E.Grossi, M.Buscema, L.Migliore. Application of artificial neural networks to investigate one-carbon metabolism in Alzheimer's disease and healthy matched individuals. PLoS One. 2013; 8(8): e74012.

[Dheeba and Selvi, 2012] J.Dheeba, S.T.Selvi. A swarm optimized neural network system for classification of microcalcification in mammograms. J Med Syst 2012;36(5):3051-3061.

[Gorunescu et al., 2012] F.Gorunescu., S.Belciug, M.Gorunescu, R.Badea. Intelligent decision-making for liver fibrosis stadialization based on tandem feature selection and evolutionary-driven neural network. Expert Syst Appl. 2012;39: 12824-12833.

[Gorunescu and Belciug, 2014] F.Gorunescu, S.Belciug. Evolutionary strategy to develop learning-based decision systems. Application to breast cancer and liver fibrosis stadialization. J Biomed Inform 2014;49:112-118.

[Haykin, 1999] S.Haykin. Neural Networks. A comprehensive foundation (2nd Ed.). Prentice Hall International. 1999.

[Kalteh et al., 2013] A.A.Kalteh, P.Zarbakhsh, M.Jirabadi, J.Addeh. A research about breast cancer detection using different neural networks and K-MICA algorithm. J Cancer Res Ther 2013;9(3):456-466.

[Karabatak and Ince, 2009] M.Karabatak, M.C.Ince. An expert system for detection of breast cancer based on association rules and neural network. Expert Syst Appl. *2009;*36:3465-3469.

[Kinney, 1998] E.Kinney. Cox Regression application communication Report with TR-10-88. Harvard University, Center for Research in computing Technology, Aiken Computation Laboratory, 1988.

[Liu and Motoda, 2007] H.Liu, H.Motoda. (Eds.), In: Computational methods of feature selection. Ed. H.Liu and H.Motoda. Chapman and Hall/CRC, 2007.

[Madasamy and Tamilselvi, 2013] B.Madasamy, J.Tamilselvi. Improving classification Accuracy of Neural Network through Clustering Algorithms. Intl J of Computer Trends and Technology. 2013;4(9):3242-3246.

[Marinakis et al., 2009] Y.Marinakis, G.Dounias, L.Jantzen. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. *Comput. Biol. Med*. 2009;39 (1):69–78.

[Nakajima et al., 2015] K.Nakajima, S.Matsuo, H.Wakabayashi et al. Diagnostic Performance of Artificial Neural Network for Detecting Ischemia in Myocardial Perfusion Imaging. Circ J. 2015;79(7):1549-56.

[Nguyen et al., 2014] L.L.Nguyen, S.Su, H.T.Nguyen. Neural network approach for non-invasive detection of hyperglycemia using electrocardiographic signals. Proc. IEEE Conf Eng Med Biol Soc. 2014:4475-4478.

[Sahin and Subasi, 2015] H.Sahin, A.Subasi. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. J. Applied Soft Computing 2015;33(C):231-238.

[Shen et al., 2009] Q.Shen, Z.Meia, B-X.Ye. Simultaneous genes and training samples selection by modified particle swarm optimization for gene expression data classification. Comput. Biol. Med. 2009;39(7):646–649.

[Spelt et al., 2013] L.Spelt, J.Nilsson, R.Andersson, B.Andersson. Artificial neural networks - a method for prediction of survival following liver resection for colorectal cancer metastases. Eur J Surg Oncol. 2013;39(6):648-54.

[Stoean et al., 2011] C.Stoean, R.Stoean, M.Lupsor, H.Stefanescu, R.Badea. Feature selection for a cooperative coevolutionary classifier in liver fibrosis diagnosis. Comput. Biol Med. 2011;41:238-246.

[Stoean et al., 2015] R.Stoean, C.Stoean, A.Sandita et al. *Ensemble of Classifiers for Length of Stay Prediction in Colorectal Cancer.* In: Intl. Work-Conference on Artificial Neural Networks (IWANN 2015), Advances in Computational Intelligence, Lecture Notes in Computer Science 2015;9094:444-457.

[Wilson and Martinez, 1997] D.R.Wilson, T.R.Martinez. Improved Center Point Selection for Probabilistic Neural Networks. In: Proc. Intl. Conference on Artificial Neural Networks and Genetic Algorithms (ICANNGA'97). 1997:514-517.

[Yang, 2010] Z.R.Yang. Machine learning approaches to Bioinformatics. World Scientific, Science, Engineering, and Biology Informatics, 2010.

[Yuan et al., 1999] H.Yuan, S-S.Tseng, W.Gangshan, Z.Fuyan. A two-phase feature selection method using both filter and wrapper, Proc. IEEE Conference on Systems, Man, and Cybernetics. 1999;2:132-136.

Authors' Information

Smaranda Belciug *–Lecturer, PhD, Department of Computer Science, Faculty of Sciences, University of Craiova, Craiova, Romania, e-mail: sbelciug@inf.ucv.ro; H-index (ISI Web of Knowledge = 6, SCOPUS = 6, Google Scholar = 9)*

Major Fields of Scientific Research: Data Mining, Artificial Intelligence, Machine Learning, Statistics

Florin Gorunescu *– Professor of Biostatistics and Computer Science, University of Medicine and Pharmacy of Craiova, Craiova Romania; e-mail: gorunef@gmail.com; H-index (ISI Web of Knowledge = 12, SCOPUS = 14, Google Scholar = 18)*

Major Fields of Scientific Research: Data Mining, Artificial Intelligence, Machine Learning, Statistics, Operational Research, Mathematics