

Multi-Objective Performance Evaluation Using Fuzzy Criteria: Increasing Sensitivity Prediction for Outcome of Septic Shock Patients

Abigail L. Horn *, Federico Cismonti *, André S. Fialho *,
Susana M. Vieira **, João M.C. Sousa **, Shane Reti ***,
Michael Howell ***, Stan Finkelstein *

* *Massachusetts Institute of Technology, Engineering Systems Division, 77
Massachusetts Avenue, 02139 Cambridge, MA, USA
(email: abbyhorn@mit.edu)*

** *Technical University of Lisbon, Instituto Superior Técnico, Dept. of
Mechanical Engineering, CIS/IDMEC – LAETA, Av. Rovisco Pais, 1049-001
Lisbon, Portugal*

*** *Beth Israel Deaconess Medical Center, Department of Medicine, 1330
Beacon St., 02446 Brookline, MA, USA*

Abstract: This work introduces a fuzzy multi-objective function for performance evaluation in feature selection for a knowledge based medical classification task. Fuzzy optimization was used to combine the sensitivity, specificity, and percentage of correct classifications as objectives, and assign flexible goals to strengthen the sensitivity. Using the same soft computing methods and publicly available septic shock database as in Fialho et al. (2010), an increase in sensitivity of 30% was achieved while maintaining the percentage of correct classifications. Feature subsets selected were different from those previously published, lending insight into factors more integrally connected to what causes risk of death for patients.

Keywords: fuzzy models; neural network models; multi-objective optimization; medical systems; decision support systems; feature selection

1. INTRODUCTION

Sepsis is a syndrome characterized by an overwhelming systemic response to infection, which can rapidly lead to loss of limbs, organ dysfunction, and ultimately death (SCCM, 1992). Septic shock, an advanced stage of sepsis, is the leading cause of death in the non-coronary intensive care unit (ICU) and carries a mortality rate of about 50%. The incidence of severe sepsis is expected to rise as the population ages, and carry with it a high cost of treatment when compared with other ICU patients. Unfortunately, the diagnosis of infection leading to mortality in critically ill patients is challenging because a unique set of infection markers does not exist (Angus and Wax, 2001). Few population-based prospective cohort studies allow accurate delineation of the risk factors for sepsis and its outcome (Muller et al., 2000).

To assist in the outcome classification of septic patients, multiple approaches have applied knowledge based methods to septic shock patient data to construct a classifying model (Fialho et al., 2010; Paetz, 2003). Like other medical diagnostic decision makers, the task is one of employing multiple features to classify patients as negative or positive cases. In describing the performance of binary classifiers, the percentage of correct classifications (PCC) cannot be considered alone¹. The sensitivity, or hit rate, and specificity, or true rejection rate, must

¹ The percentage of correct classifications is equal to the number of true positives (TP) together with the number of true negatives (TN) divided by the total number of patients, and then multiplied by 100.

also be analyzed². In medical diagnosis and in the machine learning community, the standard method for combining these two measures into the evaluation task is in the analysis of the area under the ROC curve (AUC) (Swets, 1988; Egan, 1975; Hanley and McNeil, 1982). For a full discussion on the meaning and use of the AUC index, readers are referred to Fawcett (2006).

Though the AUC is the golden standard for performance evaluation, it is not an appropriate measure of classification accuracy in specific modeling tasks (Lobo et al., 2008; Hanczar et al., 2010; Hand, 2009; Lee and Hsiao, 1996). Among other criticisms, it does not give information about the spatial distribution of model errors, and weights omission and commission errors equally. This means it fails to recognize specific phenomena, and in particular, lacks the ability to distinguish between the sensitivity and specificity.

In many binary classifications problems the cost of misclassification in one category is higher than the other, and in these applications it is desirable to employ a classifier with selective sensitivity or specificity. For example, it may be of higher relevance for an intensive care provider to determine which patients are most likely to lead to mortality so that a response action can be taken. A classifier used in this situation should strive for the highest accuracy in predicting the cases that will result in death - hits or true positives - and the lowest score for missing

² Sensitivity = $TP / (TP + FN)$; true positives over true positives and false negatives, or total positives. Specificity = $TN / (TN + FP)$; true negatives over true negatives and false positives, or total negatives.

these cases - misses or false negatives. The classifier should thus aim to maximize sensitivity, the number of hits divided by the total number of positive cases. In other applications it may be preferable for a provider to use an automatic classifier that generates a low number of false alarms, and thus maximizes the specificity. In lending importance to one measure, however, the other cannot be neglected; leaving it out of the performance evaluation will cause the model to lose the ability to return meaningful results. Therefore, to highlight one criterion, a multi-objective function must be used which includes both measures and a means to assign weights or goals to each.

Attempts have been made to adopt a multi-objective optimization approach to avoid the ambiguity inherent in using a single objective function. Garcia-Nieto et al. (2009) propose a multi-objective function that includes sensitivity, specificity, and number of features (cardinality). Though they achieve a small improvement in overall PCC, their method does not allow a means to incorporate information about the relative importance of the objectives in their aggregated function, and is thus incapable of fulfilling any desired goal. Kupinski and Anastasio (1999) take a unique approach and suggest presenting the model's solution as a vector-valued function that sweeps out over the entire ROC curve. By this method a classifier can achieve a specific goal when clinical information is applied post-model to select from one of the series of solutions returned by the objective vector. While their approach provides an insightful means to display the relation between sensitivity and specificity, it misses the importance of incorporating an objective into the model's derivation and optimization. In large databases with highly correlated features, input selection is a critical step to both reduce complexity and optimize performance. If information about an objective is incorporated into the selection criteria used for feature selection, this goal becomes part of the optimization task and is reflected in the model's results and feature sets selected. The ability to mold informative feature sets and the benefits of objective-based optimization are lost when knowledge about the objective task is incorporated after the model's derivation.

In this work, a method is proposed for incorporating an adjustable multi-objective function for performance evaluation, comprised of the sensitivity, specificity, and PCC, into the feature selection phase of a model. The flexibility in representing and manipulating the objective function is provided by fuzzy optimization, which allows interpretable rules to be used to weight the relative importance of the multiple criteria. This interpretability and clarity is of great importance in the medical environment, since the tasks of the classifier must be easily understood and validated by physicians. A further benefit of fuzzy multi-optimization is that it allows satisfaction of multiple goals, such that the maximization of one criteria will not be achieved at the expense of the other or the overall accuracy (Vieira et al., 2010). By combining the optimization measure with feature selection, optimization goals are reflected in feature subset selection while feature cardinality is accounted for. Though the classifier may be tuned to satisfy any goal, the sensitivity was emphasized in this work for the purposes of illustrating the proposed method. Though clinical input encouraging optimization around sensitivity was not expressly given, the application of a diagnostic task requiring minimization of false negatives is discussed. To evaluate the multi-objective selector, results are compared to those in Fialho et al. (2010)

and Paetz (2003), using the same publicly available database and clinical predictors.

The paper is organized as follows: Section 2 describes the modeling and feature selection techniques used and introduces the novelty of the multi-objective function for performance evaluation. Section 3 describes results, outlining the simulations undertaken, the fuzzy criteria used in experimentation, and the results themselves. Conclusions are drawn and future work delineated in Section 4.

2. TECHNIQUES

2.1 General

For a detailed description of the modeling techniques, feature selection algorithm, database, and data preprocessing steps, consult Fialho et al. (2010). To summarize, alive/deceased patient classification is performed by both fuzzy and neural network modeling, together with bottom-up wrapper method for feature selection. Bottom-up wrapper feature selection looks for single input features that may influence the output, choosing and combining those that achieve the model with the best performance evaluation criteria. The novelty introduced in this work is that of the evaluation criteria used for the feature selection task: in place of employing the AUC, the most popular construct for evaluating the performance of a diagnostic classifier, and that used in Fialho et al. (2010), a fuzzy multi-objective function is introduced.

2.2 Modeling

This paper uses fuzzy modeling and neural modeling, due to their universal function approximation properties and their ability to represent highly nonlinear problems effectively.

Fuzzy modeling. Fuzzy modeling is a tool that allows an approximation of nonlinear systems when there is little or no previous knowledge of the system to be modeled (Mendonça et al., 2007). Fuzzy models use rules and logical connectives to establish relations between the features defined to derive the model. A fuzzy classifier contains a rule base consisting of a set of fuzzy if-then rules together with a fuzzy inference mechanism.

In this work, Takagi-Sugeno (TS) fuzzy models (Takagi and Sugeno, 1985) are used, which consist of fuzzy rules where each rule describes a local input-output relation. When TS fuzzy systems are used, each discriminant function consists of rules of the type

$$\begin{aligned} \text{Rule } R_i^c : & \text{ If } x_1 \text{ is } A_{i1}^c \text{ and } \dots \text{ and } x_M \text{ is } A_{iM}^c \\ & \text{ then } d_i^c(x) = f_i^c(x), i = 1, 2, \dots, K, \end{aligned} \quad (1)$$

where f_i^c is the consequent function for rule R_i^c . In these rules, the index c indicates that the rule is associated with the output class c . Therefore, the output of each discriminant function $d_c(x)$ can be interpreted as a score (or evidence) for the associated class c given the input feature vector x . The number of rules K , the antecedent fuzzy sets A_{ij} , and the consequent parameters $f_i^c(x)$ are determined using fuzzy clustering in the product space of the input and output variables. The number of fuzzy rules (or clusters) that best suits the data must be determined for classification.

Neural Networks. Artificial Neural Networks (ANN) is a type of computational model formed by an interconnected group of artificial neurons. There are typically three parts in a neural network: an input layer with units representing the input variables, one or more hidden layers, and an output layer with one or more units representing the output variables(s). The units are joined with varying connection strengths or weights. Each connection has an associated weight (synaptic strength) which determines the effect of the incoming input on the activation level of the unit. The weights may be positive (excitatory) or negative (inhibitory).

The neuron output signal is given by the following relationship:

$$\sigma = f(w^T x) = f\left(\sum_{j=1}^n (w_j x_j)\right) \quad (2)$$

where $w = (w_1, \dots, w_n)^T \in \mathbb{R}^n$ is the weight vector, and $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ is the vector of neuron inputs. The function $f(w^T x)$ is often referred to as the activation (or transfer) function. Its domain is the set of activation values, *net*, of the neuron model, and is often represented by $f(\text{net})$. The variable *net* is defined as a scalar product of the weight and input vectors:

$$\text{net} = \sum_{j=1}^n (w_j x_j) = w_1 x_1 + \dots + w_n x_n. \quad (3)$$

Training a neural network can be defined as the process of setting the weights of each connection between units in a way that allows the network to best approximate the underlying function, thus turning it into an optimization problem. In this work, the Levenberg-Marquardt optimization method is used.

2.3 Bottom-up feature selection

A detailed description of the bottom-up approach used here may be encountered in Sugeno and Yasukawa (1993). However, it is important to note that a more recent algorithm that minimizes the computational time with similar performance was proposed in Mendonça et al. (2007). The bottom-up approach looks for single inputs that may influence the output, and combines them in order to achieve the model with the best performance. Two subsets of data are used in this stage, *T* (train) and *V* (validation). Using the train data set, a model is built for each of the *n* features in consideration, and evaluated using the performance criterion described in the next subsection upon the validation data set. The feature that returns the best value of performance is the one selected. Next, other feature candidates are added to the previous best model, one at a time, and evaluated. Again, the combination of features that maximizes the performance criterion is selected. When this second stage finishes, the model has two features. This procedure is repeated until the value of the performance criterion stops increasing. In the end, the subset of all relevant features for the considered process is obtained.

2.4 Fuzzy decision criteria

Fuzzy set theory enables the representation of flexible goals in a way such that the flexibility can be exploited to obtain desired trade-offs to satisfy contradictory goals. The use of fuzzy optimization was therefore introduced to improve the targeted

performance of the model, as it allows for optimization of incommensurable criteria. A detailed description of the general formulation for fuzzy optimization in the presence of flexible goals and constraints can be found in Vieira et al. (2010).

General formulation. Consider a decision making problem where the decision alternatives are $\mathbf{a} \in \Lambda$. A fuzzy goal F_j , $j = 1, 2, \dots, l_1$ is a fuzzy subset of Λ . Its membership function $F_j(\mathbf{a})$, with $F_j : \Lambda \rightarrow [0, 1]$ indicates the degree of satisfaction of the decision goal by the decision alternative $\mathbf{a}$. Similarly, a number of fuzzy constraints G_i , $i = 1, 2, \dots, l_2$ can be defined as fuzzy subsets of Λ . Their membership functions $G_i(\mathbf{a})$, denote the degree of satisfaction of the fuzzy constraint G_i by the decision alternative $\mathbf{a} \in \Lambda$. According to Bellman and Zadeh's fuzzy decision making model Bellman and Zadeh (1970), the fuzzy decision *D* is defined as the confluence of fuzzy goals and constraints, i.e.

$$D(\mathbf{a}) = F_1(\mathbf{a}) \circ \dots \circ F_{l_1}(\mathbf{a}) \circ G_1(\mathbf{a}) \circ \dots \circ G_{l_2}(\mathbf{a}), \quad (4)$$

where $\circ$ denotes an aggregation operator for fuzzy sets. Since the goals and the constraints must be satisfied simultaneously, Bellman and Zadeh proposed to use an intersection operator, i.e. a fuzzy t-norm for the aggregation. The optimal decision alternative $\mathbf{a}^*$ is then the argument that maximizes the fuzzy decision, i.e.

$$\mathbf{a}^* = \arg \max_{\mathbf{a} \in \Lambda} D(\mathbf{a}). \quad (5)$$

Note that both the goals and the constraints are aggregated. Hence, the goals and the constraints are treated equivalently, which is why the model is said to be symmetric.

Performance evaluation using fuzzy criteria. The goal in this fuzzy optimization problem is the simultaneous satisfaction of three different measures: sensitivity, specificity, and PCC. Let C_i , with $i = 1, \dots, l$, be a fuzzy criterion, which can be a goal or a constraint, characterized by its membership function μ_{C_i} , which is a mapping from the space of the criteria C_i to the interval $[0, 1]$. The membership value μ_s for the subset solution *s* is obtained using the aggregation operator $\circ$ to combine the fuzzy criteria

$$\mu_s = \mu_{C_1} \circ \dots \circ \mu_{C_l}. \quad (6)$$

A variety of operators can be used to combine fuzzy objectives. Averaging operators are suitable for modeling compensatory aggregation, while product or parametric triangular norms (t-norms) can control the degree of compensation between different criteria (Vieira et al., 2010). Thus, t-norms must be used to model the conjunctive aggregation desired here. The specific operators considered in this work are the product t-norm, Yager t-norm, and Hamacher t-norm, as given in Vieira et al. (2010).

The weight, or membership value, attributed to each objective in the function will not be equal, and is defined based on its relative importance to the model's task; a membership function (MF) is assigned to each objective that should reflect this importance. Membership values are then achieved by finding the maximization of fulfillment of the objective to its function. Thus, the optimal subset s^* is found by the maximization of μ_s :

$$s^* = \arg \max_s \mu_s. \quad (7)$$

Membership functions can have arbitrary shapes, and in this work trapezoidal, sinusoidal, and exponential functions were

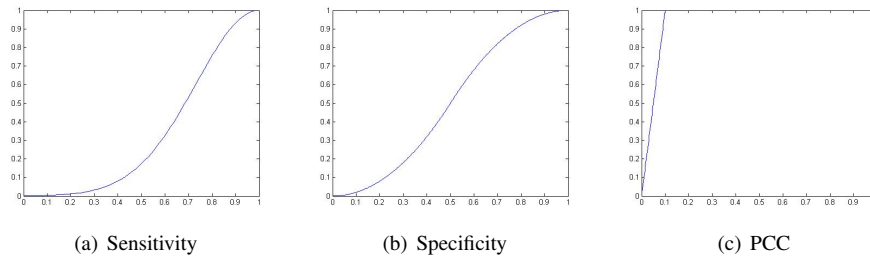


Fig. 1. Membership functions of fuzzy criteria. Y-axis represents membership grade and x-axis represents domain of criteria for each objective – PCC: percentage of correct classifications.

considered. Because the goal inherent to this work is to emphasize the model's sensitivity, the MF assigned for sensitivity to satisfy was the steepest, given by an exponential function. Sinusoidal and trapezoidal membership functions were tried with both specificity and PCC. The membership values for each decision criterion were then combined into a fuzzy objective function using a t-norm operator:

$$D(\mathbf{x}) = C_{sens}(\mathbf{x}) \circ C_{spec}(\mathbf{x}) \circ C_{PCC}(\mathbf{x}) \quad (8)$$

where C_{sens} , C_{spec} and C_{PCC} corresponds to the three different goals: sensitivity, specificity and PCC, respectively. The result is a scalar, which is used as the performance criterion, in place of AUC, to determine which feature would be selected by the bottom-up wrapper method. No defuzzification is necessary after the objective function has been applied.

3. RESULTS AND DISCUSSION

This paper follows the work in Fialho et al. (2010), using the same database and preprocessing techniques. In terms of the simulations, the same two different initial subsets of features were chosen as model inputs. The first, a subset of 12 features found in a total of 121 patients, the same as in Paetz (2003), was used for purposes of comparison. Fialho et al. (2010) considered this subset too narrow, and defined a second subset including 28 variables present in 89 patients. As suggested in Fialho et al. (2010), the percentages of training, test and validation subsets were 60, 30, and 10% of the original dataset, respectively. Bottom up feature selection was applied to each of these subsets using both fuzzy and neural network models. The criteria used to evaluate the model's predictive power were the same as those included in the fuzzy objective function for feature selection: sensitivity, specificity, and PCC. The innovation here is the application of adjustable fuzzy criteria to enable clear manipulation of the three objectives involved. Though the goal of this work was to heighten the sensitivity of the model, through appropriate control of the fuzzy criteria, PCC or specificity may also be manipulated.

3.1 Modeling experimentation and results

Fuzzy criteria. Of the specific operators considered, described in Section 2, the Yager t-norm allowed the best combinations of the three objectives. As contended in Vieira et al. (2010), this may be attributed to the behavior of the intersection of two or more goals satisfying as much as possible their criteria. In this region of interest, the satisfaction relation is nonlinear, allowing an increase in one goal to correspond to a decrease in another goal that is not necessarily proportional.

Table 1. Feature combinations to achieve highest sensitivity and percentage of correct classifications, and number of features selected FM - fuzzy modeling; NN – neural networks; PCC - percentage of correct classifications

Best feature combination		
12 features		
	This work	Fialho et al.
NN		
Features Selected	3 to 6	2 to 8
Sensitivity	[28, 10, 6, 26, 17, 16]	—
PCC	[28, 10, 6, 26, 17, 16]	—
FM		
Features Selected	3 to 6	2 to 6
Sensitivity	[26,28,16,1,7]	[28,26,8, 14]
PCC	[28, 26, 2, 8, 10, 16]	—
28 features		
	This work	Fialho et al.
NN		
Features Selected	3 to 5	4 to 8
Sensitivity	[18, 28, 22, 85, 32]	—
PCC	[18, 28, 22, 85, 32]	—
FM		
Features Selected	3 to 7	2 to 7
Sensitivity	[22, 28, 8, 35, 26, 41, 5]	[22,28,26,41,8,10]
PCC	[22, 28, 26, 29, 35, 85]	—

Various MF types were used with each objective. By an ad hoc method, different combinations of shapes (trapezoidal, sinusoidal, and exponential), domains (representing the score of the objective), spreads, and weights were tried for each membership function. The function chosen for sensitivity was an exponential function, for specificity a sigmoidal function, and for PCC a trapezoidal function, shown in Figures 1a, 1b, and 1c, respectively.

3.2 Specific results

Classification accuracy. Table 2 shows the sensitivity, specificity, and PCC achieved by the neural and fuzzy modeling methods for each initial feature subset, for the methods developed in Paetz (2003), Fialho et al. (2010), and in this work. It is possible to observe from these results that increases in sensitivity were statistically significant in all cases, and are most prominent for those in which the sensitivity was the lowest in Fialho et al. (2010) and Paetz (2003). The greatest differences are observed when using neural models, with a $29.74 \pm 1.4\%$

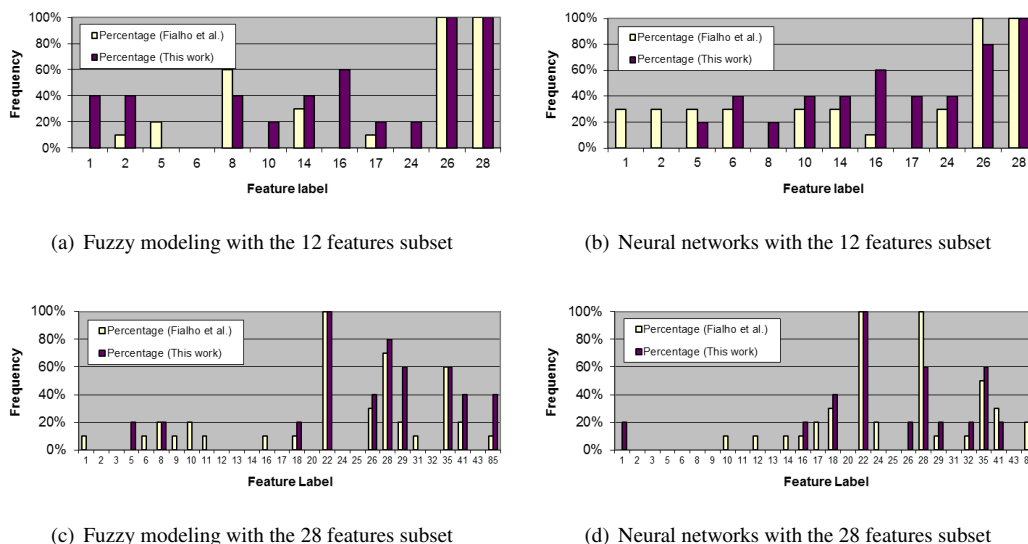


Fig. 2. Histograms with selection rate of each feature.

improvement between this work and those in Fialho et al. when using 12 features, and $23.34 \pm 1.45\%$ when using the 28 feature subset. The increase in sensitivity classification between this work and that in Paetz (2003) is $69.26 \pm 1.4\%$. This steep improvement is particularly notable because the author of Paetz (2003) has attributed low sensitivity scores to the database itself, speculating that the fact that patients who will die are not in a critical state throughout their entire stay at the ICU makes it difficult to predict their outcome. Using fuzzy criteria, we have shown that the low scores for sensitivity cannot be attributed to the database itself, but rather can be adjusted for the algorithm used for modeling.

At the expense of the increase in sensitivity, the models in this work do experience statistically significant decreases in specificity from those found in Fialho et al. (2010), in all cases but one (fuzzy model 12 features). In the mentioned cases for which the sensitivity increased the most, the specificity dropped by $16.12 \pm 1.23\%$ and $16.48 \pm 2.88\%$. This decrease is inherent to the trade-off, which was intended in this work, and is desirable depending on the task intended of the classifier. The difference in PCC, however is much less appreciable, decreasing at most by 2%; using the 12 feature subsets, the differences were not statistically significant.

Features selected. Table 1 presents the feature combination chosen for the best model in terms of sensitivity and PCC, the total number of features selected, and compares with Fialho et al. (2010) when relevant information is available. The combination of features that leads to the highest sensitivity and PCC is the same for neural models but varies for many, though not all, for the fuzzy models. This shows that the neural models are more stable in their feature selection, which could be due to the nature of the FCM clustering method used in the fuzzy models.

Figure 2 show histograms for features selected by the fuzzy and neural models, compared with Fialho et al. (2010), for the 12 and 28 feature subsets. From Figure 2a and 2b, it can be seen that features 26 and 28, corresponding to Calcium and Creatinine lab test results, were chosen by both models in both Fialho et al. (2010) and Paetz (2003) for the 12 feature subset. In Figure 2c and 2d, features 22 and 28, where 22 represents

Table 2. Mean sensitivity, specificity, PCC, and statistical significance, for fuzzy and neural models with 12 and 28 features

12 features						
Sensitivity						
	NN			FM		
	Mean	Std	P-value	Mean	Std	P-value
Paetz	15.01	-	-	-	-	-
Fialho et al.	54.53	5.42	-	79.89	2.60	-
This work	84.27	1.4	<0.0001	84.18	1.24	0.0043
Specificity						
	NN			FM		
	Mean	Std	P-value	Mean	Std	P-value
Paetz	92.26	-	-	-	-	-
Fialho et al.	81.65	3.61	-	71.16	2.86	-
This work	65.53	1.23	<0.0001	68.81	1.94	0.1244
PCC						
	NN			FM		
	Mean	Std	P-value	Mean	Std	P-value
Paetz	69.00	4.37	-	-	-	-
Fialho et al.	73.24	2.03	-	74.10	1.37	-
This work	72.78	1.37	0.6577	74.45	1.68	0.6714
28 features						
Sensitivity						
	NN			FM		
	Mean	Std	P-value	Mean	Std	P-value
Fialho et al.	64.16	3.92	-	77.26	4.40	-
This work	87.5	1.45	<0.0001	88.41	2.62	0.0002
Specificity						
	NN			FM		
	Mean	Std	P-value	Mean	Std	P-value
Fialho et al.	90.33	2.05	-	83.30	2.62	-
This work	73.85	2.88	<0.0001	76.21	1.73	0.0001
PCC						
	NN			FM		
	Mean	Std	P-value	Mean	Std	P-value
Fialho et al.	81.23	1.97	-	82.27	1.56	-
This work	79.13	2.24	0.0851	80.39	1.24	0.0361

Antithrombin III (AT3) level, are chosen by both models and both studies for the 28 feature subsets.

Among subtle differences between features chosen by the different models, the greatest added value from the heightened sensitivity study is to show which of the features selected correspond to this altered task. In other words, the subsets selected in this work show the biological relevance of these features to indicate true positive cases. Feature 16, corresponding to Haemoglobin, stands out for this purpose in Figure 2c) and d), being selected 60% of the time for both models in this work and only 10% for one model for comparison in Fialho et al. (2010). Feature 29, Potassium, stands out in a similar way for the fuzzy model in Figure 2c, though it does not carry the same clout in the neural model using the same feature subset. Even so, further medical expertise revision about the importance of this feature in indicating the association with highlighting true positive cases should be carried out.

General discussion on increasing sensitivity. The goal of the classifier described in this work is to predict the cases that would result in mortality, meaning the sensitivity is maximized. Analyzing a model with high sensitivity can give specific information about which features are most closely related with mortality. If the model exhibits very high sensitivity and low specificity, the task can be seen as serving as a red flag, indicating to medical professionals that chosen patients *might* lead to a state of septic shock and should be monitored more closely, instead of indicating that these patients *will* evolve to that condition. Furthermore, various modeling techniques and professional experience have shown that there is no ultimate set of features used to predict who will die to septic shock. Therefore, until such a set is identified, using a classifier to indicate who is *at high risk*, as opposed to who definitely will die, perhaps approaches the problem more realistically.

4. CONCLUSIONS

This work continued the efforts of Fialho et al. (2010) in applying soft computing methods to a publicly available septic shock database. A method for wrapper feature selection based on a multi-objective function was introduced, combining as objectives the model's sensitivity, specificity, and percentage of correct classifications. This is the first case in which fuzzy criteria are used to impose goals on these measures, and the impact on classification accuracy scores was much greater than what has been presented in previous approaches to multi-objective feature selection, as in Garcia-Nieto et al. (2009). For the purpose of illustrating the modeling task, the sensitivity was chosen to be highlighted. Improvements in sensitivity were notable, achieving up to 30% better performance from that of Fialho et al. (2010). Feature subsets selected by the heightened sensitivity models were different from those found in Fialho et al. (2010), lending insight into biological factors more integrally connected to those patients at risk of death.

This work provides an introductory study and many future analyzes lie ahead. (1) Though the method can be used to achieve any trade-off between the sensitivity and the specificity, results are likely to be of greater significance if one or the other is maximized. A first step in future work will be to highlight the specificity by the same optimization method demonstrated to highlight sensitivity. This will provide a further proof of concept for the method and a basis for comparison with results and features of medical significance. (2) Neural networks give more consistent feature selection subsets; applying other optimization techniques may provide further insight into selected feature

combinations. (3) Biologically: The feature sets selected should be validated by medical professionals and the rules selected by the fuzzy model can be analyzed for their medical significance.

ACKNOWLEDGEMENTS

This work was supported by the MIT Portugal Program, and by Fundação para a Ciência e a Tecnologia, through IDMEC under LAETA, by the project PTDC/SEM-ENR/100063/2008 and by the FCT grants SFRH / 43043 / 2008, SFRH / 43081 / 2008 and SFRH / BPD / 65215 / 2009, Fundação para a Ciência e a Tecnologia, Ministério do Ensino Superior, da Ciência e da Tecnologia, Portugal.

REFERENCES

- Angus, D. and Wax, R. (2001). Epidemiology of sepsis: an update. *Critical Care Medicine*, 7, 109–116.
- Bellman, R.E. and Zadeh, L.A. (1970). Decision-making in a fuzzy environment. *Management Science*, 17(4), 141–164.
- Egan, J. (1975). *Signal Detection Theory and ROC Analysis*. Academic Press, New York.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27, 861–874.
- Fialho, A., Cismondi, F., Vieira, S., Sousa, J., Reti, S., Howell, M., and Finkelstein, S. (2010). Predicting outcomes of septic shock patients using feature selection based on soft computing techniques. In R.K. E. Huellermeier and F. Hoffmann (eds.), *Applications 13th International Conference, IPMU 2010, Dortmund, Germany, Proceedings, Part II*, volume 81 of *Communications in Computer and Information Science (CCIS)*, 65–74. Springer-Verlag Berlin Heidelberg.
- Garcia-Nieto, J., Alba, E., Jourdan, L., and Talbi, E. (2009). Sensitivity and specificity based multiobjective approach for feature selection: Application to cancer diagnosis. *Information Processing Letters*, 109, 887–896.
- Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., and Dougherty, E. (2010). Small-sample precision of roc-related estimates. *Bioinformatics*, 6, 822–830.
- Hand, D. (2009). Measuring classifier performance: A coherent alternative to the area under the roc curve. *Machine Learning*, 77, 103–123.
- Hanley, J. and McNeil, B. (1982). The meaning and use of the area under a roc curve. *Radiology*, 143, 29.
- Kupinski, M. and Anastasio, M. (1999). Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves. *IEEE Transaction on Medical Imaging*, 8, 675–685.
- Lee, W. and Hsiao, C. (1996). Alternative summary indices for the roc curve. *Epidemiology*, 7, 605–611.
- Lobo, J., Jimnez-Valverde, A., and Real, R. (2008). Auc: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 145–151.
- Mendonça, L., Vieira, S., and Sousa, J. (2007). Decision tree search methods in fuzzy modeling and classification. *International Journal of Approximate Reasoning*, 44, 106–123.
- Muller, B., Becker, K., Schachinger, H., Rickenbacher, P., Huber, P., and Zimmerli, R. (2000). Calcitonin precursors are reliable markers of sepsis in a medical intensive care unit. *Critical Care Medicine*, 4, 977–983.
- Paetz, J. (2003). Knowledge-based approach to septic shock patient data using a neural network with trapezoidal activation functions. *Artificial Intelligence in Medicine*, 28, 207–230.
- SCCM (1992). Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Critical Care Medicine*, 20, 864–874.
- Sugeno, M. and Yasukawa, T. (1993). A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1, 7–31.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285.
- Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15, 116–132.
- Vieira, S., Sousa, J., and Kaymak, U. (2010). Fuzzy criteria for feature selection. *Preprint submitted to Elsevier*.