Expert Systems With Applications 79 (2017) 244–253

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Daily prediction of ICU readmissions using feature engineering and ensemble fuzzy modeling

Rita Viegas[a], Cátia M. Salgado[a,*], Sérgio Curto[b], João P. Carvalho[b], Susana M. Vieira[a], Stan N. Finkelstein[c,d]

[a] IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Portugal
[b] INESC–ID, Instituto Superior Técnico, Universidade de Lisboa Lisboa, Portugal
[c] Massachusetts Institute of Technology, Engineering Systems Division, Cambridge, MA, USA
[d] Massachusetts Institute of Technology, Harvard-MIT Division of Health Sciences & Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 7 June 2016
Revised 23 February 2017
Accepted 24 February 2017
Available online 28 February 2017

Keywords:
Feature engineering
Ensemble modeling
Fuzzy clustering
ICU
Readmissions

ABSTRACT

This research is focused on the prediction of ICU readmissions using fuzzy modeling and feature selection approaches. There are a number of published scores for assessing the risk of readmissions, but their poor predictive performance renders them unsuitable for implementation in the clinical setting. In this work, we propose the use of feature engineering and advanced computational intelligence techniques to improve the performance of current models. In particular, we propose an approach that relies on transforming raw vital signs, laboratory results and demographic information into more informative pieces of data, selecting a subset of relevant and non–redundant variables and applying fuzzy ensemble modeling to the feature–engineered data for deriving important nonlinear relations between variables. Different criteria for selecting the best predictor from the ensemble and novel evaluation measures are explored. In particular, the area under the sensitivity curve and area under the specificity curve are investigated. The ensemble approach combined with feature transformation and feature selection showed increased performance, being able to predict early readmissions with an AUC of 0.77 ± 0.02. To the best of our knowledge, this is the first computational intelligence technique allowing the prediction of readmissions in a daily basis. The high balance between sensitivity and specificity shows its strength and suitability for the management of the patient discharge decision making process.

1. Introduction

A clinician's decision on the discharge of a patient is always a delicate task. But when such decision takes place in intensive care units (ICUs) where the patients' health condition is extremely fragile and the case of early discharge and unplanned readmission can result in severe consequences, an even greater amount of responsibility is at stake. Previous studies have shown how ICU readmissions can cause significant distress to patients and result in considerable higher financial costs (Allaudeen, Schnipper, Orav, Wachter, & Vidyarthi, 2011), and how in–hospital mortality rates can be 2 to 10 times greater in patients readmitted to the ICU than in patients that are not readmitted (Rosenberg & Watts, 2000).

Moreover, hospital 30–day readmissions payment penalties imposed by the Centers for Medicare and Medicaid (CMS) in the United States are leading to a worldwide increased interest in finding solutions that tell which patients are going to be readmitted in a short period of time if they are discharged at a certain moment (Futoma, Morris, & Lucas, 2015). Information stored in large repositories of health records might provide solutions for this problem, yet, much of the published data–based risk models are unsuitable for use in the clinical setting due to their poor predictive power (Futoma et al., 2015). The question remains in how to make most of this data and translate it into clinical useful knowledge.

Traditionally, logistic regression models of independent variables, typically chosen by hand, have been used to develop readmission risk scores (Badawi & Breslow, 2012; Campbell, Cook, Adey, & Cuthbertson, 2008; Frost et al., 2010; Gajic et al., 2008; Ouanes et al., 2012). Statistically significant risk factors for ICU readmission have been systematically reported and the most commonly identified factors include: patient location before ICU admission; SAPS II score at the time of ICU admission; APACHE II score at the time of

* Corresponding author.
E-mail addresses: ritadovi@gmail.com (R. Viegas), catiasalgado@live.com (C.M. Salgado), scurto@gmail.com (S. Curto), joao.carvalho@inesc-id.pt (J.P. Carvalho), susana.vieira@tecnico.ulisboa.pt (S.M. Vieira), snfinkel@mit.edu (S.N. Finkelstein).

http://dx.doi.org/10.1016/j.eswa.2017.02.036
0957-4174/

admission; age; length of stay in the ICU; physiologic abnormalities at the time of ICU discharge or on the ward; ICU discharge at night or after hours; discharge to another critical care area or hospital; shock index (heart rate/systolic blood pressure); respiratory rate and Glasgow Coma Score and higher Nursing Activity Score at the time of discharge; unstable vital signs, especially respiratory and heart rate abnormalities, and the presence of poor pulmonary function at time of ICU discharge (Rosenberg, Hofer, Hayward, Strachan, & Watts, 2001; Rosenberg & Watts, 2000).

1.1. Approaches to predicting patients' outcomes

Other successful and more sophisticated learning methods have been employed in clinical classification tasks (Deo, 2015; Fialho et al., 2012; Kim, Kim, & Park, 2011; Miotto, Li, Kidd, & Dudley, 2016; Zheng et al., 2015). Some algorithms can be very robust in differentiating between two classes using nonlinear boundaries, especially when the dataset includes complex characteristics that cannot be linearly separable in the feature space, as in the case of health care data. In this work, nonlinearity is explored using the combination of local linear models and fuzzy logic. The main advantage of fuzzy modeling over conventional approaches such as logistic regression, lies in its ability to obtain transparent insights of the system being modeled. Moreover, fuzzy models are particularly suited for developing clinical decision support systems, where knowledge representation with exact mathematical reasoning is almost always poor and inadequate (Minutolo, Esposito, & De Pietro, 2016).

The key point in multimodeling, or ensemble modeling, is the creation of many models such that the combination of their output improves the performance of a single model (Chadli & Borne, 2012; Polikar, 2006). Ensemble modeling has been applied in a wide range of complex real world problems with great success (Dietterich, 2000; Ghosh, 2002; Kittler, Hatef, Duin, & Matas, 1998; Sharkey & Sharkey, 1997). In particular, in the ICU domain, studies reporting ensemble approaches have shown promising results. In a recent study by Salgado, Vieira, Mendonça, Finkelstein, and Sousa (2015), significantly better results were obtained using an ensemble classifier scheme to predict the necessity of vasopressors in sepsis patients, in comparison with the traditional whole–population modeling strategy. In Cismondi et al. (2012), an ensemble approach with two models was developed to predict the survival of ICU septic shock patients, where one model aims to maximize sensitivity and the other aims to maximize specificity.

As the clinical condition of patients changes over time, the risk of an adverse event such as mortality or readmission will vary day by day; single predictions, usually calculated within 24h of ICU admission, fail to account for the clinical dynamics and are therefore less effective in providing treatment guidelines. Studies have tried to address this issue using models capable of adjusting prediction estimates over the length of stay. In Huang et al. (2013), discrete time event logistic regression models were developed to predict daily probability of mortality from day 3 to day 28 after ICU admission. In Cai et al. (2016), a model for predicting the daily probability of a hospitalized patient to remain in the hospital, be discharged, readmitted or dead within 7 days is proposed. The model uses electronic health records of data collected in the emergency department and is based on a Bayesian network that is updated each time a new laboratory result is available. Last, Tosas, Cassarino, Kozlakidis, and Edgeworth (2016) also propose a model that is refined daily as new attributes are recorded; logistic regression, decision trees and the incremental information network algorithms are used to predict mortality in the ICU.

In the ICU, clinicians do not know *a priori* when is the last day of a patient in the ICU and have to evaluate their state continuously in order to make the decision of discharge. Rather than using one model per day, we propose a feature transformation approach that allows the same model to be trained for different time points of the ICU stay, allowing a continuous evaluation of patients. Thus, we propose a scheme to predict if a patient is going to be readmitted to the ICU within 24 to 72 h if discharged at the evaluation moment. The scheme is divided in three main steps: feature construction, feature selection and ensemble prediction, as presented in Fig. 1. In the feature construction step, the time series data from the patients' electronic health records are discretized into several features per variable. In the second step, feature selection is performed in order to find a subset of relevant and non–redundant features that maximizes both sensitivity and specificity. In order to enforce the maximization of sensitivity in one model and the maximization of specificity in the other model, three criteria are investigated (feature selection criteria). In the third step, an ensemble approach is proposed, where the final classification is obtained either by combining both models in the ensemble, or by selecting the best one (decision criteria). This way, we aim to make the most of the information available about the patients' full hospitalization in the ICU and to develop models more suitable for real world applications.

2. Ensemble modeling

In binary classification, one is generally concerned with achieving a good trade–off between sensitivity and specificity. However, it may not be possible to simultaneously improve both measures using only one objective function. The ensemble model implemented in this work integrates two independent models, one developed with the aim of maximizing sensitivity, i.e. the amount of readmitted patients correctly classified, and the other with the aim of maximizing specificity, i.e. the amount of patients correctly classified as not readmitted, similarly to what was proposed in Cismondi et al. (2012). The role of the ensemble is mainly to provide a means of integrating two objective functions, so that one model is tuned to easily identify a readmission and the other model is better at identifying a no–readmission.

When classifying a new patient, both models are tested separately and the final output corresponds to the output of the individual model with less associated uncertainty, or to the aggregation of the contribution of both models, depending on the selected criterion.

2.1. Decision criterion

To obtain the final output, three decision criteria were considered to either select one or combine the two outputs generated by the ensemble.

1. The first criterion consists in averaging the outputs of each model. This is a common strategy for combining the outputs of individual classifiers in ensemble–based approaches (Polikar, 2006; Salgado, Azevedo, Garibaldi, & Vieira, 2015). Being an aggregation–type of strategy, it reduces the chances of selecting a poor classifier from the ensemble, a situation that might occur when using selection–type strategies. Using this criterion, for an output $y_{se,\,n}$ of the sensitivity model and an output $y_{sp,\,n}$ of the specificity model for patient n, the output $y_{\delta,\,n}$ provided by the ensemble is given by:

$$y_{\delta,n} = \frac{y_{se,n} + y_{sp,n}}{2} \tag{1}$$

The threshold α that gives the best balance between sensitivity and specificity ($c^{se}_{Si\alpha}$ and $c^{sp}_{Si\alpha}$ described in Section 2.3) is used to turn the continuous output of $y_{\delta,\,n}$ into a binary classification.

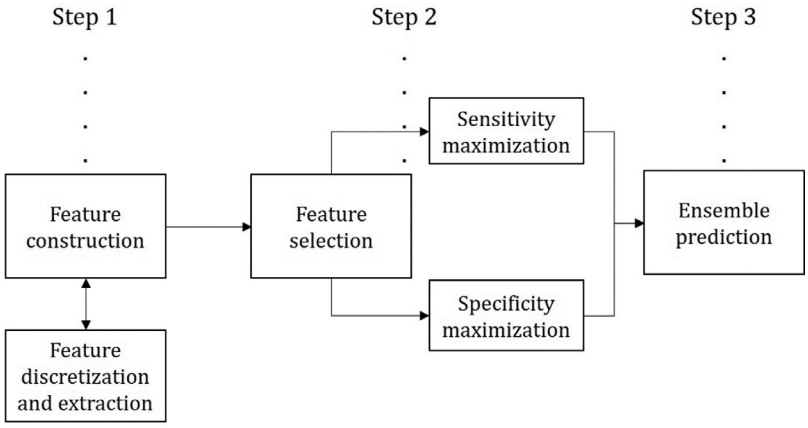

Fig. 1. Proposed scheme for predicting readmissions.

2. The second criterion consists in choosing the predicted outcome with the lower associated uncertainty. This is called an *a posteriori* decision criterion, and follows the procedure proposed in Salgado, Vieira et al. (2015) and Fernandes, Silva, Vieira, and Sousa (2014). For each patient n, the difference between the threshold and the predicted value is calculated, and the model with the higher difference (more discrimination) is selected. This way, the decision is performed according to:

$$l = \arg\max_{se,sp}(|\alpha_{se} - y_{se,n}|, |\alpha_{sp} - y_{sp,n}|) \tag{2}$$

$$y_{\delta,n} = y_l \tag{3}$$

When in the presence of a readmission test patient, the model trained in the feature set leading to the higher sensitivity will, ideally, be certain over its classification, i.e., it will give a prediction far from the boundary α. On the other hand, the model that was trained in the feature set leading to the higher specificity will not be as good at predicting the output; in this case, selecting the model with lower uncertainty will result in a correct classification.

The threshold α used to turn the continuous output of $y_{\delta,\,n}$ into a binary classification corresponds to the threshold selected in the feature selection step (please refer to Section 2.3).

3. The third criterion consists in combining the two predicted outcomes with different weights, inversely proportional to their degree of uncertainty (Polikar, 2006; Salgado, Azevedo, et al., 2015). This way, more importance is given to the outcome predicted with a lower degree of uncertainty. The weights, w_{se} and w_{sp}, are computed as:

$$w_{se,n} = \begin{cases} \dfrac{|\alpha_{se} - y_{se,n}|}{\alpha_{se}} & , y_{se,n} < \alpha_{se} \\ \dfrac{|\alpha_{se} - y_{se,n}|}{1 - \alpha_{se}} & , y_{se,n} \geq \alpha_{se} \end{cases} \tag{4}$$

$$w_{sp,n} = \begin{cases} \dfrac{|\alpha_{sp} - y_{sp,n}|}{\alpha_{sp}} & , y_{sp,n} < \alpha_{sp} \\ \dfrac{|\alpha_{sp} - y_{sp,n}|}{1 - \alpha_{sp}} & , y_{sp,n} \geq \alpha_{sp} \end{cases} \tag{5}$$

And the output $y_{\delta,\,n}$ is given by:

$$y_{\delta,n} = w_{se}y_{se,n} + w_{sp}y_{sp,n} \tag{6}$$

Similarly to the first criterion, the threshold α that gives the best balance between sensitivity and specificity is the one selected ($c^{se}_{Si\alpha}$ and $c^{sp}_{Si\alpha}$ described in Section 2.3).

The next section describes the algorithm used to derive the specificity and sensitivity models.

2.2. Fuzzy modeling

In contrast with traditional set theory, fuzzy modeling attempts to mimic human reasoning producing what is often referred to as approximate reasoning (Engelbrecht, 2007). In fuzzy sets, elements are assigned to sets to a degree, indicating the certainty (or uncertainty) of their membership. Fuzzy logic is used to deal with such uncertainties in order to infer new facts with a respective degree of associated uncertainty.

In this study, Takagi–Sugeno (TS) fuzzy models (Takagi & Sugeno, 1985) were used. The TS fuzzy model is a nonlinear system represented by fuzzy rules translating the relationships between inputs and outputs, of the type:

R_i: If x_1 is A_{i1} and ... x_Q is A_{iQ} then $y_i = a_i^T x + b_i$, where $i = 1, \ldots, C$ corresponds to the rule number, $x = (x_1, \ldots, x_Q)$ is the input vector, Q is the total number of features, A_{iq} is the fuzzy set for rule R_i and the q^{th} feature, y_i is the consequent function of rule R_i, a_i is the parameter vector of rule R_i and b_i is a scalar offset of rule R_i. The degree of activation β_i for the i^{th} rule is calculated using the product operator:

$$\beta_i = \prod_{q=1}^{Q} \mu_{A_{iq}}(x), \tag{7}$$

where $\mu_{A_{iq}}(x) : \mathbb{R} \to [0, 1]$.

In TS fuzzy modeling, decisions are based on the testing of all of the rules y_i in the model. The final output y of a test instance is determined through the weighted average of the individual outputs y_i obtained by each rule:

$$y = \frac{\sum_{i=1}^{C} \beta_i y_i}{\sum_{i=1}^{C} \beta_i} \tag{8}$$

The number of rules C is defined by the number of groups (clusters) objects are sorted into, and the antecedent fuzzy sets A_{iq} are determined using fuzzy clustering in the product space of the input and output variables (Sousa & Kaymak, 2002). In order to determine the antecedent fuzzy sets A_{iq}, the partition matrix obtained with the objective function–based Gustafson–Kessel algorithm is projected onto the input variable axes. The consequent parameters for each rule are obtained as a weighted ordinary least–square estimate.

Since the output y is a linear consequent that results in a continuous output, a threshold α is required to transform y into a final binary output. In this work, label "0" indicates that the patient is classified as not readmitted and label "1" that the patient is classified as readmitted.

2.2.1. Gustafson–Kessel

The TS fuzzy models implemented in this work were based on the Gustafson–Kessel (GK) fuzzy clustering algorithm (Gustafson & Kessel, 1979). This algorithm extends the widely known fuzzy c–means (FCM) (Bezdek, Ehrlich, & Full, 1984) by replacing the Euclidean distance with the Mahalonobis distance, which allows the detection of clusters with different geometrical shapes and orientations.

In the GK algorithm, similarly to what is done in the FCM, a predefined number of clusters C is computed, where each cluster is characterized by a prototype v_i defined as:

$$v_i = [v_1, \ldots, v_Q], \tag{9}$$

where $i = 1, 2, \ldots, C$. Each training sample x_n, where $n = 1, 2, \ldots, N$, is assigned with a normalized membership degree μ_{in} to a cluster i. The prototypes v_i represent the centers of the clusters and can be computed through:

$$v_i = \frac{\sum_{n=1}^{N} \mu_{in}^m x_n}{\sum_{n=1}^{N} \mu_{in}^m}, \tag{10}$$

where $m \in [0, \infty[$ is the weighting exponent that determines the degree of fuzziness of the clusters.

The degree of membership of each training sample n to each cluster i is proportional to the distance d_{in} between the cluster prototype and the sample, and is given by:

$$\mu_{in} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d_{in}}{d_{kn}}\right)^{\frac{2}{m-1}}}, \tag{11}$$

where $\mu_{in} \in [0, 1]$. A value of 0 implies that the sample n does not belong at all to cluster i, whereas a value of 1 implies that the sample n belongs with absolute certainty to cluster i.

The problem of assigning a dataset, X, into C clusters is considered optimal when the following objective function is minimized:

$$J(X, U, V) = \sum_{i=1}^{C} \sum_{n=1}^{N} \mu_{in}^m d_{in}^2, \tag{12}$$

where V is a matrix containing all the cluster prototypes v_i and U is a matrix containing all the μ_{in}.

Note that the Gustafson–Kessel algorithm is only used to determine the antecedent fuzzy sets A_{iq} during the training of the TS fuzzy model, and not directly as a classification algorithm.

2.2.2. Model assessment

In supervised learning, several performance measures can be used to assess the performance of a given model (Steyerberg et al., 2010). In this work two classes were considered: readmitted patients and not readmitted patients for readmissions within 24 to 72 h after discharge. To assess the performance of the models, we used sensitivity, which measures the proportion of patients correctly classified as readmitted; specificity, which measures the proportion of patients correctly classified as not readmitted; accuracy, which measures the proportion of correct classifications and area under the receiver–operating characteristic (ROC) curve (AUC). The AUC is obtained by plotting the values of sensitivity against the values of 1 - specificity for a range of thresholds, and then computing the area under it. Values of AUC of 0.50 correspond to a random classifier and 1 to a perfect classifier.

2.3. Feature selection

Feature selection is mainly focused on the goal of selecting relevant and informative features, but it can bring a variety of other advantages, such as diminishing the processing and storing requirements or facilitating the data visualization and understanding, by reducing the dimensionality of the dataset (Guyon, Gunn, Nikravesh, & Zadeh, 2008).

In this work, considering the complexity and large size of the dataset, a greedy search feature selection strategy was implemented in a wrapper way, i.e. based on the performance of a machine learning algorithm, in this case the sequential forward selection (SFS) algorithm. The SFS algorithm is an iterative process where the subset of features that best predicts the output is obtained by sequentially selecting features until there is no improvement in prediction. In SFS, the criterion used to select features and to determine when to stop is chosen based on the objectives of the problem.

Once the criterion is defined, a TS fuzzy model is built for each of the available features and its performance is evaluated in terms of the chosen criterion. The feature that returns the best performance is kept and, in a second round combined with each of the remaining features. A TS fuzzy model is built and evaluated for each new combination and the subset giving higher performance is kept. This process is repeated successively, always increasing the number of selected features, until there is no increase in performance, in which case the process is stopped and the best subset of features found so far is kept.

The sensitivity and specificity of a certain model are intrinsically dependent on the threshold α selected to turn the continuous output of the TS fuzzy model into a 0 or 1 classification. Higher thresholds will have a positive impact in specificity and a negative impact in sensitivity, whereas lower thresholds will work inversely. In order to achieve a good balance between both measures, α should be chosen with caution, and in regard with the objectives of the task. Choosing the model that gives higher sensitivity/specificity is ultimately a task of evaluating its performance after application of a certain threshold. We investigate what features are selected when the threshold that provides the minimum difference (higher balance) between sensitivity and specificity is applied. We then apply an increment or decrement to this threshold, so that the model will ideally penalize one measure over the other, even though at a small cost. We also propose one measure that gives a performance of the model that is independent of the selected threshold - the area under the sensitivity/specificity curve. Each measure is composed of a c^{se} criterion, used to select the most appropriate features for the model highlighting sensitivity and a c^{sp} criterion, used to select the most relevant features for the model highlighting specificity. Therefore, two different sets of features are selected at each round of SFS, under each criterion. The three sets of criteria implemented are discussed in more detail next:

1. Area under the sensitivity and specificity curves (c^{se}_{AUS}, c^{sp}_{AUS}).
2. Sensitivity and specificity at the intersection threshold ($c^{se}_{Si\alpha}$, $c^{sp}_{Si\alpha}$).
3. Sensitivity and specificity close to the intersection threshold ($c^{se}_{S\varepsilon\alpha}$, $c^{sp}_{S\varepsilon\alpha}$).

1. In the first case, the specificity, sp_α, and sensitivity, se_α, values for the test samples are computed for each threshold α in the test phase. The threshold is varied between 0 and 1 and the overall sum of sensitivity and specificity values is considered, as shown in Fig. 2.

For the model aiming to highlight sensitivity, more weight w_ϕ is given to the sum of sensitivity values and less weight w_β is given to the sum of specificity values. The opposite is true for the model aiming to highlight specificity, as described, respectively, in Eqs. (13) and (14). In this way, a greater importance is given to the measure aimed to highlight, but at the same time, a compromise is

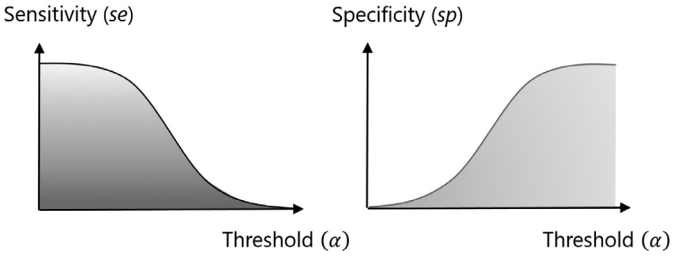

Fig. 2. Area under the sensitivity and specificity curves.

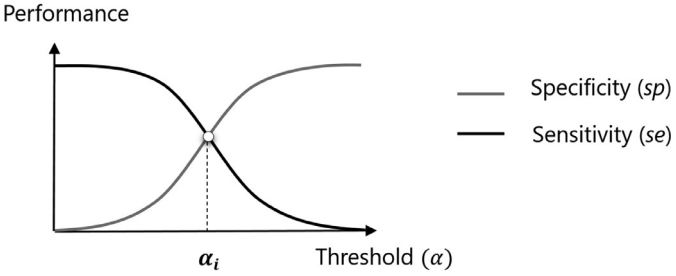

Fig. 3. Sensitivity and specificity at the intersection threshold.

found so that the opposite measure does not present too poor results. Hence, the performance criteria used in feature selection for the sensitivity and specificity models, c^{se} and c^{sp}, are calculated, respectively, through Eqs. (13) and (14), where w_ϕ and w_β are weighting factors heightening the importance to be given to each performance measure.

$$c^{se}_{AUS} = w_\phi \sum_{\alpha=0}^{1} se_\alpha + w_\beta \sum_{\alpha=0}^{1} sp_\alpha \tag{13}$$

$$c^{sp}_{AUS} = w_\phi \sum_{\alpha=0}^{1} sp_\alpha + w_\beta \sum_{\alpha=0}^{1} se_\alpha \tag{14}$$

2. In the second case, the threshold is again varied between 0 and 1, this time in the training phase. The sensitivity and specificity values of the training samples are calculated for each threshold, as well as the difference between the two. The threshold, α_i, corresponding to the minimum difference is chosen and kept for the test phase, as illustrated in Fig. 3.

In the test phase, the values of sensitivity and specificity corresponding to the threshold α_i are computed and, once again, more weight, w_ϕ, is given to the sensitivity value in the model highlighting sensitivity and the opposite for the model highlighting specificity. In this way, models that result in the best values of sensitivity and specificity at the intersection point of both measures can be obtained. The performance criteria used to perform feature selection for the sensitivity and specificity models are calculated through:

$$c^{se}_{Si\alpha} = w_\phi se_{\alpha_i} + w_\beta sp_{\alpha_i} \tag{15}$$

$$c^{sp}_{Si\alpha} = w_\phi sp_{\alpha_i} + w_\beta se_{\alpha_i} \tag{16}$$

3. In the third case, a reference threshold α_i is selected through the previous method. When assessing the values of sensitivity and specificity in the test phase, a slight variation of the reference threshold is chosen. For the model highlighting sensitivity, a threshold α_{se} is obtained by subtracting 0.01 to α_i. For the model highlighting specificity, a threshold α_{sp} is obtained by adding 0.01 to α_i, as represented in Fig. 4.

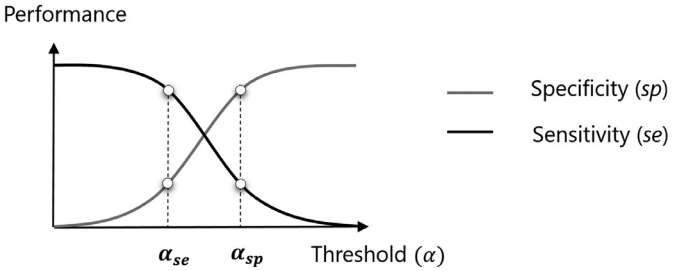

Fig. 4. Sensitivity and specificity close to the intersection threshold.

Such criteria were chosen with the aim of intensifying the importance given to each measure, since values of sensitivity increase as the threshold decreases to zero and the opposite for the values of specificity. The overall expressions to calculate the sensitivity and the specificity criteria are given by:

$$c^{se}_{S\varepsilon\alpha} = w_\phi se_{\alpha_{se}} + w_\beta sp_{\alpha_{se}} \tag{17}$$

$$c^{sp}_{S\varepsilon\alpha} = w_\phi sp_{\alpha_{sp}} + w_\beta se_{\alpha_{sp}} \tag{18}$$

A schematic representation of the proposed ensemble is presented in Fig. 5.

The next section describes the data used to derive the sensitivity and specificity TS fuzzy models.

3. Feature engineering for readmissions

3.1. MIMIC II Database description

This study used data acquired from the Multi–Parameter Intelligent Monitoring for Intensive Care (MIMIC II) database (Scott et al., 2013), a public research archive of data collected from over 32,000 ICU patients between 2001 and 2008, de–identified by removal of all Protected Health Information and containing data collected at the Beth Israel Deaconess Medical Center in Boston.

3.1.1. Data acquisition

Adult patients (older than 15 years) that survived until one year after discharge and that were admitted to the ICU for a minimum of 24 h were selected. Patients who returned to any ICU between 24 and 72 h after discharge were classified as readmitted patients and the ones that did not return to any ICU were classified as not readmitted patients. The choice of 24 h as the lower bound relies on the fact that in the MIMIC II database patients readmitted in less than 24 h are considered as belonging to the same ICU stay. For this cohort of patients, the following information was extracted from the database: body mass index (BMI), a measure of relative size based on weight and heigh; time of admission in the ICU and time and value of the variables listed in Table 1. The choice of these variables with the exception of creatinine, is fundamented on a previous work by Fialho et al. (2012), where data–based modeling was used to discover important features associated with early readmissions. The choice of creatinine inclusion relies on the opinion of physicians, who suggested this variable as an important indicator of the patients' health condition.

3.1.2. Data preprocessing

Since the application of statistical methods may lead to the removal of extreme but possible values, we empirically defined the range of admissible values according to expert recommendations. Extreme values related to the poor health condition of patients are kept, while impossible values (such as negative lactic acid) and probable outliers (such as heart rate above 250 beats/min) are removed. The dataset was restricted to patients containing at least

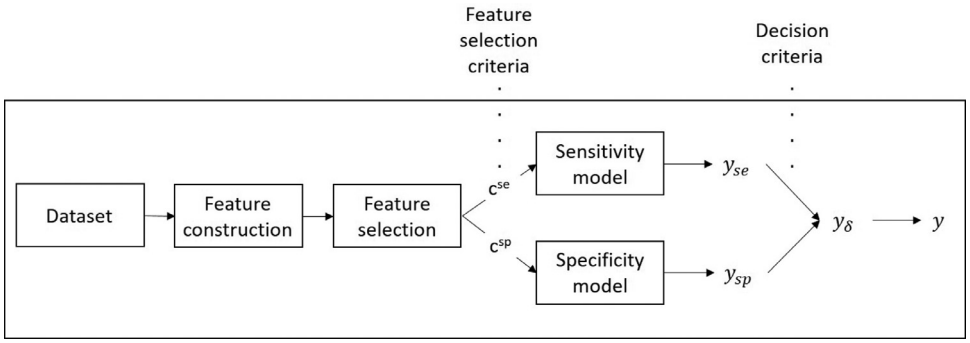

Fig. 5. Ensemble configuration. Three feature selection criteria and three decision criteria are investigated.

Table 1
Physiological variables and admissible ranges considered from the MIMIC II database (Longo, Fauci, Kasper, Hauser, Jameson, & Loscalzo, 2005).

Variable ID (*units*)	Minimum	Maximum
Heart rate (*beats/min*)	0	250
Temperature (°C)	25	42
Platelets (*cells* $\times$ $10^3/\mu L$)	3	1,000
NBP – Non–invasive blood pressure mean (*mmHg*)	10	187
SpO2 – Oxygen saturation in the blood (%)	60	100
Lactic acid (0.5–2.0) (*mg/dL*)	0	10
Creatinine (0–1.3) (*mg/dL*)	0.1	9

Table 2
Feature construction techniques applied to each variable.

v_l	Value of the last recorded measurement
v_f	Value of the first recorded measurement
$v_{\max}$	Maximum value recorded until the evaluation point
$t_{\max}$	Time elapsed between admission and maximum value recorded until the evaluation point
$v_{\min}$	Minimum value recorded until the evaluation point
$t_{\min}$	Time elapsed between admission and minimum value recorded until the evaluation point
v_m	Mean of all values recorded until the evaluation point

two measurements of each of the variables listed in Table 1 and a valid value of BMI, resulting in a dataset comprising 1499 not readmitted patients and 105 readmitted patients. In order to have all data dimensions placed within a similar range, the dataset was normalized in the interval [0,1] according to the min–max normalization (Kuncheva, 2004).

3.2. Feature construction

Feature engineering methods are particularly important when dealing with high dimensional data sets. High dimensional datasets can substantially increase the size of the search space and the chances of a data mining algorithm finding erroneous patterns (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). There are two main categories in which feature engineering methods, also known as feature extraction methods, can be divided: feature construction and feature selection (Guyon et al., 2008; Torkkola, 2003).

Feature construction addresses the problem of finding the transformation of features containing the greatest amount of useful information (Guyon, 2006). Many techniques can be used to obtain such result, from complex linear and nonlinear statistical procedures to simple mathematical operations. When dealing with dynamic variables, as in the case of variables measured throughout the patient stay in the hospital, values such as the mean, the minimum or the maximum can be extracted from the database, in an attempt to retain most of the information contained in the time series, a procedure commonly followed in medical applications (Xia, Daley, Petrie, & Zhao, 2012).

3.2.1. Inputs

In an attempt to mimic as close as possible the way clinicians evaluate their patients when deciding on their discharge, a feature construction approach was performed, using the information contained in the time series of physiological variables recorded until the moment each sample was obtained. The different feature construction techniques applied to each variable (in Table 1) is shown in Table 2.

Additionally, the number of days since admission in the ICU (t_{adm}), age, gender and BMI were included, resulting in a total of 53 features per sample (7 feature construction approaches * 7 variables + t_{adm} + age + gender + BMI).

A schematic comparison of the feature construction approach with a similar approach without feature construction is depicted in Fig. 6. In the latter, only the time since admission and the last value recorded (prime features) are used for modeling readmissions. Intuitively, the feature construction approach is able to provide a more realistic description of the clinical evolution of a patient, resulting in information that can be crucial to extract useful knowledge and improve the accuracy of the developed models.

3.2.2. Outputs

One of the main goals of this work was to develop models capable of predicting patients' readmissions at any day of their ICU hospitalization. In order to mimic real world situations, where clinicians have to evaluate continuously their patients so as to decide on the day of their discharge, it was first necessary to obtain samples representing the different stages of the patients' stay and to assign a label/output to each one.

Taking the sampling frequency of each variable in consideration, a sampling time of 24 h was considered, so that each patient is evaluated once a day until the moment of discharge from the ICU. Since, in principle, the health condition of patients at the moment of admission and in the days preceding discharge is worse than at the moment of leaving the ICU, all samples obtained more than 24 hbefore discharge were assigned to the readmitted patients' class. This approach to output construction was idealized for this specific problem by intensivist physicians. Consider the following example: even so we know beforehand that a particular patient was not readmitted after being discharged in day 4, we should expect that it is more likely that the patient is readmitted in case of discharge in days 1, 2 and 3 than otherwise, hence the model should tell that the patient should not leave yet the ICU. Fig. 7 shows the evolution of the outputs attributed to samples belonging to not readmitted and readmitted patients.

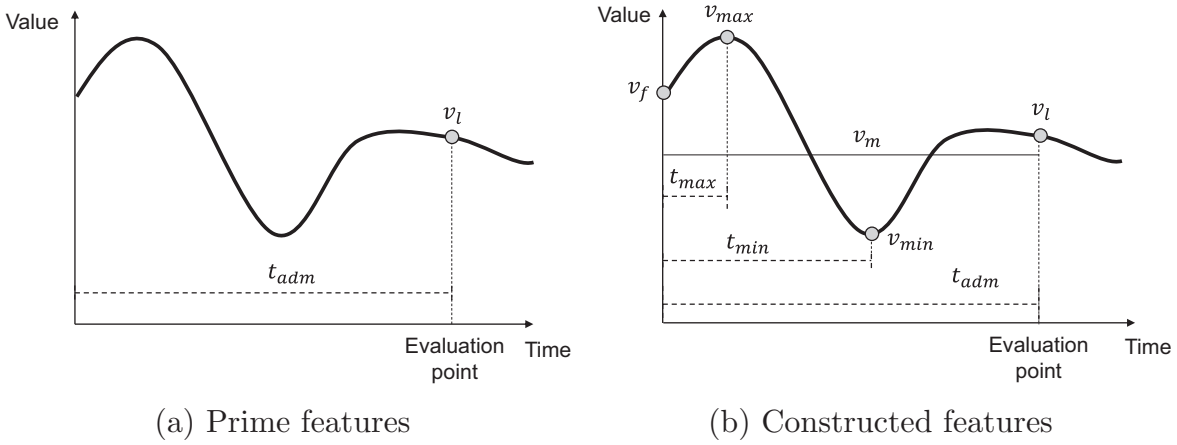

Fig. 6. Features acquired from the time series variables, using two different approaches.

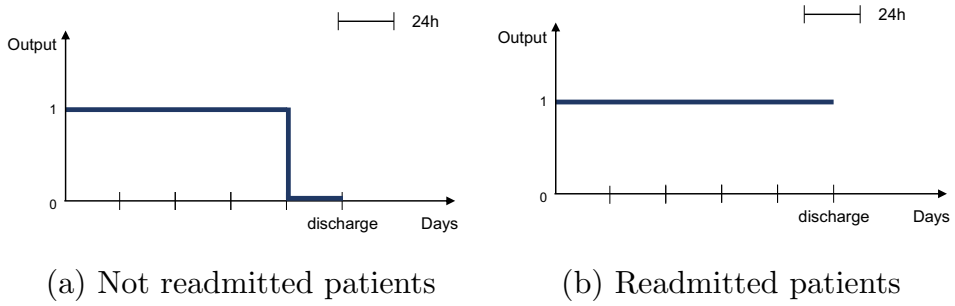

Fig. 7. Output evolution throughout the ICU stay.

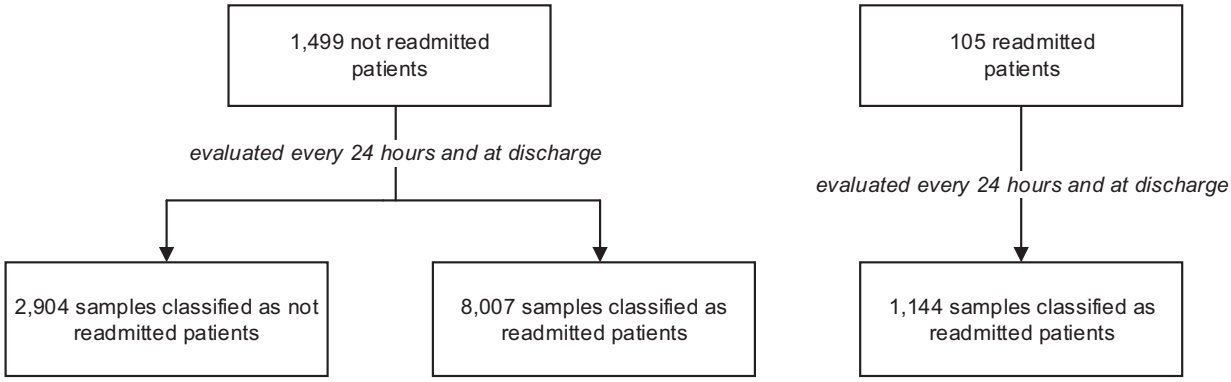

Fig. 8. Flowchart of samples distribution.

After the feature output construction, the final dataset used for modeling comprises 1144 samples from originally readmitted patients and 10,911 samples from originally not readmitted patients, distributed by the two classes as shown in Fig. 8.

4. Results and discussion

In order to build the ensemble model and assess its performance in an independent way, the initial dataset was randomly divided into two equal parts: one for feature selection (FS subset), and the other for model assessment (MA subset).

The FS subset was initially used to select the model parameters over the entire feature space. The weighting factors w_ϕ and w_β were determined empirically, through observation of sensitivity and specificity for various w_ϕ and w_β, in the feature selection set. For $(c^{se}_{AUS}, c^{sp}_{AUS})$, the values w_ϕ and w_β that gave a better trade–off between sensitivity and specificity were 0.8 and 0.2, respectively. For $(c^{se}_{Si\alpha}, c^{sp}_{Si\alpha})$ and $(c^{se}_{S\varepsilon\alpha}, c^{sp}_{S\varepsilon\alpha})$, the best w_ϕ and w_β were 0.7 and 0.3, respectively. These values differ from the first case because the sensitivity and specificity values obtained for the selected thresholds were more uneven than those obtained with the first method, thus being necessary to adjust the corresponding weighting factors. Based on literature, the weighting exponent m of the GK algorithm, which determines the degree of fuzziness of the clustering, is usually set to 2 (Babuška, 1998), but in this case an m of 1.5 produced better AUC. The number of clusters that provided better classification results in terms of AUC was $K = 2$.

Having set the model parameters, sequential forward selection was performed over the FS subset in a 5–fold cross–validation fashion. The MA subset was then used to assess the performance of the ensemble, using AUC, accuracy, sensitivity and specificity as performance measures. In order to reduce variability, 5 rounds of 5–fold cross–validation were performed using different partitions and results were averaged over the rounds. We investigated the combination of three different types of feature selection criteria and three ensemble decision criteria, which resulted in nine mod-

Table 3
Mean and standard deviation of the best results obtained after performing five rounds of 5–fold cross–validation in the MA dataset.

Feature selection criteria		Ensemble decision criteria		
		Average	Maximum distance	Weighted distance
		m_1	m_2	m_3
Area under the sensitivity and specificity curves c^{se}_{AUS}, c^{sp}_{AUS}	AUC	**0.77 ± 0.02**	**0.77 ± 0.02**	**0.77 ± 0.02**
	Accuracy	0.68 ± 0.02	**0.71 ± 0.02**	0.70 ± 0.02
	Sensitivity	0.67 ± 0.03	**0.71 ± 0.03**	0.69 ± 0.03
	Specificity	**0.73 ± 0.03**	0.69 ± 0.03	0.71 ± 0.03
		m_4	m_5	m_6
Sensitivity and specificity at the intersection threshold $c^{se}_{Si\alpha}$, $c^{sp}_{Si\alpha}$	AUC	**0.76 ± 0.02**	0.75 ± 0.02	**0.76 ± 0.02**
	Accuracy	0.67 ± 0.02	**0.71 ± 0.02**	0.69 ± 0.02
	Sensitivity	0.65 ± 0.04	**0.72 ± 0.03**	0.69 ± 0.04
	Specificity	**0.73 ± 0.04**	0.67 ± 0.03	0.70 ± 0.03
		m_7	m_8	m_9
Sensitivity and specificity close to the intersection threshold $c^{se}_{S\varepsilon\alpha}$, $c^{sp}_{S\varepsilon\alpha}$	AUC	**0.75 ± 0.02**	0.74 ± 0.02	**0.75 ± 0.02**
	Accuracy	0.67 ± 0.03	**0.70 ± 0.02**	0.69 ± 0.02
	Sensitivity	0.65 ± 0.04	**0.71 ± 0.03**	0.68 ± 0.03
	Specificity	**0.73 ± 0.03**	0.67 ± 0.03	0.70 ± 0.03

Table 4
Features selected using sequential forward selection according to each individual goal (sensitivity and specificity). Each subset is constituted by common and unique features.

Sensitivity and specificity	
Creatinine: $v_{\min}$	
Lactic acid: $t_{\max}$, $t_{\min}$	
NBP mean: $t_{\max}$, $t_{\min}$, v_l	
Temperature: $v_{\max}$, $t_{\max}$, $t_{\min}$, v_l	
Heart rate: $v_{\max}$	
SpO2: v_l	
Sensitivity	**Specificity**
Platelets: $v_{\max}$	Platelets: $v_{\min}$, v_{mean}
NBP mean: $v_{\max}$	Heart rate: v_l
	SpO2: v_f

Table 5
p-values between the developed models.

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
m_1	–	< 0.05	< 0.05	0.06	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
m_2	–	–	< 0.05	0.20	< 0.05	0.05	< 0.05	< 0.05	< 0.05
m_3	–	–	–	0.11	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
m_4	–	–	–	–	< 0.05	< 0.05	0.07	< 0.05	< 0.05
m_5	–	–	–	–	–	< 0.05	0.43	0.08	0.18
m_6	–	–	–	–	–	–	0.26	< 0.05	0.10
m_7	–	–	–	–	–	–	–	< 0.05	< 0.05
m_8	–	–	–	–	–	–	–	–	< 0.05
m_9	–	–	–	–	–	–	–	–	–

els, identified by m_x in Table 3. From here, it can be concluded that overall, c^{se}_{AUS} and c^{sp}_{AUS} performed better than the other criteria, and $c^{se}_{Si\alpha}$ and $c^{sp}_{Si\alpha}$ performed better than $c^{se}_{S\varepsilon\alpha}$ and $c^{sp}_{S\varepsilon\alpha}$.

For the same ensemble decision criteria, results across different feature selection criteria are very similar. In general, better results were achieved for the maximum distance criterion, however, this criterion tends to penalize the specificity, especially when comparing to the average criterion. Although slightly worse than the results obtained with the maximum distance, and with a slight disfavour for sensitivity, the weighted distance criterion can be chosen in alternative, as it achieves more balanced results. The specificity is more strongly favoured by the average criterion, at the cost of a relatively reduced sensitivity. For the prediction of readmissions, we highlight the ensemble m_2, developed using the area under the sensitivity and specificity curves as feature selection criteria and the maximum distance as decision criterion, as the best approach found. This ensemble achieved an AUC of 0.77, an accuracy and sensitivity of 0.71 and a specificity 0.69.

The list of feature subsets that integrate the ensemble model are present in Table 4. The combination of features selected under the sensitivity model should reflect a higher ability in identifying patients at the risk of being readmitted. In the specificity subset, on the other hand, features should improve the recognition of not readmitted patients. Notice that the selected features tend to focus more on the variation of each variable over time, covering features like the maximum and last value, or time of the maximum and minimum values, than in the absolute value of the variables, as it is done in most of the previous works.

Table 5 shows the *p*–values resulting from the combination of the AUC obtained under the nine different approaches, in the 5 × 5 cross validation partitions. Values < 0.05 enhance the statistical difference between the results obtained for each model, at a significance level of 5%.

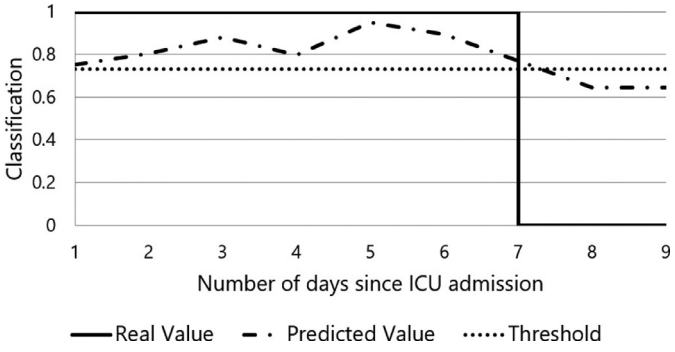

Fig. 9. Real and predicted outcomes of the **not readmitted** patient with ID number 23904.

To further assess the performance of this model, graphical examples of the evolution of the ensemble prediction versus the constructed output, for the full period of hospitalization in the ICU, are presented. Three not readmitted patients are depicted in Figs. 9–11 and a readmitted patient in Fig. 12. Predictions above the threshold indicate an impeding readmission, while predictions below the threshold indicate that the patient can be discharged from the ICU with no risk of being readmitted.

More than single point predictions, it is possible to identify trajectories that indicate expected discharge and expected readmission. In the case of not readmitted patients, the predicted values tend to decrease as the discharge day approaches, even if not in a linear way. In the case of the readmitted patient, although some oscillations can be spotted, the predicted values tend to stay above the threshold, indicating that the patient is going to be readmitted.

Because of the output construction strategy, predictions on day 8, in case of Fig. 10, and in day 5 in case of Fig. 11, are consid-

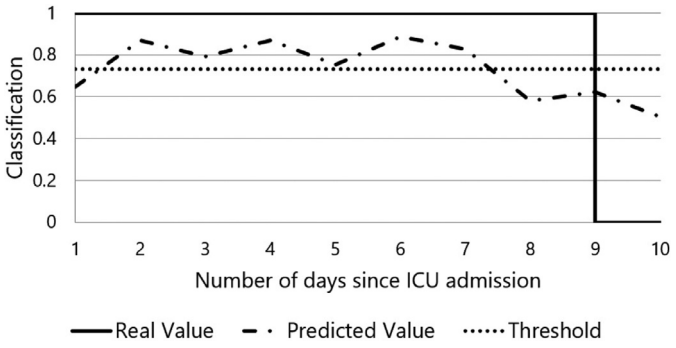

Fig. 10. Real and predicted outcomes of the **not readmitted** patient with ID number 24508.

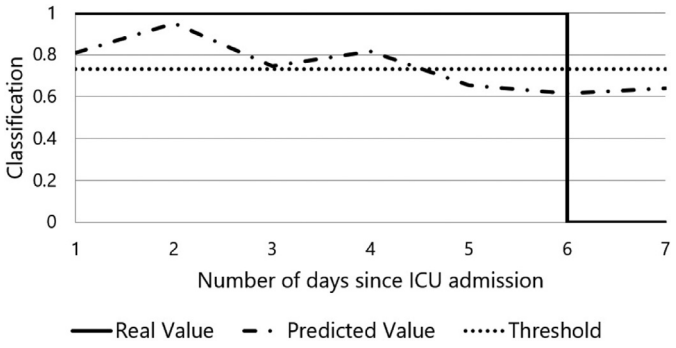

Fig. 11. Real and predicted outcomes of the **not readmitted** patient with ID number 27574.

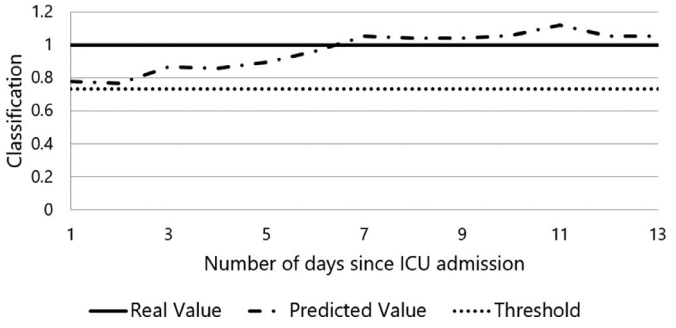

Fig. 12. Real and predicted outcomes of the **readmitted** patient with ID number 23990.

ered false negatives. Although the intersection of the curves with the threshold does not correspond with the day of discharge, the model is in fact anticipating discharge 2 days in advance. Whether or not the patient could have been discharged earlier is a matter of discussion out of the scope of this work, as it would require looking at each case individually and presume the most probable outcome. Nonetheless, while we cannot know if the patient could have been discharged earlier, we can say that the model anticipates the patient outcome in both cases (although this is penalized in the current design). Therefore, we argue that prediction trajectories give valuable information to be used in decision support systems, as their stabilization around a certain output can be viewed as a measure of prediction certainty. However, it should be stressed that these are only examples. The performance of the model can be better assessed in terms of misclassification rate, which in this case is 0.32 (please refer to Table 3).

Fig. 13 presents the results obtained with the best ensemble and with the alternative approaches: single models with prime features and single models with extracted features. From here follows that the performance is significantly improved with the development of both ensemble modeling and input and output feature engineering.

In terms of comparison with previous works, several observations can be made. Campbell et al. (2008) developed multivariate logistic regression models for predicting readmissions in 48 h, which yielded an AUC of 0.67 using 16 variables, and an AUC of 0.63 using APACHE II (which is calculated based on 12 variables). Ouanes et al. (2012) developed the Minimizing ICU Readmission (MIR) score using logistic regression, for determining whether a patient will be readmitted to the ICU within 7 days after discharge, with an AUC of 0.74. Regarding nonlinear modeling strategies, the approach proposed by Fialho et al. (2012) earned an AUC of 0.72 $\pm$ 0.04, an accuracy of 0.71 $\pm$ 0.03, a sensitivity of 0.68 $\pm$ 0.02 and a specificity of 0.73 $\pm$ 0.03, which is worse than the best results presented in this paper in terms of AUC and sensitivity (see Table 3). Fernandes et al. (2014) presented an AUC of 0.74 $\pm$ 0.06, accuracy of 0.74 $\pm$ 0.07, sensitivity of 0.74 $\pm$ 0.16 and specificity of 0.74 $\pm$ 0.09, which means that the proposed ensemble model with feature engineering provides only better AUC and smaller standard deviations in comparison.

More importantly, the works mentioned provide only a means of telling if the patient was readmitted, not a means of telling if the patient is going to be readmitted if discharged at a certain 24 h period, which are very important distinctions from a clinical point of view.

5. Conclusions

This work proposes a feature engineering approach combined with ensemble modeling for the prediction of readmissions in a daily basis. Instead of using raw measurements of physiological variables, we construct features that retain more important aspects of the patients' condition at different points in time, in particular peak values in their lab results and vital signs, and time elapsed between events. Feature selection is performed in order to determine which of these features are more relevant in identifying patients that are going to be readmitted and which features are more relevant in identifying those that will not be readmitted. Thus, a model is constructed for each subset of features, according to certain criteria oriented for sensitivity and specificity maximization. In order to classify each patient once per day in an autonomous way, the prediction of both sensitivity and specificity models is combined, again, in a goal–oriented fashion. In comparison with single models and prime features, the proposed methodology provides a significant increase in the performance, being able to classify readmissions with an AUC of 0.77. These results also outperform results reported in previous studies.

By providing a daily classification we enable the visualization of patient prediction trajectories throughout the ICU stay. Daily prediction provides more valuable information about the patient evolution than single predictions, and can thus better support decision making. As future research we suggest the development of individual models for each day, where the predictions of previous days are fed into the next models.

This is the first study addressing the prediction of readmissions in a daily basis. In the future, smaller time spans should be investigated, for instance for the prediction of readmissions in an hourly basis; current models should be tested using hourly data and new models trained and tested with samples collected by the hour. The feature construction and feature selection steps should be extended to other variables.

Acknowledgment

This work was supported by FCT, through IDMEC, under project iDecision4Care (IF/00833/2014/CP1238/CT0002) and under LAETA,

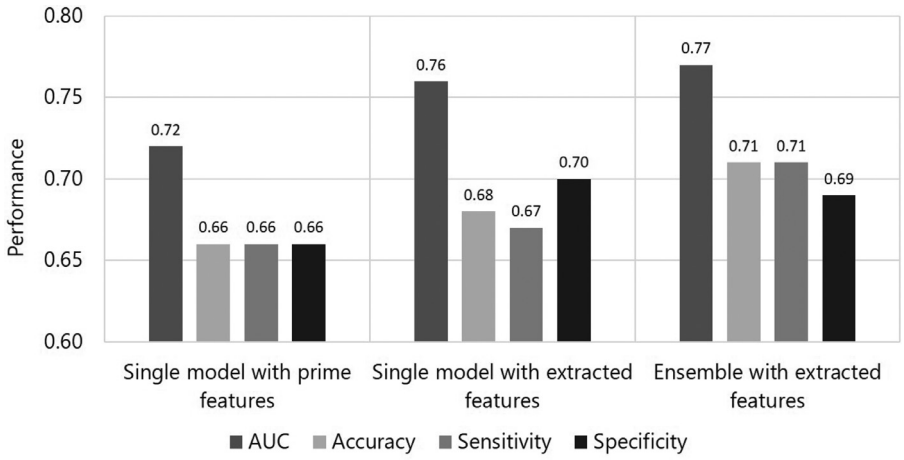

Fig. 13. Performance measures obtained with different approaches.

project UID/EMS/50022/2013. The work of C. M. Salgado was supported by the PhD scholarship SFRH/BD/121875/2016 from FCT. S. Vieira acknowledges support by Program Investigador FCT (IF/00833/ 2014) from FCT, co–funded by the European Social Fund (ESF) through the Operational Program Human Potential (POPH).

References

Allaudeen, N., Schnipper, J. L., Orav, E. J., Wachter, R. M., & Vidyarthi, A. R. (2011). Inability of providers to predict unplanned readmissions. *Journal of General Internal Medicine, 26*(7), 771–776.

Babuška, R. (1998). *Fuzzy modeling for control*: Vol. 12. Springer Science & Business Media.

Badawi, O., & Breslow, M. J. (2012). Readmissions and death after ICU discharge: Development and validation of two predictive models. *PloS One, 7*(11), e48758.

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences, 10*(2), 191–203.

Cai, X., Perez-Concha, O., Coiera, E., Martin-Sanchez, F., Day, R., & Roffe, D. (2016). Real-time prediction of mortality, readmission, and length of stay using electronic health record data. *Journal of the American Medical Informatics Association, 23*(3), 553–561.

Campbell, A. J., Cook, J. A., Adey, G., & Cuthbertson, B. H. (2008). Predicting death and readmission after intensive care discharge. *British Journal of Anaesthesia, 100*(5), 656–662.

Chadli, M., & Borne, P. (2012). *Multiple models approach in automation: Takagi—Sugeno fuzzy systems*. John Wiley & Sons.

Cismondi, F., Horn, A. L., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M., et al. (2012). Multi-stage modeling using fuzzy multi-criteria feature selection to improve survival prediction of icu septic shock patients. *Expert Systems with Applications, 39*(16), 12332–12339.

Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920–1930.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Springer.

Engelbrecht, A. P. (2007). *Computational intelligence: An introduction*. John Wiley & Sons.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37.

Fernandes, M. P., Silva, C. F., Vieira, S. M., & Sousa, J. (2014). Multimodeling for the prediction of patient readmissions in intensive care units. In *Fuzzy systems (fuzz-IEEE), 2014 IEEE international conference on* (pp. 1837–1842). IEEE.

Fialho, A. S., Cismondi, F., Vieira, S. M., Reti, S. R., Sousa, J. M., & Finkelstein, S. N. (2012). Data mining using clinical physiology at discharge to predict ICU readmissions. *Expert Systems with Applications, 39*(18), 13158–13165.

Frost, S. A., Tam, V., Alexandrou, E., Hunt, L., Salamonson, Y., Davidson, P. M., et al. (2010). Readmission to intensive care: Development of a nomogram for individualising risk. *Critical Care and Resuscitation, 12*(2), 83.

Futoma, J., Morris, J., & Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics, 56*, 229–238.

Gajic, O., Malinchoc, M., Comfere, T. B., Harris, M. R., Achouiti, A., Yilmaz, M., et al. (2008). The stability and workload index for transfer score predicts unplanned intensive care unit patient readmission: Initial development and validation*. *Critical Care Medicine, 36*(3), 676–682.

Ghosh, J. (2002). Multiclassifier systems: Back to the future. In *Multiple classifier systems* (pp. 1–15). Springer.

Gustafson, D., & Kessel, W. (1979). Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings of IEEE CDC* (pp. 761–766).

Guyon, I. (2006). *Feature extraction: Foundations and applications*: Vol. 207. Springer Science & Business Media.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2008). *Feature extraction: Foundations and applications*: Vol. 207. Springer.

Huang, Y. C., Chang, K. Y., Lin, S. P., Chen, K., Chan, K. H., & Chang, P. (2013). Development of a daily mortality probability prediction model from intensive care unit patients using a discrete-time event history analysis. *Computer Methods and Programs in Biomedicine, 111*(2), 280–289.

Kim, S., Kim, W., & Park, R. W. (2011). A comparison of intensive care unit mortality prediction models through the use of data mining techniques. *Healthcare Informatics Research, 17*(4), 232–243.

Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(3), 226–239.

Kuncheva, L. I. (2004). *Combining pattern classifiers: Methods and algorithms*. John Wiley & Sons.

Last, M., Tosas, O., Cassarino, T. G., Kozlakidis, Z., & Edgeworth, J. (2016). Evolving classification of intensive care patients from event data. *Artificial Intelligence in Medicine, 69*, 22–32.

Longo, J., Fauci, A. S., Kasper, D. L., Hauser, S. L., Jameson, J. L., & Loscalzo, J. (2005). *Harrison's principles of internal medicine*: Vol. 2. McGraw-Hill, USA.

Minutolo, A., Esposito, M., & De Pietro, G. (2016). A fuzzy framework for encoding uncertainty in clinical decision-making. *Knowledge-Based Systems, 98*, 95–116.

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports, 6*.

Ouanes, I., Schwebel, C., Français, A., Bruel, C., Philippart, F., Vesin, A., et al. (2012). A model to predict short-term death or readmission after intensive care unit discharge. *Journal of Critical Care, 27*(4). 422–e1

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine, 6*(3), 21–45.

Rosenberg, A. L., Hofer, T. P., Hayward, R. A., Strachan, C., & Watts, C. M. (2001). Who bounces back? Physiologic and other predictors of intensive care unit readmission. *Critical Care Medicine, 29*(3), 511–518.

Rosenberg, A. L., & Watts, C. (2000). Patients readmitted to ICUs*: A systematic review of risk factors and outcomes. *CHEST Journal, 118*(2), 492–502.

Salgado, C. M., Azevedo, C. S., Garibaldi, J., & Vieira, S. M. (2015). Ensemble fuzzy classifiers design using weighted aggregation criteria. In *Fuzzy systems (fuzz-ieee), 2015 IEEE international conference on* (pp. 1–5). IEEE.

Salgado, C. M., Vieira, S. M., Mendonça, L. F., Finkelstein, S., & Sousa, J. M. C. (2015). Ensemble fuzzy models in personalized medicine: Application to vasopressors administration. *Engineering Applications of Artificial Intelligence*.

Scott, D. J., Lee, J., Silva, I., Park, S., Moody, G. B., Celi, L. A., et al. (2013). Accessing the public mimic-ii intensive care relational database for clinical research. *BMC Medical Informatics and Decision Making, 13*(1), 9.

Sharkey, A., & Sharkey, N. (1997). Diversity, selection, and ensembles of artificial neural nets. *Neural Networks and their Applications (NEURAP97)*, 205–212.

Sousa, J. M., & Kaymak, U. (2002). *Fuzzy decision making in modeling and control*: Vol. 27.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., et al. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology, 21*(1), 128. (Cambridge, Mass.)

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on*, (1), 116–132.

Torkkola, K. (2003). Feature extraction by non parametric mutual information maximization. *The Journal of Machine Learning Research, 3*, 1415–1438.

Xia, H., Daley, B. J., Petrie, A., & Zhao, X. (2012). A neural network model for mortality prediction in ICU. *Computing in Cardiology, 39*, 261–264.

Zheng, B., Zhang, J., Yoon, S. W., Lam, S. S., Khasawneh, M., & Poranki, S. (2015). Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Systems with Applications, 42*(20), 7110–7120.