Predicting ICU readmissions based on bedside medical text notes

Sérgio Curto, Joao P. Carvalho
INESC-ID
Instituto Superior Técnico, Universidade de Lisboa
Lisboa, Portugal
scurto@gmail.com, joao.carvalho@inesc-id.pt

Cátia Salgado, Susana M. Vieira, João M. C. Sousa
LAETA, IDMEC
Instituto Superior Técnico, Universidade de Lisboa
Lisboa, Portugal
catiasalgado@live.com, susana.vieira@tecnico.ulisboa.pt, jmsousa@tecnico.ulisboa.pt

***Abstract*— Patients are often discharged prematurely from Intensive Care Units (ICU) due to clinical resource limitations, economic pressure or poor discharge planning. The readmission of such patients is associated with an increased risk of death and is currently viewed as a marker for poor quality care. Several studies have focused on predicting which patients are likely to be readmitted, using techniques such as logistic regression or machine learning algorithms, and based on physiological data measured during the patients' stay at the ICU. So far, no published algorithms have been able to predict readmissions to a satisfactory degree. In this work we hypothesize that physicians' and nurses' notes could give a better explanation of both ICU discharges and readmissions, and propose using the text notes in an ICU database in order to build classification models for the prediction of readmissions. We tested the use of Fuzzy Fingerprints and other traditional text classifiers and compared them to a previously proposed model based on numerical data, obtaining very relevant improvements in the classification results, namely an AUC=0.8.**

Keywords—ICU readmissions; Fuzzy Fingerprints; Text based classification; MIMIC II; Weka.

I. Introduction

Improving quality, clinical effectiveness and reducing costs are nowadays the main concerns of healthcare delivers. In the intensive care unit (ICU), readmissions represent a type of adverse event that receives a lot of attention from the general medical community. Patients readmitted to the ICU have an increased risk of death. As such, it is of interest to keep patients in the ICU until such risk is minimal. However, longer stays are usually a marker for poor quality care, are dissatisfying for patients and family, and represent increased health care costs [4][27]. These contradictory aspects must be conciliated to minimize the readmission rate whenever a discharge decision is taken.

Varying definitions of ICU readmission exist, but many authors consider as readmission the return to the ICU within a time period of 72 hours [23]. In this work we will abide to such definition.

A recent review by Elliott et al. [23] has shown that despite decades of research, overall ICU readmission rates changed little over the last years, ranging from 1.3% to 13.7%. Some readmissions can be attributed to premature discharge from the ICU, either due to clinical resource limitations or poor discharge planning.

In the United States, 30-day readmissions rates of 18% are estimated to cost between 15 and 17 billion annually among Medicare beneficiaries [37]. The Centers for Medicare & Medicaid Services (CMS) recently began using readmission rates as a publicly reported metric and to apply financial penalties to hospitals with rates above a pre-determined risk-standardized goal [9].

To reach the goal of preventing readmissions and death, identifying the group of patients at risk prior to the discharge from the ICU is of paramount importance. Early identification might allow these patients to be kept in the ICU for a longer period, to triage the patient to an appropriate level of ongoing care, and to focus efforts in identifying early signs of deterioration [3].

From a pure clinical point of view, the importance of analyzing which patients should be discharged or kept at the ICU can be explained by the need to balance a prolonged and cautious health care delivery, with the drawbacks associated to their stay, such as increased risk of delirium and exposure to multi-resistant bacteria.

There are different studies focusing on logistic regression by means of multivariate and univariate analysis to assess the risk of readmissions and evaluate outcomes in the critically ill [3][12][20][33][34][36]. Statistically significant risk factors for ICU readmission have been systematically reported in these prospective and retrospective cohort studies. The most commonly identified factors include: patient location before ICU admission; acute physiology score at the time of ICU admission; APACHE II score, age; co-morbidities; ICU length of stay; 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. In spite of the growing popularity of these models among the research community, the role they play in supporting the physicians' decisions and in improving patients' outcomes remains uncertain. Currently, there has been an attempt to improve these conventional standard logistic regression techniques using machine learning algorithms such as artificial neural networks, fuzzy logic and decision trees, which resulted in predictive models with promising results in different ICUs [1][2][6][8]. However, there is no consensus on an ICU discharge risk stratification tool, and so far neither inpatient providers nor published algorithms were able to accurately predict outcomes [9][12][27][28].

Previous methods have used numerical data obtained from physiological variables measured during the patients stay before discharge [1][2][24] to build predictive models. In this work we hypothesize that physicians' and nurses' notes could give a better explanation of both discharges and readmissions, as they are a direct representation of the experts' views on the observable data and thus contain valuable knowledge that should be used to improve the results obtained using only physiological variables. Hence, this work proposes using the text notes in an ICU database in order to build classification models for the prediction of early readmissions. We tested the use of Fuzzy Fingerprints [10][11][29] and other traditional text classifiers, and compare them to a previously proposed model based on numerical data [24].

This paper is organized as follows: Section 2 presents the MIMIC II ICU database and the used dataset. Section 3 presents the numerical data classifier used as a benchmark. Section 4 details the text modeling approach. Evaluation and results are shown in Section 6. Finally, Section 7 presents the overall conclusions and discusses future work.

II. MIMIC II DATABASE AND THE USED DATASET

This study uses data from the Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC II) Database which is a de-identified publicly available ICU database composed of detailed information of more than 32,000 patients [25]. Patients were admitted between 2001 and 2008 to the Beth Israel Deaconess Medical Center (BIDMC), an academic medical center in Boston with 620 beds, 77 of which are exclusive for critical care. MIMIC II includes information regarding patient demographics, physiological measures, procedures, medications, laboratory tests results, fluid balance and nursing notes, organized into a relational database. The MIMIC II database is formed by 32,535 patients, of which 24,580 are adults (>15 years old at time of admission).

Inclusion criteria for the dataset used in this work included adult patients that were ICU patients for at least 24 hours and readmissions back to any ICU of the same medical center between 24 and 72 hours. The reason for choosing 24 hours as the lower bound for the readmission time window is related to how MIMIC II is structured – patients readmitted to the ICU less than 24 hours after their discharge are considered to belong to the same ICU stay. Patients that died within one year, either in hospital or after discharge, were excluded from the not readmitted patients cohort, and stays with less than two measurements in any of the numerical variables described in section III.A, were discarded.

In total, 12,091 not readmitted and 775 readmitted patients were considered. 825 admissions were considered in the latter group due to the fact that some patients were readmitted more than once.

III. NUMERICAL MODELING

Fuzzy modeling is a tool that allows approximation of nonlinear systems when there is no previous knowledge of the problem to be modeled [26]. "Grey box" and transparent models, that allow linguistic interpretation in the form of if-then rules, which are particularly useful in health care scenarios, are obtained using this approach. Takagi-Sugeno (TS) fuzzy models [40] have been previously used to develop readmissions classifiers and were used in this work as a benchmark model for classification based on numerical data.

A. Data and data preprocessing

The following variables were used for numerical classification among those present in MIMIC II: heart rate (beats/min), temperature (ºC), platelets (cells × $10^3/\mu$L), non-invasive blood pressure mean (mmHg), oxygen saturation in the blood (%), lactic acid (mg/dL) and creatinine (mg/dL). These variables were selected based on a previous study where feature selection was used to determine the most relevant features in predicting readmissions [19].

Values outside the physiological ranges presented in TABLE I. were considered as outliers and deleted. Values were normalized between 0 and 1 for modeling purposes, except for for NBP mean, where the considered range was 10-186.7mmHg.

A meta-analysis has shown that the gradient of risk of readmission to ICU is similar regardless of whether severity of illness is measured at admission or at discharge [38]. As so, and following the expert suggestions, the mean values of each physiological variable during the first day of the ICU stay were used as the modeling data. Further testing confirmed that results were not hindered by this choice.

B. Fuzzy Modeling

Takagi-Sugeno (TS) fuzzy models consist of input-output rules of the type:

$$R_i : \text{If } x_1 \text{ is } A_{i1} \text{ and ... and } x_n \text{ is } A_{in}$$
$$\text{then } y_i = a_{i1}x_1 + ... + a_{in}x_n + b_i,$$

where i = 1, 2, ..., K is the rule number in the rule base, R_i is the ith rule, y_i is the output of the ith rule, A_{i1},...,A in are the antecedent fuzzy sets, $a_i = [a_{i1}, ..., a_{in}]$ is a parameter vector, b_i is a scalar offset and $x=[x_1, x_2, ..., x_n]^T$ is the input vector, where n is the number of input variables.

The degree of activation of the ith rule is given by:

$$\beta_i(x) = \prod_{j=1}^{n} \mu_{A_{ij}}(x), \quad (1)$$

where μ_{Aij}: R→[0,1] is the membership function of the fuzzy set A_{ij} in the antecedent of R_i.

The model output is determined through the weighted average of the individual rule outputs.

The number of rules K and the antecedent fuzzy sets A_{ij} are determined using fuzzy c-means clustering [13] in the product space of the input and output variables [14], with A_{ij} being approximated by Gaussian functions.

The consequent parameters of y_i, namely b_i and a_i, are obtained using a weighted ordinary least-square estimate. The threshold t selected to turn the continuous output into a binary classification is determined for each model by evaluation of the train set. This way, the predicted output is 1 if $y \geq t$ and 0 if $y < t$.

The ensemble strategy proposed in [7][24], based on the "a priori" criterion, was used to develop a numerical predictive model. Patients are initially divided in subgroups using fuzzy c-means, and then a TS fuzzy model is created for each subgroup/cluster separately. Upon evaluation, the cluster closer to the patient is the one selected, and the classification of the ensemble is given by the classification of the model of that cluster.

IV. TEXT MODELING

The MIMIC II database includes medical and nurse text reports associated to the stays of patients in the ICU. We approached the problem of predicting ICU readmissions using the text reports via two different classifier alternatives: one based on text fuzzy fingerprints, as used previously in problems such as Authorship Identification[29], Twitter Topic detection [10][11], or textual event detection [22]; and the other based on traditional classifiers implemented on the Weka library[1]. However, several issues had to be solved before even any classification attempt since the reports have several particularities that set them apart from other text sources:

- The reports are not structured as a typical written text – sentences are short, have many abbreviations, a reduced number of function words and most of the words are specific and relevant within the context;
- The reports have a large number of medical technical terms and specific technical abbreviations;
- There are many numerical values associated with physiological variables readings;
- Many different ways of expressing/representing the same information. E.g., dates (23-06-2014; 6/23/14; June, 23 2014, etc.), time (10:14PM; 22:04; 2204, etc.), etc.;
- Text contains a huge number of typographical and other word errors;
- Text contains many other artifacts, such as misplaced control characters that break sentences into paragraphs, escape sequences, etc;
- Text anonymization replaced information with ids and randomly shifted timestamps.

The following excerpts show examples of text extracted from ICU text reports where it is easy to understand the difficulties involved in processing and understanding the report, even for a human (note that it is impossible to show all the mentioned issues in such small excerpts):

"Pt placed on a spont breathing trial @ 13:00, pt resp one time within 10 sec -- unfortunatly his SBP droppd from 100 to 70 rapidly and therefore the trail was d/c'ed."

"Cardiac: BP stable 120-130/60. Pt is on Amiodarone via NGT TID. Tolerating this well. HR 80-95 most of the shift. Has rare to occ. PVC/APC. Swan numbers done Q6hrs as ordered and probably swan will come out today. CVP 7-9, PCW 16-20, CO [**6-2**] and SVR 800-900. He remains on heparin drip which needed to be decreased to 750u/hr at 11PM for PTT 110. Repeat PTT will be sent at 5AM."

The presented particularities prevent the effective use of common Natural Language Processing (NLP) techniques, and hinder the use of those texts as "automatic information providers". As such several steps had to be performed in order to clean and pre-process the text reports.

A. Text Preprocessing

Text preprocessing involved the following steps: text cleaning and extraction, word error correction, medical information normalization and medical information fuzzification. Most of the steps were performed automatically or using semi-supervised methods.

Text cleaning followed a procedure described in [18] involving the use of several regular expressions to normalize white spaces, repeated characters, punctuation, etc. Text extraction consisted in extracting sentences using a segmenter and extracting unigrams using a tokenizer, both recurring to components of the Stanford Parser[2].

Word error correction was deemed necessary due to the huge amount of errors found in the MIMIC II database. As an example of the extent of such typing errors, here is a non-extensive list of the different misspelled variants of the word "abdomen" found in the MIMIC II database: abadomen, abdaomen, abndomen, badomen, abdaomen, abdeomen, abdcomen, abdemon, abdeom, abdoem, abdmoen, abdemon, abdiomen, abdman, abdmen, abdme, abddmen, abbomen, abdmn, abdme, abdmonen, abdonem, abdoben, abdodmen, abdoemen, abdoem, abdoem, abdomin. Just out of curiosity, the incorrect form "abdomin" appears 1968 times in the database. An automatic procedure was developed to detect and correct typographical and other word errors in the MIMIC II text corpus [19]. The procedure uses the Fuzzy Uke Word Similarity (FUWS) [16][17], a similarity measure, that unlike edit distance and common subsequence metrics, takes in consideration linguistically driven misspellings from phonetics of the string, or mistakes resulting from the several input devices used to create the text.

Medical information normalization consisted in detecting different units and abbreviations and normalizing them

[1]http://www.cs.waikato.ac.nz/ml/weka

[2]http://nlp.stanford.edu/software/lex-parser.shtml

(example, convert all temperature measures from Fahrenheit or Kelvin, to Celsius degrees) [18].

Medical information fuzzification was a novel technique [18] that enabled the application of bag-of-words NLP techniques to the MIMIC II text corpus. Sparse medical numerical information prevents proper classification given the number of relevant cases and the large number of possible features (bag-of-words techniques rely heavily on word counts). For example, in bag-of-words, systolic blood pressure (SBP) values of, respectively, 183mmHg, 189mmHg, 181mmHg, 190mmHg are all considered as different features since they are different tokens. If, for classification purposes, we group and account all these occurrences as "Very high SBP" by fuzzifying this information, then we can extract more easily the implicit information regarding what happens when SBP is very high.

With this in mind, a set of fuzzy linguistic terms was attributed to each of the variables presented in TABLE I. We used expert opinions in order to distribute the linguistic terms "extremely low" (below the minimum value), "low", "medium", "high" and "extremely high" (over the maximum value) over the universe of discourse indicated in the table. As an example, Fig. 1 shows the linguistic terms attributed to Temperature. All numerical information concerning those variables within the MIMIC II database was then fuzzified accordingly. Since we are dealing with ICU patients, sometimes the values are out of the regular ranges, so we kept those entries in both extremely low and high ranges (as long as they are not too far from the minimum and maximum values for the range).

B. *Fuzzy Fingerprint Classifier*

Fingerprint identification is a well-known and widely documented technique in forensic sciences. In computer sciences a fingerprint is a procedure that maps an arbitrarily large data item (such as a computer file, or author set of texts) to a much compact information block, its fingerprint, that uniquely identifies the original data for all practical purposes, just as human fingerprints uniquely identify people for practical purposes. Fingerprints are a fast and compact way to identify items.

For text classification purposes, we consider a set of texts associated with a given class to build the class fingerprint. Each word in each text represents a distinctive event in the process of building the class fingerprint. Distinct word frequencies are used as a proxy for the class associated with a specific text. The set of the fuzzy fingerprints of all classes is known as the fingerprint library.

In the case of the ICU medical reports, we use a dual fingerprint system based on unigram (word) occurrence on the texts composing our training set: one fingerprint represents the patients that were readmitted and another the ones that were not readmitted.

The full set of known texts (i.e. the properly classified texts that compose the training set) is processed to compute the top-k words list for each class (readmitted/non-readmitted): consider F_j as the set of all words for all texts belonging to class j; the processing result consists of an ordered k-sized list containing the most frequent distinct words, i.e., a list of k tuples $\{v_i, n_i\}$ where v_i is the i-th most frequent word and n_i the corresponding word; words present on the English NLTK[3] stop words corpus are discarded. For extensive databases the top-k list can be approximated without any significant loss in accuracy in order to improve performance [31][30].

Each top-k list is then fuzzified in order to obtain the class fingerprint. We assign a membership value to each word in the set based only on the <u>order</u> in the list. The reason for using the order instead of the frequency results from empirical experiments that show that the order of the frequency seems more relevant than the frequency actual value [32][39]. The choice

TABLE I. LIST OF VARIABLES FUZZIFIED IN MIMIC II TEXT REPORTS AND RESPECTIVE RANGE.

Variable	Min	Max	units
Respiratory rate	0	250	breaths/min
Heart rate	0	250	beats/min
Temperature	25	42	Celsius
White blood cell count	400	50K	$\times 10^3$cells/μL
Blood urea nitrogen	4	500	mg/dL
Creatinine	0.1	9	mg/dL
Systolic blood pressure	90	180	mmHg
Diastolic blood pressure	60	110	mmHg
Mean blood pressure	70	110	mmHg
Oxygen saturation	60	100	%
Lactic acid	0	10	mg/dL
Platelets	3K	1M	cells/L
Red blood cell count	2K	8K	$\times 10^6$cells/μL
Hematocrit	19	60	%
Sodium	120	160	mEq/L
Potassium	2.2	8	mEq/L
Calcium	4.8	12	mg/dL
Magnesium	0	10	mg/dL
Albumin	0.5	18	mg/dL
Arterial pH	6.8	7.8	
Urine out foley	0	1K	
Mechanical ventilation	--	--	Yes/No
Weight	20	200	Kilograms

of the fuzzifying function is critical, and for this work we used the following three functions (Fig. 2):

- erfc – gives preference to the words that occur more often;
- Pareto – same as erfc but highlights even more the words with higher occurrence;

[3] http://www.nltk.org/

- Pyramid – gives preference to words closer to the average occurrence for the used k value.

Erfc and Pareto functions were previously successfully used in multiclass classification problems, and give preference to the words that occur more often in a given class. Here we are dealing with a dual class problem (a patient is either readmitted or not), hence we decided to develop to a new function specifically tailored for dual class problems. The pyramid function was chosen since we noticed that the initial part of both readmitted and non-readmitted fingerprints usually contain many common non-medical related words, and as such there is not as much discrimination as intended.

The resulting fingerprint, Φ, which is based on the top-k list, consists on a size-k fuzzy vector where each position i contains an element v_i and a membership value μ_i representing the fuzzified value of v_i's rank (the membership of the rank).

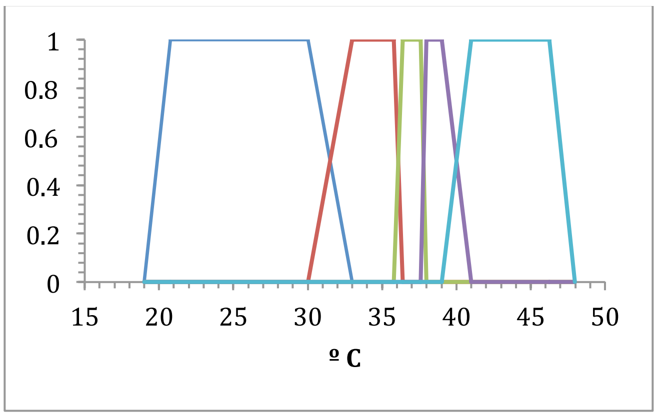

Fig. 1. Membership functions for variable Body Temperature

A class j is represented by its fingerprint $\Phi_j = \Phi(F_j)$. Formally, fingerprint $\Phi_j = \{(v_{ji}, \mu_{ji}) | i = 1..k_j\}$, has length k_j, with $S_j = \{v_{ji} | i = 1..k_j\}$ representing the set of v's in Φ_j. The set of all class fingerprints constitutes the fingerprint library.

In order to find the class of an unknown patient T, we start by computing the size-k fingerprint of T, which we refer as Φ_T. Then we compare Φ_T with the fingerprints Φ_j of both classes present in the fingerprint library. The patient is classified as j if his fingerprint is more similar to Φ_j than to the other class. Fingerprint comparison, $sim(\Phi_T, \Phi_j)$, is calculated using (4):

$$sim(\Phi_T, \Phi_j) = \sum_{v \in (S_T \cup S_j)} \min(\mu_v(\Phi_T), \mu_v(\Phi_j)), \quad (4)$$

where Φ_T is the fingerprint of the patient to be classified, Φ_j the class fingerprint, and $\mu_v(\Phi_x)$ is the membership value associated with the rank of element v in fingerprint x. This function is based on the fuzzy AND. In this case we use the minimum or Gōdel t-norm in accordance with [29], but other t-norms could also be used.

Due to the fuzzy fingerprints mechanic, the unbalanced dataset does not create a bias towards a specific class, and compared to the Weka classifiers presented on next section, the classification of a new patient is a lot faster, since new patients can be added incrementally and there is no need to retrain a model [32].

C. *Weka Classifier*

The second approach to the creation of the patient reports classifier was the use of several algorithms provided by the Weka java library. The dataset was composed of all unigrams present on the patient reports, consisting on 231,852 features/attributes. Since not all unigrams are present on every report we used Weka's *SparseInstance* to represent more efficiently the test set data.

Many Weka algorithms were tested, but several either used too much memory or took too much time to run due to the very high number of features/attributes. The following algorithms were chosen for testing since all were able to complete a 10-fold cross validation: Decision Stump, IBK, MNB (Multinomial Naive Bayes) [5], One R, Random Forest, SVM/SMO (Support Vector Machines using Sequential Minimal optimization) [15][21]. All algorithms were fully tested and optimized, but as explained below, only MNB and SVM/SMO produced acceptable results.

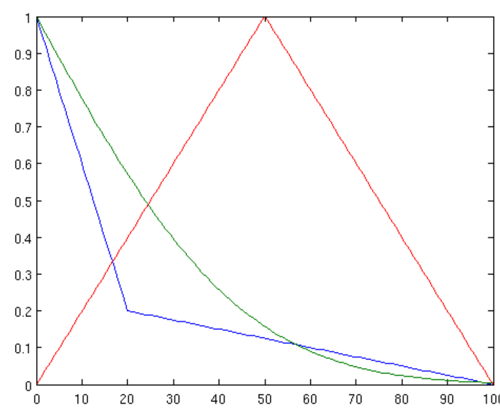

Fig. 2. Fuzzifying functions: erfc (green), Pareto (blue), Pyramid (red)

V. EVALUATION AND RESULTS

A. *Model Assessment*

Cross validation is used to assess the validity and robustness of the models since it avoids bias possibly introduced by selection of a specific training and test set. Cross validation was performed by using 90% of the set as the training set, and the remaining 10% as the test set (10-fold cross validation). Results were averaged over the rounds.

The performance of the models is evaluated using the most relevant and widely used statistical measures when addressing unbalanced datasets, as usual in medical problems: Accuracy (ACC); Specificity or True Negative Rate (TNR); Sensitivity or True Positive Rate (TPR); and Area Under the ROC Curve (AUC).

B. *Optimization and Results*

TABLE II. shows the obtained results with the different successfully tested classifiers.

In what concerns the numerical classifier, the optimal threshold t, selected to turn the continuous output into a binary classification, was found by balancing sensitivity and specificity. In each round of the cross validation, a grid search was performed in order to select the number of clusters *c* that maximize the AUC, such that $K = 2, 3, ..., 10$ and $m = 1.2, 1.3, …,2$. The AUC was integrated over a range of thresholds $t = 0, 0.01, ..., 1$. The optimal threshold is found by balancing sensitivity and specificity. We repeated this procedure for different number of clusters in the ensemble and found that the best results were obtained when using 4 models. The obtained results are in line with previous ICU readmission classification approaches using numerical physiological data. It should be noted that the previous studies also used MIMIC II, but the dataset was not exactly the same due to varying criteria in the choice of what patients are supposed to be considered readmitted.

TABLE II. CLASSIFIER RESULTS USING 10-FOLD CROSS VALIDATION.

Classifier	ACC	TNR	TPR	AUC
Fuzzy Modeling Numerical Data	0.69	0.64	0.63	0.64
FFP, Pyramid, 2000	0.85	0.87	0.56	0.78
FFP, erfc, 4500	0.84	0.85	0.60	0.79
FFP, Pareto, 2500	0.84	0.86	0.58	**0.80**
FFP, Pyramid, 6000	0.42	0.40	**0.71**	0.76
FFP, erfc, 10500	0.83	0.84	0.61	0.78
FFP, Pareto, 500	0.80	0.81	0.60	**0.80**
Naive Bayes	0.82	0.83	0.67	0.76
SVM/SMO	**0.94**	**0.98**	0.32	0.66

In order to optimize the fuzzy fingerprints classifier, only two parameters are relevant: the fingerprint size, *k*, and the fuzzifying function. We tested the classifier for *k* values between 250 and 20,000 words for each of the presented fuzzifying functions. Due to the dual architecture of the developed classifier, it was possible to tailor the parameters to either maximize the True Positive Rate, or the True Negative Rate. The second and third sections of TABLE II. show the best results for each fuzzifying function when maximizing the former and the latter. The ROC curves were generated using Roc[4] software.

The results obtained using Weka were highly variable since certain classifiers perform very poorly due to the very unbalanced dataset. Such classifiers prefer to mark nearly all instances as not readmitted in order to achieve better results. Overall, the best performing Weka algorithms were Naïve Bayes and SVM/SMO, whose results also improved those obtained using the numerical data classifier.

Overall, the obtained results are surprising since they show a significant performance improvement in all measures when using the textual reports data instead of the measured physiological numerical data. This is certainly unusual, but proves the hypothesis concerning the existence of very relevant expert knowledge within those reports. The most accurate classifier and the one with the highest sensitivity (TPR) is the text based SVM/SMO. However these values are achieved with a very low specificity (TNR), which in turn results in the lowest AUC among the successfully tested text based classifiers. The best AUC (0.8) is obtained with two versions of the Fuzzy Fingerprint Classifier using the Pareto inspired membership function.

The best Fuzzy Fingerprint classifiers are very balanced in the sense that they achieve very good values of accuracy and specificity (around 0.85), while keeping acceptable levels of sensitivity (close to 0.6). This results in very good AUC values, well above any other published results using numerical models.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we approached the problem of improving the detection of patients that are likely to be readmitted to an ICU by taking advantage of bedside medical text annotations associated to each patient.

We created and adapted several text classifiers, including balanced classifiers such as fuzzy fingerprint erfc 4500 and 10500, and others that give preference to identification of readmitted patients such as Naive Bayes or fuzzy fingerprint pyramid 6000. The obtained results are very encouraging and show an improved performance over previous classification attempts based on numerical physiological data.

The fuzzy fingerprint classifiers present an additional advantage: without any performance optimization, the entire dataset is analyzed in a few seconds, while the Weka classifiers can take a few minutes to process the data set. And it can easily be almost twice as faster by distributing the computation over 2 processors, as in practice the classifier is composed by two sub-classifiers, one that compares the new patient with the readmitted patients fingerprint and the other with the not readmitted patients.

Fuzzy fingerprint classifier performance could be improved in the future by exploring the usage of bigrams. This would explore the fact that often a small change on words order in a

[4] https://kboyd.github.io/Roc/

sentence may alter its meaning [35]. This approach would have a strong negative impact on the Weka classifiers (time and memory wise) as it would add several new attributes, but would have nearly no impact in the fuzzy fingerprint based approach.

In the case of the Weka classifiers, the use of meta classifiers could improve the results by giving different scores to the correct classification of the readmitted class (with fewer instances) than the not readmitted class. This bonus would be proportional to the size of each class, so that getting a readmitted patient correct would be worth more than a not readmitted one, but getting all instances of a class correct would have the same final score of getting all the other class instances correct, without restricting the training set even further.

Further future work includes experimenting with the fusion of both textual and numerical models in order to increase the accuracy of the individual parts, since each model provides complementary information about the same problem.

ACKNOWLEDGMENT

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, under project IC4U, PTDC/EMS-SIS/3220/2012, and by a FCT post-doctoral grant SFRH/BPD /65215/2009, Ministério da Educação e Ciência, Portugal. S. M. Vieira acknowledges the support by Program Investigador FCT (IF/00833/2014), co-funded by the European Social Fund (ESF) through the Operational Program Human Potential (POPH).

REFERENCES

[1] A. Fialho, F. Cismondi, S. Vieira, S. Reti, J. Sousa, S. Finkelstein, “Data mining using clinical physiology at discharge to predict ICU readmissions”, Expert Systems with Applications 39 (18) (2012) 13158–65, ISSN 0957- 4174, doi:http://dx.doi.org/10.1016/j.eswa.2012.05.086.

[2] A. Fialho, U. Kaymak, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. C. Sousa, S. N. Finkelstein, “Predicting intensive care unit readmissions using probabilistic fuzzy systems”, IEEE International Conference on Fuzzy Systems ISSN 10987584.

[3] A. J. Campbell, J. A. Cook, G. Adey, B. H. Cuthbertson, “Predicting death and readmission after intensive care discharge”, Br J Anaesth 100 (5) (2008) 656–62, ISSN 1471-6771 (Electronic); 0007-0912 (Linking), doi: 10.1093/bja/aen069.

[4] A. L. Rosenberg, C. Watts, “Patients readmitted to ICUs: a systematic review of risk factors and outcomes”, Chest 118 (2) (2000) 492–502, ISSN 0012-3692 (Print); 0012-3692 (Linking).

[5] A. McCallum, K. Nigam, "A comparison of event models for naive bayes text classification." In *AAAI-98 workshop on learning for text categorization*, vol. 752, pp. 41-48. 1998.

[6] B. Zheng, J. Zhang, S. W. Yoon, S. S. Lam, M. Khasawneh, S. Poranki, “Predictive modeling of hospital readmissions using metaheuristics and data mining”, Expert Systems with Applications 42 (20) (2015) 7110–20, ISSN 0957-4174.

[7] C. M. Salgado, C. S. Azevedo, J. Garibaldi, S. M. Vieira, Ensemble fuzzy classifiers design using weighted aggregation criteria, FUZZ-IEEE, 2015 IEEE International Conference on Fuzzy systems, IEEE, 1–5, 2015.

[8] C. Walsh, G. Hripcsak, “The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions”, Journal of biomedical informatics 52 (2014) 418–26.

[9] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, S. Kripalani, “Risk prediction models for hospital readmission: A systematic review”, JAMA 306 (15) (2011) 1688–98, doi: 10.1001/jama.2011.1515.

[10] H. Rosa, J.P. Carvalho, F. Batista, “Detecting a Tweet’s Topic within a Large Number of Portuguese Twitter Trends”, In SLATE'14 - 3rd Symposium on Languages, Applications and Technologies, Schloss Dagstuhl, vol. 4659, series OpenAccess Series in Informatics (OASIcs), pages 185-199, doi: http://dx.doi.org/10.4230/OASIcs.SLATE.2014.185, June 2014

[11] H. Rosa, F. Batista, J.P. Carvalho, “Twitter Topic Fuzzy Fingerprints”, Proc. of the WCCI2014 – World Congress of Computational Intelligence, 2014, Beijing, China, IEEE Xplorer

[12] I. Ouanes, C. Schwebel, A. Francais, C. Bruel, F. Philippart, A. Vesin, L. Soufir, C. Adrie, M. Garrouste-Orgeas, J. Timsit, B. Misset, “A model to predict short-term death or readmission after intensive care unit discharge”, J Crit Care 27 (4) (2012) 422e1–9, ISSN 1557-8615 (Electronic); 0883-9441 (Linking).

[13] J. C. Bezdek, R. Ehrlich, W. Full, “FCM: The fuzzy c-means clustering algorithm”, Computers & Geosciences 10 (2) (1984) 191–203.

[14] J. M. C. Sousa, U. Kaymak, “Fuzzy decision making in modeling and control”, vol. 27, World Scientific Singapore, 2002.

[15] J. Platt, "Fast training of support vector machines using sequential minimal optimization." *Advances in kernel methods—support vector learning* 3, 1999.

[16] J.P. Carvalho, “On the semantics and the use of fuzzy cognitive maps in social sciences”, In WCCI2010, 2010 IEEE World Congress in Computational Intelligence, FUZZ-IEEE 2010, Jul. 2010, pp. 2456-2461, IEEE-Xplorer

[17] J.P. Carvalho, L. Coheur, “Introducing UWS – A Fuzzy Based Word Similarity Function with Good Discrimination Capability: Preliminary results”, In FUZZ-IEEE2013 - 2013 IEEE International Conference on Fuzzy Systems, IEEE Xplore, pages 1369-1376, July 2013

[18] J.P. Carvalho, S. Curto, “Fuzzy Preprocessing of Medical Text Annotations of Intensive Care Units Patients”, In IEEE 2014 Conference on Norbert Wiener in the 21st Century, IEEE, series IEEE Xplorer, Boston, MA, USA, June 2014

[19] J.P. Carvalho, S. Curto, “Towards Unsupervised Word Error Correction in Textual Big Data”, In FCTA2014 - 6th International Conference on Fuzzy Computation Theory and Applications, pages 181-186, Rome, Italy, October 2014

[20] K. Strand, H. Flaatten, “Severity scoring in the ICU: a review”, Acta anaesthesiologica Scandinavica 52 (4) (2008) 467–78, ISSN 1399-6576.

[21] Keerthi, S., Shevade, S.K., Bhattacharyya, C., Murthy, K., "Improvements to Platt's SMO algorithm for SVM classifier design." *Neural Computation* 13, no. 3 2001: pp. 637-649.

[22] L. Marujo, et al, “Textual Event Detection using Fuzzy Fingerprints”, IEEE Intelligent Systems IS’14, Advances in Intelligent Systems and Computing, Sep. 2014 , Vol 322, pp. 825-836, Springer.

[23] M. Elliott, L. Worrall-Carter, K. Page, “Intensive care readmission: a contemporary review of the literature”, Intensive Crit Care Nurs 30 (3) (2014) 121–37, ISSN 1532-4036 (Electronic); 0964-3397 (Linking), doi: 10.1016/j.iccn.2013.10.005.

[24] M. P. B. Fernandes, C. F. Silva, S. M. Vieira, J. M. C. Sousa, “Multimodeling for the Prediction of Patient Readmissions in Intensive Care Units”, in: Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on Fuzzy Systems, ISBN 9781479920723, 1837–42, 2014.

[25] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, R. G. Mark, “Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database”, Critical Care Medicine 39 (2011) 952–60.

[26] M. Sugeno, T. Yasukawa, “A fuzzy-logic-based approach to qualitative modeling”, IEEE Transactions on Fuzzy Systems 1 (1) (1993) 7–31.

[27] N. Allaudeen, J. L. Schnipper, E. J. Orav, R. M. Wachter, A. R. Vidyarthi, “Inability of providers to predict unplanned readmissions”, J Gen Intern Med 26 (7) (2011) 771–6, ISSN 1525-1497 (Electronic); 0884-8734 (Linking), doi:10.1007/s11606-011-1663-3.

[28] N. Allaudeen, J. L. Schnipper, E. J. Orav, R. M. Wachter, A. R. Vidyarthi, “Inability of providers to predict unplanned readmissions”, Journal of general internal medicine 26 (7) (2011) 771–6, ISSN 1525–1497.

[29] N. Homem, J.P. Carvalho, “Authorship Identification and Author Fuzzy Fingerprints”, Proc. of the NAFIPS2011 - 30th Annual Conference of the North American Fuzzy Information Processing Society, 2011, IEEE Xplorer

[30] N. Homem, J.P. Carvalho, “Finding top-k elements in a time-sliding window” Evolving Systems, 2(1), pp. 51-71, Jan. 2011, Springer.

[31] N. Homem, J.P. Carvalho, “Finding top-k elements in data streams”; Information Sciences, 180(24), pp. 4958-4974, Dec. 2010, Elsevier

[32] N. Homem, J.P. Carvalho, “Web User Identification with Fuzzy Fingerprints”, FUZZ-IEEE 2011 - 2011 IEEE International Conference on Fuzzy Systems, Jun. 2011, pp. 2622-2629 , IEEE Xplorer. 10.1109/FUZZY.2011.6007446.

[33] O. Badawi, M. J. Breslow, “Readmissions and Death after ICU Discharge: Development and Validation of Two Predictive Models”, PLoS ONE 7 (11) (2012)

[34] O. Gajic, M. Malinchoc, T. B. Comfere, M. R. Harris, A. Achouiti, M. Yilmaz, M. J. Schultz, R. D. Hubmayr, B. Afessa, J. C. Farmer, “The Stability and Workload Index for Transfer score predicts unplanned intensive care unit patient readmission: Initial development and validation”, Critical care medicine 36 (3) (2008) 676–82.

[35] P. Fialho, L. Coheur, S, Curto, P. Cláudio, A. Costa, A. Abad, H. Meinedo, I. Trancoso, “MEET EDGAR, A TUTORING AGENT AT MONSERRATE”, In ACL, series Proceedings of the 51st Annual Meeting of the Association f, Sofia, Bulgaria, August 2013

[36] S. A. Frost, V. Tam, E. Alexandrou, L. Hunt, Y. Salamonson, P. M. Davidson, M. J. Parr, K. M. Hillman, “Readmission to intensive care: development of a nomogram for individualising risk”, Crit Care Resusc 12 (2) (2010) 83–9.

[37] S. F. Jencks, M. V. Williams, E. A. Coleman, “Rehospitalizations among Patients in the Medicare Fee-for-Service Program”, New England Journal of Medicine 360 (14) (2009) 1418–28, doi:10.1056/NEJMsa0803563, pMID: 19339721.

[38] S.A. Frost, E. Alexandrou, T, Bogdanovski et al., “Severity of illness and risk of readmission to intensive care: a meta-analysis”. Resuscitation 2009; 80: 505-10

[39] T. Sousa et. al, “A hybrid simulated annealing approach to handle energy resource management considering an intensive use of electric vehicles, Energy (67), (2014) 81-96.

[40] T. Takagi, M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control”, IEEE Transactions on Systems, Man and Cybernetics (1) (1985) 116–32.