WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia FUZZ IEEE

Probabilistic Fuzzy Prediction of Mortality in Intensive Care Units

A. S. Fialho*, U. Kaymak†, R. J. Almeida‡, F. Cismondi*, S. M. Vieira§, S. R. Reti¶,
J. M. C. Sousa§, S. N. Finkelstein*
*Massachusetts Institute of Technology, Engineering Systems Division
77 Massachusetts Avenue, 02139 Cambridge, MA, USA, Email: afialho@mit.edu
†Eindhoven University of Technology, School of Industrial Engineering
PO Box 513, 5600 MB Eindhoven, The Netherlands
‡Erasmus University of Rotterdam, Econometric Institute
PO Box 1738, 3000 DR Rotterdam, The Netherlands
§Technical University of Lisbon, Instituto Superior Técnico, Dept. of Mechanical Engineering, CIS/IDMEC – LAETA
Av. Rovisco Pais, 1049-001 Lisbon, Portugal
¶Division of Clinical Informatics, Department of Medicine
Beth Israel Deaconess Medical Centre, Harvard Medical School, Boston, MA, USA

***Abstract*—In the present work, we propose the application of probabilistic fuzzy systems (PFS) to model the prediction of mortality in septic shock patients. This technique is characterized by the combination of the linguistic description of the system with the statistical properties of data. Preliminary results for this particular clinical problem point that PFS models, besides performing as accurately as first order Takagi-Sugeno fuzzy models, also provide probability measures that provide additional clinical information upon which physicians can act on.**

I. Introduction

Prediction tools, such as logistic regression, neural networks or fuzzy systems, have been employed in intensive care units (ICU) of hospitals since the 1980s to assist clinicians in the process of decision-making [1].

Fuzzy systems can be described as a tool that significantly differs from logistic regression and neural networks by allowing to develop models of human reasoning (also referred to as approximate reasoning) and by dealing with uncertainty within the data [2], [3]. This approach is particularly appealing as it provides not only transparent, non-crisp models, but also a linguistic interpretation in the form of rules and logical connectedness. Ultimately, these models classify each instance of a dataset as pertaining with a certain degree, to one of the possible classes defined for the specific problem being modeled [3]. However, these output scores cannot usually be directly translated into the real probability of an outcome. Contrarily, probabilistic fuzzy systems (PFS) may help to model a probabilistic outcome while keeping linguistic interpretability [4]. For the process of decision-making of physicians this is of particular interest as it provides statistical additional information and support relative to the rules described.

In this way, the present work aims to explore the use of probabilistic fuzzy systems to the clinical problem of mortality prediction in septic shock patients. These models are an extension of fuzzy systems, in which the deterministic fuzzy rules are replaced with probabilistic fuzzy rules. Interpretable fuzzy rules are initially determined to deal with the uncertainty associated with the linguistic terms used in the model. These rules increase the transparency of the learned system for humans and provide an additional mean to validate the fuzzy classifier by experts' knowledge regarding the system. The determination of fuzzy rules is then followed by the estimation of the probabilities of observing a class label [5], [4].

The outline of the paper is as follows. In Section II, we briefly introduce the clinical problem of mortality prediction in septic shock patients. The principles behind PFS are discussed in Section III. In Section IV we present both implementation considerations and results from an empirical study involving the application of PFS to the clinical problem described. Discussion is done in section V and conclusions are drawn in Section VI.

II. Septic Shock

According to [6], sepsis is defined as the state where there is a systemic inflammatory response syndrome (SIRS) resultant of a confirmed infectious process. This SIRS is a clinical condition defined by the presence of at least two of the following findings: low ($< 36\,^{\circ}\mathrm{C}$) or high ($> 38\,^{\circ}\mathrm{C}$) body temperature (hypothermia or fever, respectively), high heart rate (> 90 beats per minute), high respiratory rate (> 20 breaths per minute), and low ($< 4,000$ cells/mm3) or high ($> 12,000$ cells/mm3) white blood cell count. Ultimately, a patient is considered to be in septic shock when the hypotensive state related to a sepsis condition persists, despite adequate fluid resuscitation [6].

From the clinical point of view, being able to model this clinical problem of mortality prediction of septic shock patients has great potential for: (i) refining clinical trials among this well-defined class of patients; (ii) to evaluate the performance of individual ICUs in treating septic shock patients [7]; and (iii) to plan the remaining lifetime of these highly critical patients.

Previous works have explored this problem of modeling mortality in septic shock patients. In [8] the authors compared logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. Ribas et al. used factor analysis for feature selection combined with logistic regression over the resulting latent factors [9]. In [10], a neural network model was developed to predict mortality of abdominal septic shock patients. However, the use of the two types of data mining techniques mentioned may present a few drawbacks such as difficulty to fit highly non-linear data (logistic regression) or lack of interpretability (neural networks) [11]. Thus, in the present work an hybrid technique that incorporates probability in fuzzy logic is used.

III. Probabilistic Fuzzy Systems

In this section, we briefly overview the principles behind fuzzy rule based classification and introduce the the main concepts behind probabilistic fuzzy systems. Namely, the methods for estimating the antecedent and probabilistic parameters of PFS are described.

A. Fuzzy Systems

Fuzzy rule based classification consists of a classification system based on a set of fuzzy if-then rules together with a fuzzy inference mechanism [12]. Three general methods to design a fuzzy classifier can be distinguished: (i) regression method, (ii) discriminant method, and (iii) maximum compatibility method [12], [13]. For the present work, given that the two output classes (patient will die/patient will not die) will be characterized differently, the second fuzzy classifier method is used.

In the discriminant method, a discriminant function $d_c(x)$ is associated with each class $c = 1, 2$ (binary problems), and their combination allows the implementation of fuzzy inference systems. When Takagi-Sugeno fuzzy systems are used, each discriminant function consists of rules of the type

$$\begin{aligned}\text{Rule } R_j^c : \text{If } x_1 \text{ is } A_{j1} \text{ and ... and } x_N \text{ is } A_{jN} \\ \text{then } d_c(x) = f_j^c \text{ , } j = 1, 2, ..., J\end{aligned} \quad (1)$$

where N corresponds to number of inputs, J to the number of rules, A_{jn} is the fuzzy set of the j^{th} rule and input n, f_j is the consequent function for rule R_j and the index c indicates that the rule is associated with the output class c. Typically, the antecedent parts of the rules are the same for different discriminants, but the consequents may be different. 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** [12]. The classifier assigns the class label corresponding to the maximum value of the discriminant functions, i.e

$$\max_c \; d_c(x). \quad (2)$$

If the discriminant function is the probability that an input feature vector is assigned to class w_c, then (2) reduces to the probabilistic decision function

$$\max_c \; \Pr(w_c|x). \quad (3)$$

The structure of a discriminant classifier is depicted in Fig. 1 [14]. Note that the classification is based on the largest discriminant function $d_c(x)$ regardless of the values or definitions of other discriminant functions.

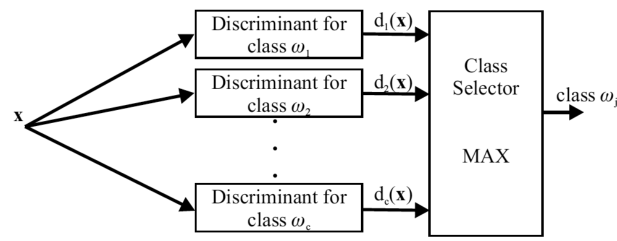

Fig. 1. Classifier design by discriminant method.

B. Probabilistic Fuzzy Systems

A *probabilistic fuzzy system* consists of a set of rules whose antecedents are fuzzy conditions similar to the ones previously described, and whose consequents are probability distributions.

Considering a binary problem of determining the class y $\in \{C_1,C_2\}$ to which a data point $\mathbf{x} = (x_1, ..., x_N)$ belongs, the binary probabilistic fuzzy classifier will have the following form [12]:

$$\begin{aligned}\text{Rule } R_j : \text{If } \mathbf{x} \text{ is } A_j \text{ then } y = C_1 \text{ with probability } p_{j1} \text{ and} \\ y = C_2 \text{ with probability } p_{j2}\end{aligned} \quad (4)$$

In the present work, C_1 and C_2 correspond to the binary classes 0 and 1, which relate respectively to patients surviving and patients dying. A similar reasoning can be extended to a largest number of C classes.

The antecedent fuzzy sets A_j $(j = 1, ..., J)$ in (4) are defined in the n-dimensional input space X. For each fuzzy set there is a corresponding probabilistic fuzzy rule. Furthermore, the probability parameters p_{jc} in equation 4 satisfy

$$p_{jc} \geq 0, \text{ for } j = 1, ..., J \text{ and } c = 1, 2 \quad (5)$$

and

$$\sum_{c=1}^{2} p_{jc} = 1 \text{ , for } j = 1, ..., J. \quad (6)$$

Let $\mu_{A_j}(\mathbf{x})$ denote the membership function of a fuzzy set A_j. A probabilistic fuzzy classifier with rules given by (4) provides an estimate of $Pr(y|\mathbf{x})$, the conditional probability distribution of y given **x**. The estimate $\hat{p}(C_c|x)$ of a conditional probability $Pr(C_c|\mathbf{x})$ is obtained by

$$\hat{p}(C_k|\mathbf{x}) = \sum_{j=1}^{J} \bar{\mu}_{A_j}(\mathbf{x}) p_{jc} \quad (7)$$

where $\bar{\mu}_{A_j}(\mathbf{x})$ denotes a normalized membership function given by

$$\bar{\mu}_{A_j}(\mathbf{x}) = \frac{\mu_{A_j}(\mathbf{x})}{\sum_{j=1}^{a} \mu_{A_j}(\mathbf{x})}. \tag{8}$$

Given $\mathbf{x}$, the normalized membership functions $\bar{\mu}_{A_j}(\mathbf{x})$ determine the degrees of fulfilment of the probabilistic fuzzy rules. This is similar to the reasoning mechanism in Takagi-Sugeno fuzzy systems.

The estimated conditional probability distribution $\hat{p}(y|\mathbf{x})$ can be used for classifying a data point $\mathbf{x}$. The following classification rule minimizes the probability of misclassification

$$\hat{y} = \underset{y \in \{C_1, C_2\}}{\text{argmax}} \; \hat{p}(y|\mathbf{x}). \tag{9}$$

In this way, in a probabilistic fuzzy classifier, one has to determine (i) the parameters and type of the antecedent membership functions and (ii) the probability parameters of the rule consequents. The first probabilistic fuzzy classifiers used a sequential method for determining the parameters in a probabilistic fuzzy classifier [15], [4], [16]. A more recent approach was developed in [17], where the parameters are simultaneously estimated using maximum likelihood. Some studies also considered the use of rule weights which corresponds to the probabilities under certain conditions [18]. The sequential method for parameter estimation is now briefly overviewed, followed by a more detailed explanation of the maximum likelihood parameter estimation.

1) Sequential Method For Parameter Estimation: The principle behind the sequential method is that the set of parameters for the antecedent membership functions are estimated first, using for instance, unsupervised learning methods or expert knowledge. Then, assuming that these first parameters are constant, the probability parameters for the rule consequents can be estimated based on probability measures of fuzzy events [19].

In the present work, we consider each antecedent membership function $\mu_{A_j}(\mathbf{x})$ as the product of n univariate Gaussian membership functions, one for each dimension of the input space X. We then obtain

$$\phi = \exp\left(- \sum_{n=1}^{N} \frac{(x_n - v_{jn})^2}{\sigma_{jn}^2} \right). \tag{10}$$

In this way, the parameters that need to be estimated for the antecedent membership functions are given by a vector $\mathbf{v}_j = \{v_{j1}, ..., v_{jN}\}$ and a vector $\mathbf{\Sigma}_j = \{\sigma_{j1}, ..., \sigma_{jN}\}$. These vectors indicate, respectively, the center and the width of the membership function in each dimension of the input space. Unsupervised methods such as clustering may be used for determining $\mathbf{v}_j$ and $\mathbf{\Sigma}_j$. Following [20], a method based on fuzzy c-means (FCM) clustering was considered for the estimation of the parameters for the antecedent membership functions. First, the data set available for parameter estimation should be normalized. For the n^{th} feature $(n = 1, ..., N)$ of a data point $\mathbf{x}_i (i = 1, ..., I)$, this is done according to

$$\bar{x}_{il} = \frac{x_{in} - \mu_n}{\sigma_n} \tag{11}$$

where μ_n and σ_n denote, respectively, the mean and the standard deviation of the n^{th} feature over the entire data set. The FCM algorithm is then applied to the normalized data points $\bar{x}_n$ in order to identify a predefined number of cluster centers. The FCM algorithm uses the standard Euclidean distance measure. The cluster centers obtained using FCM serve as the centers v_j of the Gaussian membership functions $\bar{\mu}_{A_j}(\mathbf{x})$. The vectors $\mathbf{\Sigma}_j$, which contain the widths of the membership functions, then remain to be estimated. In this work, we used the nearest neighbour heuristic for estimating these vectors, i.e.

$$\sigma_{jn} = \min \left\| v_j - v_j' \right\| \text{ , for } (n = 1, ..., N) \tag{12}$$

where $\left\| \mathbf{v}_j - \mathbf{v}_j' \right\|$ denotes the Euclidean distance between $\mathbf{v}_j$ and $\mathbf{v}_j'$. Note that a membership function is given the same width in each dimension according to this heuristic. After determining the parameters $\mathbf{v}_j$ and $\mathbf{\Sigma}_j$, the parameters p_{jc} have to be determined in such a way that the constraints in (5) and (6) are satisfied. In [5], [4], [17], it has been proposed to set these parameters equal to the estimates of the conditional probabilities $Pr(C_c|A_j)$. According to [21] this results in

$$p_{jc} = \frac{\sum_{i=1}^{I} \bar{\mu}_{A_j}(\mathbf{x}_i) \chi_{C_c}(y_i)}{\sum_{i=1}^{I} \bar{\mu}_{A_j}(\mathbf{x}_i)}, \tag{13}$$

where the characteristic function χ_{C_c} equals 1 if $y = C_c$ and equals 0 otherwise. The use of (13) for estimating the parameters p_{jc} is commonly referred to as the *conditional probability estimation*.

2) Maximum Likelihood For Parameter Estimation: The sequential method just described does not usually provide an optimal estimate of the antecedent and probability parameters [17]. Some of the reasons include: (i) unsupervised learning of the antecedent parameters does not take class labels into account, which may decrease the performance of the classifier; and (ii) equation (13) generally does not provide maximum likelihood (ML) estimates of the probability parameters in a probabilistic fuzzy classifier, i.e. (13) does not maximize the probability of observing the data set available for parameter estimation. As proposed in [17], a ML method can be used to estimate both the antecedent parameters and the probability parameters for a probabilistic fuzzy classifier. This method has the advantage that the likelihood of the data set available for parameter estimation is maximized and that all parameters are estimated simultaneously to satisfy the identification objective.

Given that,

$$\hat{p}(y|\mathbf{x}) = \sum_{c=1}^{2} p(y|C_c) \sum_{j=1}^{J} \bar{\mu}_{A_j}(\mathbf{x}) p_{j,c}, \tag{14}$$

the likelihood of a data set is given by

$$L = \prod_{i=1}^{I} \hat{p}(y_i|x_i). \quad (15)$$

Maximization of the likelihood is equivalent to minimization of the negative log-likelihood. Therefore the following error function is minimized:

$$E = -\sum_{i=1}^{I} \ln \hat{p}(y_i|x_i). \quad (16)$$

For minimizing the error function in (16), a gradient descent optimization algorithm is used. The stochastic variant of gradient descent is applied, which means that the available training classification examples are processed one by one and that updates are performed after each example.

Finding the parameters $\mathbf{v}_j$, $\mathbf{\Sigma}_j$, and p_{jc} that minimize the error function in (16) is a constrained optimization problem, since the probability parameters p_{jc} must satisfy the conditions in (5) and (6). By using the auxiliary variables $u_{jc}(j = 1, ..., J$ and $c = 1, ..., C)$, it is possible to convert this constrained optimization problem into an unconstrained optimization problem. The relation between these variables and the probability parameters p_{jc} is described by the *softmax* function, i.e.

$$p_{jc} = \frac{e^{u_{jc}}}{\sum_{c=1}^{2} e^{u_{jc}}}. \quad (17)$$

ML estimates of the parameters v_j, Σ_j, and p_{jc} can then be obtained by the unconstrained minimization of (16) with respect to $\mathbf{v}_j$, $\mathbf{\Sigma}_j$ and u_{jc}.

The approach developed in [17] and used in the present work starts by finding the initial values of the parameters using the sequential parameter estimation method, and then applies the maximum likelihood procedure for their optimization. An outline of this method is given in the Algorithm 1.

Algorithm 1 Maximum likelihood method to determine parameters $\mathbf{v}_j$, $\mathbf{\Sigma}_j$, and p_{jc}

1) Initialize $\mathbf{v}_j$, $\mathbf{\Sigma}_j$, and p_{jc} :
FCM clustering to find $\mathbf{v}_j$ and $\mathbf{\Sigma}_j$
Determine $p_{jc} = \frac{\sum_{i=1}^{I} \bar{\mu}_{A_j}(x_i)\chi_{C_c}(y_i)}{\sum_{i=1}^{I} \bar{\mu}_{A_j}(x_i)}$

2) Maximum Likelihood:
for $iter = 1 \rightarrow$ no. of iterations **do**
  for $i = 1 \rightarrow$ no. of training classification examples **do**
    Simultaneous re-estimation of the parameters $\mathbf{v}_j$, $\mathbf{\Sigma}_j$ and p_{jc} through minimization of error function in (16)
  end for
  Determine $E(iter) = -\sum_{i=1}^{I} \ln \hat{p}(y_i|x_i)$
end for

IV. Mortality Prediction

The present work aimed at using probabilistic fuzzy systems to model the prediction of mortality in septic shock patients. Namely, these models were used to assess the probability of a patient to decease following 72 hours of a given time point. Simultaneously, in order to assess differences in performance, first-order Takagi-Sugeno fuzzy models (FM) were applied to the same dataset. The correspondent dataset, preprocessing steps and models setup are now described. The respective results are then presented.

A. Dataset

This retrospective cohort study used data from the MEDAN database [22]. This is a database of intensive care unit (ICU) abdominal septic shock patients admitted to 70 different hospitals in Germany, collected from 1998 to 2002. All information was de-identified by removal of all protected health information. This MEDAN database is formed by 583 patients and 103 different variables including personal records, physiological parameters, procedures, diagnosis/therapies, medications, and respective outcomes (survived or deceased).

In the present work, we performed our experiments in a subgroup of 131 patients that met the criteria of containing data relatively to the following physiological parameters: systolic blood pressure, platelets, spO2, white blood cell count, arterial pO2 and PTT. Systolic blood pressure relates to the amount of pressure that blood exerts on vessels while the heart is beating; platelets are a type of blood cells responsible, along with other substances, to form clots to stop bleeding; spO2 refers to the amount of oxygen bound to red blood cells in bloodstream; white blood cells are a type of cell in the blood involved in defending the body against both infectious disease and foreign materials; arterial pO2 (partial pressure of oxygen) relates to the amount of dissolved oxygen in the blood; and PTT corresponds to the amount of time it takes for blood to clot; i.e. it is a test that helps in determining whether there are bleeding or clotting problems. This choice of parameters derived from a previous process of feature selection detailed in [23].

B. Data Preprocessment

In order to improve the quality of the data, a few preprocessing steps were performed using the complete dataset (with all 103 variables). Given that all variables were collected with different sampling periods, a template variable was used to align them. This process allows having all samples of all variables aligned at the same point in time as the template variable. We chose to use heart rate as the template, once it is the most frequently measured variable (in average one sample every 60 minutes) and thus, the one introducing fewer artefacts in the data [24]. Regarding missing data, the last available value was used to impute values in recoverable missing segments. According to [24], [25] a recoverable missing segment is the one where a variable was not measured during a certain period of time because of an intentional reason.

C. Model Setup

With regards to the experimental conditions, ten simulations were performed. For each of the simulations, the dataset was randomly divided into two equally sized subsets: one for training and one for testing. The two models (fuzzy systems and probabilistic fuzzy systems) were first trained and tested upon these subsets. Then, still part of the same simulation, the training subset became the test subset and vice-versa, and models were again trained and tested. Thus, a total of twenty runs were performed (ten simulations times two runs).

In order to model PFS, several types of parameters need to be identified. The first kind of parameters relate to the clustering process which, as previously pointed, was based on fuzzy c-means clustering (FCM) [20]. The two clustering parameters to be defined include the number of clusters (which translates into the number of PFS rules) and the degree of fuzziness of the clustering (weighting exponent of the clustering algorithm). The first was determined by testing different number of clusters and simultaneous estimation of the partition index (SC) [26]. This index allows comparing different partitions with the same number of elements. The lower this value the better is the partition. A local optimum number of six clusters was found. Based on literature, the weighting exponent of the FCM algorithm, which determines the degree of fuzziness of the clustering, was given a value of 2 [27]. The second kind of parameter relates to the width of the membership functions. For all experiments the constraint $\sigma_{nj} \geq 0.10$ with, $(j = 1, ..., J)$ and $(n = 1, ..., N)$, was imposed. This was done to avoid numerical instability. Finally, the third type of parameters to be identified relate to the gradient descent optimization used during the maximum likelihood method. Different combinations were tested for the number of iterations and learning rate. A final combination of 50 iterations with a learning rate of 0.03 was sufficient to guarantee the error function (16) to converge.

For fuzzy systems, the same partition index *SC* was used to find the optimum number of clusters and the clustering was also based on fuzzy c-means algorithm (FCM) [20].

D. Model Assessment

Since this work consists of classifying a set of instances to one of two possible classes (patient will die or patient will not die), the area under the receiver-operating characteristic (ROC) curve (AUC) can be used to assess its discrimination performance [28]. This is a function of the true positive rate (proportion of correctly classified positive cases) versus the false positive rate (proportion of incorrectly classified positive cases), integrated over all thresholds. The true positive rate and true negative rate correspond, respectively, to the sensitivity and specificity of the model and, in this particular problem, represent the cases where the patient was correctly classified as dying and the cases where the patient was correctly classified as surviving. The accuracy of the models was also assessed by determining the proportion of the total number of predictions that were correct.

E. Results

As previously mentioned, instead of using a convergence criteria for the gradient descent method, we preferred to fix the number of iterations to a number that guaranteed the convergence of the error function (16). As depicted in Fig. 2 it is possible to observe that this function converges for less than 50 iterations. Thus, using a a limit of 50 iterations guaranteed the gradient descent method to converge.

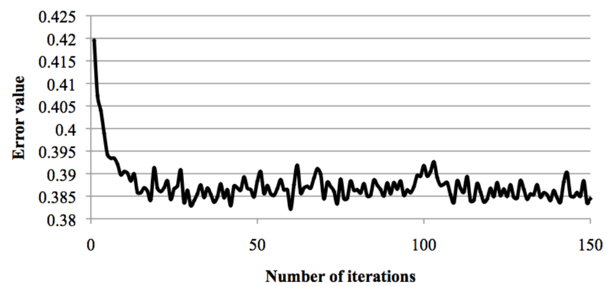

Fig. 2. Plot of the error function (16) along the gradient descent method.

Results obtained after training PFS and FM over a training data subset and running them over a test data subset are shown by the ROC in Fig. 3, and by the respective confusion matrices in Tables I and II. Note that both the plots and confusion matrices presented here refer to one of the 20 simulations performed (differences in the plots and confusion matrices obtained for the remaining runs were not significant). Results averaged over all 20 simulations are presented in Table III. Generally, models returned good values of AUC (good discrimination), specificity, sensitivity and accuracy. No significant differences were found between the results obtained by PFS and by FM models.

TABLE I
CONFUSION MATRIX FOR PFS MODEL.

	PFS	
	Predicted 0	Predicted 1
Real 0	0.72	0.06
Real 1	0.02	0.20

TABLE II
CONFUSION MATRIX FOR FM MODEL.

	FM	
	Predicted 0	Predicted 1
Real 0	0.73	0.05
Real 1	0.03	0.19

Parameters v_j and σ_j obtained for the membership functions of the PFS model obtained are presented in Table IV. The plots of the respective membership functions are presented in Fig. 4. Finally, the probability parameters p_{jc} associated to each PFS rule are presented in Table V. Given the three types of parameters obtained, an example of a rule is given by:

TABLE III
RESULTS OBTAINED AFTER RUNNING THE PROBABILISTIC FUZZY SYSTEMS (PFS) MODELS AND FUZZY MODELS (FM) OVER THE TEST DATA SUBSET.

	PFS	**FM**
AUC	0.80 ± 0.02	0.81 ± 0.03
Specificity	0.79 ± 0.03	0.79 ± 0.02
Sensitivity	0.81 ± 0.04	0.82 ± 0.03
Accuracy	0.78 ± 0.03	0.80 ± 0.02

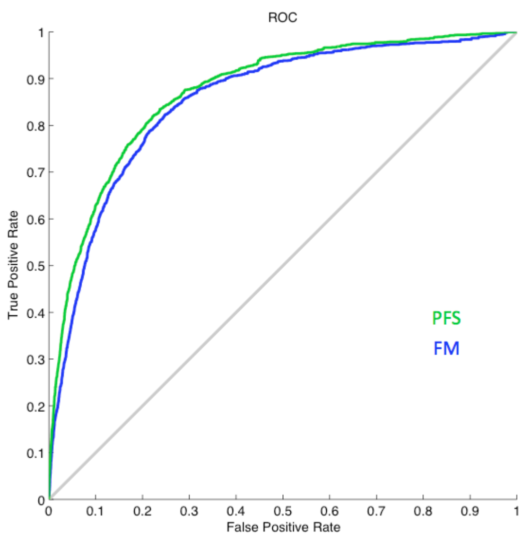

Fig. 3. Receiving operator curve (ROC) for PFS and FM models.

$$\begin{aligned} R_1\text{: } &\textbf{If } x_1 \text{ is } A_{11} \text{ and } x_2 \text{ is } A_{21} \text{ and } x_3 \text{ is } A_{31} \\ &\text{and } x_4 \text{ is } A_{41} \text{ and } x_5 \text{ is } A_{51} \text{ and } x_6 \text{ is } A_{61} \\ &\textbf{then } y = 1 \text{ (patient will die)} \\ &\text{with a probability of } 90\% \end{aligned} \tag{18}$$

Note that all membership functions A_{jn} for Rule R_1, i.e. all A_{1n} are depicted in blue in Fig. 4.

Figure 5 depicts the histogram of probability parameters for class 1 (probability of the patient to die after 72 hours) using the testing samples. This histogram was divided into three regions. Region number 1 and 3 correspond to the areas where the probability of a patient to die is either very low or very high. Region 2 corresponds to the uncertain area where the probability for a patient to die is between 0.4 and 0.6. The respective confusion matrices (in proportions) for each of the three regions is presented in Tables VI, VII and VIII.

V. DISCUSSION

For this specific problem of mortality prediction of septic shock patients, PFS and FM models returned, in general, good results of discrimination, specificity, sensitivity and accuracy. The difference in the results between the two types of models was not statistically significant and thus, both models are comparable from the performance point of view.

TABLE IV
PARAMETERS v_j AND σ_j FOR EACH RULE.

Variable	**Rule Number**	v_j	Σ_j
Systolic blood pressure	Rule 1	0	0.57
	Rule 2	-0.16	0.47
	Rule 3	1.07	0.94
	Rule 4	0.55	0.85
	Rule 5	0.59	0.74
	Rule 6	0.34	0.50
Platelets	Rule 1	0.31	0.21
	Rule 2	-0.40	0.15
	Rule 3	0.47	0.10
	Rule 4	-0.07	0.11
	Rule 5	0.85	0.26
	Rule 6	0.30	0.11
FiO2	Rule 1	0.77	0.28
	Rule 2	0.46	0.35
	Rule 3	0.03	0.71
	Rule 4	0	0.36
	Rule 5	1.06	0.32
	Rule 6	-0.14	0.29
White blood cell	Rule 1	0.78	0.18
	Rule 2	0.41	0.30
	Rule 3	0.12	0.18
	Rule 4	0.05	0.34
	Rule 5	1.00	0.29
	Rule 6	-0.26	0.52
Arterial pO2	Rule 1	0.64	0.21
	Rule 2	0.22	0.33
	Rule 3	-0.40	0.10
	Rule 4	0.08	0.34
	Rule 5	0.99	0.34
	Rule 6	0	0.29
PTT	Rule 1	0.87	0.36
	Rule 2	0.48	0.39
	Rule 3	0.02	0.65
	Rule 4	0.04	0.43
	Rule 5	1.07	0.35
	Rule 6	-0.27	0.59

TABLE V
PROBABILITY PARAMETERS p_{jc} FOR EACH RULE.

Rule number	**Class 0**	**Class 1**
1	0.10	0.90
2	0.61	0.39
3	0.95	0.05
4	0.96	0.04
5	0,78	0.22
6	0.94	0.06

From the analysis of PFS confusion matrices several points can be discussed. First, it is clear that a higher proportion of misclassifications is obtained in region 2. This finding is in accordance to what was expected as this region represents the samples with probability parameters laying in between 0.4 and 0.6, i.e. in the area of higher uncertainty. From the point of view of medical decision making, it is preferable to have values of probability outside this interval, carrying higher certainty power for any of the outcomes. Second, it is relevant to note that the proportion of incorrect diagnosis with high probabilities attached is low (region 1 and 3). The underlying idea is that an incorrect diagnosis with a probability of 0.6 attached to it is less meaningful than an incorrect diagnosis

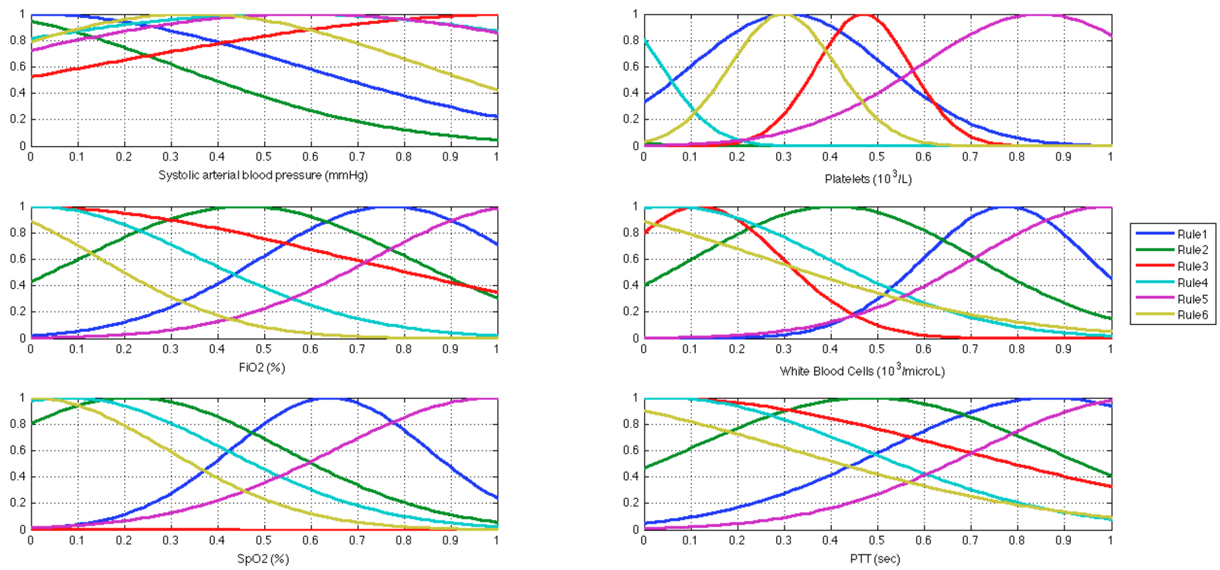

Fig. 4. Plot of the membership functions.

TABLE VI
CONFUSION MATRIX FOR REGION 1.

Region 1	Predicted 0	Predicted 1
Real 0	0.99	0
Real 1	0.01	0

TABLE VII
CONFUSION MATRIX FOR REGION 2.

Region 2	Predicted 0	Predicted 1
Real 0	0.54	0.38
Real 1	0.02	0.06

TABLE VIII
CONFUSION MATRIX FOR REGION 3.

Region 3	Predicted 0	Predicted 1
Real 0	0	0.28
Real 1	0	0.72

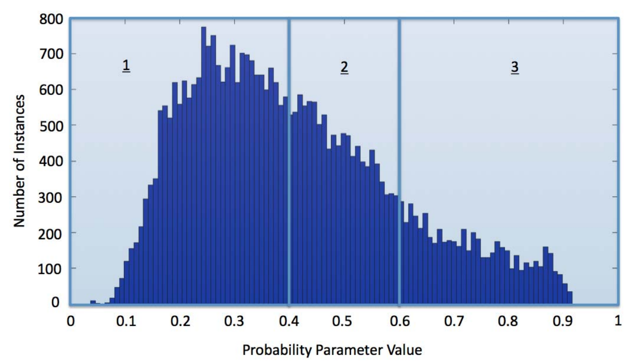

Fig. 5. Histogram of probability parameters for class 1 (probability of the patient to die after 72 hours) using the testing samples.

with a probability of 0.9 attached to it and, in the same way, a correct diagnosis is more valuable when with a probability of 0.9 than with a probability of 0.6. Finally, it is also important to note that in any of the three regions, there are very few cases of real 1s, predicted as 0s. This means that the models rarely classified a patient as surviving when it actually died. From the medical point of view, a misclassification of this type is the one with the more severe consequences and thus, the one models should primarily avoid.

As shown in the present work, PFS and FM models can be used in clinical modeling problems, as they are able to deal with the linguistic uncertainty by providing rules for clinicians. However, PFS have a fundamental difference with ordinary fuzzy systems where, given some input, an ordinary fuzzy system returns a single output value. Contrarily, PFS also have the capacity to deal with the probabilistic uncertainty by returning a probability distribution over the possible output values. In general, determining these probabilities is of particular interest for healthcare providers as it provides statistically interpretable additional information and support relative to the rules described. These probability estimates can assist clinicians in making decisions in four distinct ways: (i) by providing greater certainty of the expected effects of treatment, (ii) by improving understanding of specific prognostic elements and their relative influence on outcomes, (iii) by reducing reliance on commonly used clinical rules that may be biased, and (iv) by providing an explicit opportunity to review and compare explicit probability thresholds for important clinical decisions [29].

In the present work of mortality prediction, the obtained probability parameters also point to the valuable possibility of stratifying patients according to their risk of death. According to Table V, a patient whose physiological conditions falls under the conditions of rule number 1 will be in high risk of death in 72 hours (90% probability). Contrarily a patient in a condition described by rules 3, 4 or 6 will have a very low risk of death (95% probability). Together with physicians' input about the rules obtained (i.e. the interpretation of the membership functions and probability parameters), this stratification of patients according to their mortality risk might be particularly useful for resource management and planning of clinical trials of septic shock patients in ICUs. This can also be seen as a first step towards personalized medicine according to patient's mortality risk.

VI. Conclusion

In the present work PFS models were used to predict mortality among septic shock patients, after 72 hours of a given time point. Probabilistic fuzzy logic is a new approach for incorporating probability in fuzzy logic in order to better represent non-deterministic real world systems. This technique involves two types of reasoning: one about the degree of truth of a proposition, and the other about the probability of truth, in a combination that can provide 'the best of the two worlds'.

Obtained results for this particular clinical problem point that PFS models not only perform comparably to first order Takagi-Sugeno fuzzy models, but also that they can provide additional information regarding the specific mortality probability of patients. This additional information upon which physicians can act on has the potential to make PFS models more useful within the clinical setting.

Future work will involve the further exploration and comparison of PFS with the currently established methods in the medical framework (e.g. logistic regression), and the merge of these preliminary results with experts' knowledge (e.g. physicians) to validate the obtained rules and probabilities.

Acknowledgment

The authors would like to acknowledge the help and space provided by the Division of Clinical Informatics of the Beth Israel Deaconess Medical Center, the Massachusetts Institute of Technology and Eindhoven University of Technology. Both human and technical resources were available through them, and were critical for the development of this work. This work is supported by the Portuguese Government under the programs: project PTDC/SEMENR/100063/2008, Fundação para a Ciência e Tecnologia (FCT), and by the MIT-Portugal Program and FCT grants SFRH / BPD / 65215 / 2009, SFRH / 43043 / 2008 and SFRH / 43081 / 2008, Ministério da Educação e da Ciência, Portugal.

References

[1] L. Ohno-Machado, F. Resnic, and M. Matheny, "Prognosis in critical care," *Ann. Rev. of Biomedical Engineering*, vol. 8, pp. 567–599, 2006.

[2] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modelling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[3] J. M. C. Sousa and U. Kaymak, *Fuzzy Decision Making in Modeling and Control*. Singapore: World Scientific, 2002.

[4] U. Kaymak, W.-M. van den Bergh, and J. van den Berg, "A fuzzy additive reasoning scheme for probabilistic mamdani fuzzy systems," in *Proc. 12th IEEE Int. Conf. Fuzzy Syst.*, 2003, pp. 331–336.

[5] U. Kaymak and J. van den Berg, "On probabilistic connections of fuzzy systems," in *Proc. 15th Belgian-Dutch Conf. Artif. Intell.*, 2003, pp. 187–194.

[6] Members of the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference Committee, "Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis," *Crit Care Med*, vol. 20, 1992.

[7] Y.Arabi, N. A. Shirawi, Z. Memish, S. Venkatesh, and A.Al-Shimemeri, "Assessment of six mortality prediction models in patients admitted with severe sepsis and septic shock to the intensive care unit: a prospective cohort study," *Crit Care Med*, vol. 7, pp. R116–R122, 2003.

[8] F. Jaimes, J. Farbiarz, D. Alvarez, and C. Martnez, "Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room," *Crit Care Med*, vol. 9, 2005.

[9] V. Ribas, A. Vellido, J. Ruiz-Rodriguez, and J. Rello, "Severe sepsis mortality prediction with logistic regression over latent factors," *Expert Systems with Applications*, vol. 39, 2012.

[10] E. Hanisch, R. Brause, J. Paetz, and B. Arlt, "Predicting death for patients with abdominal septic shock," *Journal of Intensive Care Medicine*, vol. 26, 2009.

[11] J. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *J Clin Epidemiol*, vol. 49, pp. 1225–1231, 1996.

[12] J. van den Berg, U. Kaymak, and W.-M. van den Bergh, "Fuzzy classification using probability-based rule weighting," in *Proc. 11th IEEE Int. Conf. Fuzzy Syst.*, 2002, pp. 991–996.

[13] L. I. Kuncheva, *Fuzzy Classifier Design*. Heidelberg: Springer, 2000.

[14] V. Kecman, *Learning and Soft Computing*. MIT Press, Cambridge, MA, 2001.

[15] A. H. Meghdadi and M.-R. Akbarzadeh-T, "Probabilistic fuzzy logic and probabilistic fuzzy systems," in *IEEE International Fuzzy Systems Conference*, 2001, pp. 1127–1130.

[16] Z. Liu and H.-X. Li, "A probabilistic fuzzy logic system for modeling and control," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 6, pp. 848–859, 2005.

[17] L. Waltman, U. Kaymak, and J. van den Berg, "Maximum likelihood parameter estimation in probabilistic fuzzy classifiers," in *Proc. 14th IEEE Int. Conf. Fuzzy Syst.*, 2005, pp. 1098–1103.

[18] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rulebased classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 506–515, 2001.

[19] L. Zadeh, "Probability measures of fuzzy events," *Journal of Mathematical Analysis and Applications*, vol. 23, pp. 421–427, 1968.

[20] J. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, 1981.

[21] J. van den Berg, U. Kaymak, and W.-M. van den Bergh, "Probabilistic and statistical fuzzy set foundations of competitive exception learning," in *Proc. 10th IEEE Int. Conf. Fuzzy Syst.*, 2001, pp. 1035–1038.

[22] E. Hanisch, R. Brause, B. Arlt, J. Paetz, and K. Holzer, "The medan database," Tech. Rep., 2009.

[23] L. F. Mendonça, S. M. Vieira, and J. M. C. Sousa, "Decision tree search methods in fuzzy modeling and classification," *International Journal of Approximate Reasoning*, vol. 44, no. 2, pp. 106–123, 2007.

[24] F. Cismondi, A. Fialho, S. Vieira, J. Sousa, S. Reti, M. Howell, and S. Finkelstein, "New methods to processing misaligned, unevenly sampled time series containing missing data," in *Proc. IEEE Symposium Series on Computational Intelligence - SSCI*, 2011, pp. 224–231.

[25] A. Fialho, F. Cismondi, S. Vieira, J. Sousa, S. Reti, M. Howell, and S. Finkelstein, "Missing data in large intensive care units databases," *Crit Care Med*, vol. 38, 2010.

[26] A. Bensaid, L. Hall, J. Bezdek, L. Clarke, M. Silbiger, J. Arrington, and R. Murtagh, "Validity-guided (re)clustering with applications to image segmentation," in *IEEE Trans. on Fuzzy Systems*, 1996, pp. 112–123.

[27] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-mean model," in *IEEE Transactions on Fuzzy Systems*, 1995, pp. 370–379.

[28] A. J. V. E. W. Steyerberg, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan, "Assessing the performance of prediction models: a framework for traditional and novel measures," *Epidemiology*, vol. 21, no. 1, pp. 128–138, 2010.

[29] W. Knaus, D. Wagner, and J. Lynn, "Short-term mortality predictions for critically ill hospitalized adults: science and ethics," *Science*, vol. 18, pp. 389–394, 1991.