Predicting Hospital Mortality for ICU Patients: Time Series Analysis

Aya Awad[1,2] and Mohamed Bader-El-Den[1] and James McNicholas[3] and Jim Briggs[1] and Yasser El-Sonbaty[4]

Abstract
Current mortality prediction models and scoring systems for Intensive Care Unit (ICU) patients are generally usable only after at least 24 or 48 hours of admission as some parameters are unclear at admission. However, some of the most relevant measurements are available shortly following admission. It is hypothesized that outcome prediction may be made using information available in the earliest phase of ICU admission. This study aims to investigate how early hospital mortality can be predicted for ICU patients. We conducted a thorough time-series analysis on the performance of different data mining methods during the first 48 hours of ICU admission. The results showed that the discrimination power of the machine learning classification methods after 6 hours of admission outperformed the main scoring systems used in intensive care medicine (Acute Physiology and Chronic Health Evaluation, Simplified Acute Physiology Score and Sequential Organ Failure Assessment) after 48 hours of admission.

Keywords
critically ill, missing values, mortality prediction, patient mortality, time-series analysis

INTRODUCTION

Early physiological monitoring and laboratory surveillance can aid clinicians in making effective interventions to improve patient outcome. Existing severity scoring systems and machine learning approaches give rise to challenges in integrating a comprehensive panel of physiologic variables, and presenting to clinicians interpretable models early in a hospital admission. This problem has particular importance in the Intensive Care Unit (ICU) as patients are necessarily very unwell and there is considerable complexity. Early hospital mortality prediction for ICU patients (EMPICU) remains an open challenge as the majority of the severity of illness scores developed provide risk assessments for ICU patients based on the first 24, 48 or 72 hours of a patient's ICU stay (Luo and others 2016; Celi and others 2012; Pirracchio and others 2015; Ribas and others 2011; Kim and others 2011; Delen and others 2005; Crawford and others 2000; Le Gall and others 1984; Knaus and others 1985; Lemeshow and others 1993; Vincent and others 1996). According to research conducted in (Luo and others 2016), many measurements are not yet available during the first half of the first day (i.e. first 12 hours), as a result data from this time period is usually missing and so excluded from analysis. However, patients receive a great deal of intervention in this period, imposing a burden upon them, and conferring a cost. It is in the interest of both patients, and providers, that intensive care intervention is delivered only where it is likely to be effective. The early identification of patients who are more likely to survive, and more likely therefore to benefit, may help both patients and providers to make informed choices about their care.

Therefore, this study presents a thorough time-series analysis for hospital mortality prediction during the first 48 hours of ICU admission together with examining the impact of missing values on the performance of mortality prediction in order to establish the most effective model for early mortality prediction for ICU patients (EMPICU). The question that emerges is: "Given the ICU patients' medical records, how early in the ICU admission can data mining (DM) methods help in predicting hospital mortality considering the impact of missing measurements, and what are the most effective data mining methods for EMPICU?"

This paper is organized as follows: section II introduces previous work that has been done in ICU mortality prediction, section III presents challenges in ICU data. Section IV introduces the time-series analysis for ICU mortality prediction presented in this research. Section V introduces a framework for early mortality prediction in the ICU. Section VI discusses the results

[1]School of Computing, University of Portsmouth, Buckingham Building, Portsmouth, UK
[2] Department of Business Information Systems, Arab Academy for Science and Technology, Egypt
[3] Critical Care Unit, Queen Alexandra Hospital, Portsmouth Hospitals NHS Trust, UK
[4] Department of Computer Science, Arab Academy for Science and Technology, Egypty

Corresponding author:
Mohamed Bader-El-Den, School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK
Email: Mohamed.Bader@port.ac.uk

and finally section VII concludes the work done in this research.

RELATED WORK IN ICU MORTALITY PREDICTION

This section highlights some data mining challenges in ICU mortality prediction facing medical doctors and data scientists. It provides a review of similar solutions for mortality prediction, including severity scoring systems, real-time models, daily models and data mining approaches.

Scoring systems for mortality prediction

Traditional scoring systems for mortality prediction In this section, we will discuss the following traditional ICU scoring systems: (1) Acute Physiology and Chronic Health Evaluation (APACHE) (Knaus and others 1985), (2) Simplified Acute Physiology Score (SAPS) (Le Gall and others 1993) and (3) Sequential Organ Failure Assessment (SOFA) (Vincent and others 1998).

Several publications in the literature have discussed and compared mortality prediction models for ICU patients that rely on panels of experts or statistical models (Le Gall and others 1984; Knaus and others 1985; Le Gall and others 1993; Poole and others 2012; Lemeshow and others 1993; Rosenberg 2002; Vincent and Singer 2010; Gilani and others 2014). For example, APACHE (Knaus and others 1985) and SAPS (Le Gall and others 1993) assess disease severity to predict outcome. The objective of these models is to characterize disease severity from patient demographics and physiological variables obtained within the first 24 hours after ICU admission in order to assess ICU performance. These models have been refined for use within specified geographical areas, such as France, Southern Europe and Mediterranean countries, and to Central and Western Europe Knaus and others (1991); Le Gall and others (2005); Metnitz and others (2009); Moreno and others (2005); Pirracchio and others (2015); Vincent and Singer (2010); Gilani and others (2014). Using a very different strategy. Hoogendoorn and others (2016) built two prediction models. The methods used were: (1) extraction of high-level (temporal) features from Electronic Medical Records (EMRs) and to build a predictive model; (2) definition of a patient similarity metric with prediction based on the outcome observed for similar patients. Neither approach gave optimal discrimination but the first model, using temporal features (AUROC 0.84), was superior to the patient similarity model (AUROC 0.68). In a recent study (Awad and others 2017), the authors looked at use of random forest in early ICU mortality prediction, however, study does not provide time analysis of the proposed framework.

Prediction systems have evolved since their inception, but have not always led to improved discrimination. APACHE III (Knaus and others 1991) was developed in 1991 and in 2002/2003 APACHE IV (Zimmerman and others 2006) was developed, which provides length of stay prediction equations, in addition to the prediction capability of earlier iterations.. A more detailed comparison of the current APACHE scoring systems is available in (Vincent and Singer 2010). Research in (Le Gall and others 2005) introduced an expanded SAPS II by adding six admission variables: age, gender, length of pre-ICU hospital stay, patient location before ICU, clinical category and presence of drug overdose. Results show that the expanded SAPS II performed better than the original and a customized SAPS II, with an AUROC of 0.879. However, a study conducted by Gilani and others (2014) comparing APACHE scores and SAPS II score, showed that the discrimination of APACHE II (as measured by the area under the receiver operating characteristic curve – AUROC) was excellent (AUROC: 0.828) and acceptable for APACHE III (AUROC: 0.782) and SAPS II (AUROC: 0.778) scores. In addition Kramer and others (2014) found that the discrimination of APACHE IVa was superior with area under receiver operating curve (0.88) compared with Mortality Probability Model (MPM) III (Higgins and others 2007) (0.81) and ICU Outcomes Model/National Quality Forum (0.80) (Kramer and others 2014).

Another traditional scoring systems is the SOFA score (Vincent and others 1998), which is limited to 6 organ systems by looking at respiration, coagulation, liver, cardiovascular, central nervous system, and renal measurements. For each organ system, the score provides an assessment of derangement between 0 (normal) and 4 (highly deranged).

According to the clinical review conducted by Vincent and Singer (2010), the different types of score should be seen as complementary, rather than competitive and mutually exclusive. Scoring systems have focused on providing increasingly refined methods for benchmarking ICU performance, and have laid the foundation for robust systems of quality control, but the use of such tools for individual decision assist, remains unproven.

Early scoring systems for mortality prediction The MPM (Lemeshow and others 1985) was described by Lemeshow et al. in 1985. Initially 137 variables were considered; using statistical techniques the relative importance of each variable was determined and only those with a strong association with outcome retained. This resulted in 7 variables collected at admission and 7 at 24 hours. Unlike APACHE and SAPS, this model could be applied at the time of admission. Further, the physiological variables are recorded as affirmative or negative rather than as an actual number. Lemeshow published an updated form of the model, the MPM II in 1993 (Lemeshow and others 1993). This resulted in two models, $mpm0$ at admission and $mpm24$ at 24 hours. $mpm0$ requires the collection of 15 variables and $mpm24$ a further 8 variables. Both models were shown to be good systems for reliably estimating hospital mortality. At that time (1993) $mpm0$ was, by definition, the only model for estimating early hospital mortality which was independent of treatment.

Another scoring system for early mortality prediction is SAPS-III (Moreno and others 2005). The objective of

the development of SAPS-III was the evaluation of the effectiveness of ICU practices; therefore the focus of the model was on data available at ICU admission or within a day of admission. Missing values were coded as the reference of "normal" category for each variable. In data collection, maximum and minimum values were recorded during a certain time period; missing maximum values of a variable were replaced by the minimum and vice versa. Some regression imputations were performed if noticeable correlations of available values could be exploited. Selection of variables was done according to their association with hospital mortality, together with expert knowledge and definitions used in other severity of illness scoring systems. The objective of using this combination of techniques rather than regression-based criteria alone was to reach a compromise between over-sophistication of the model and knowledge from sources beyond the sample with its specific case mix and ICU characteristics. The study conducted by Poole and others (2012) compared the predictive ability of SAPS-II (originally developed from data collected in 1991/1992) and SAPS-III (developed from data collected in 2002) scores on a sample of critically ill patients. Both scores provided unreliable predictions, but unexpectedly SAPS-III turned out to overpredict mortality compared to SAPS-II.

The MPM and SAPS III attempt early mortality prediction, however they were not used in comparison with our model as most of their attributes are not available in the MIMIC II database and are complex to calculate. On the other hand, the traditional scoring systems - APACHE-II, SAPS-I and SOFA scores were used.

Data Mining Techniques for Mortality Prediction

Various studies have advocated the use of Data Mining techniques for predicting ICU mortality, such as the one proposed by Calvert and others (2016) which attempts to predict mortality 12 hours before in-hospital death. Although the work conducted shows strong predictive accuracy, however we question the practical utility of the tool, which predicts at a point twelve hours from the sampling. It is not clear at what stage in the evolution of a critical care episode that this tool should be employed to best effect. If it were used continuously until such time as a death, it would be very high risk for patients, and for many of them, there already have been a protracted ICU course with the attendant burdens of treatment. Whilst this delay is acceptable where the intended purpose is unit quality benchmarking, it is slow for the purpose of decision assist. In contrast, the model proposed in our study attempts to predict in-hospital mortality shortly after ICU admission. It is our hypothesis that accurate prediction of hospital mortality is possible using data collected in the earliest phase of admission.

Another study that attempted early mortality prediction was proposed by Sadeghi and others (2018) which focuses on specific patient diagnosis. The study proposed a novel method to predict mortality using 12 features extracted from the heart signals of patients within the first hour of ICU admission using the MIMIC-III database. Similar to our work, their study showed that the Random Forests classifier satisfies both accuracy and interpretability better than the other classifiers - linear discriminant, logistic regression, SVM, random forest, boosted trees, Gaussian SVM, and K-nearest neighborhood, producing an F1-score and AUC of 0.91 and 0.93 respectively. The study indicates that heart rate signals can be used for predicting mortality in patients in the ICU. In addition, Crawford and others (2000) concluded that a DT used in their study provided a clinically acceptable mining result in predicting susceptibility of prostate carcinoma patients at low risk for lymph node spread. On the other hand, Ramon and others (2007) reported that the AUROCs of DT based algorithms (DT learning, 65%; first order RF, 81%) yielded smaller areas compared to those of NB networks (AUROC, 85%) and tree-augmented NB networks (AUROC, 82%) in their study on a small dataset containing 1,548 mechanically ventilated ICU patients. Also, the work conducted by Yakovlev and others (2018) showed that overall prediction accuracy was highest (90.0%) for naive Bayes in predicting in-hospital mortality for patients with Acute Coronary Syndrome.

Unlike these models, the framework proposed by our study attempts mortality prediction from physiological data including chart variables, lab tests, vital signs and patient demographics, that are not necessarily related to one specific organ/ diagnosis as the ICU is a very complex environment and normally patients get admitted suffering from several conditions. Early mortality prediction is motivated by the intention to assist clinicians and patients in the assessment of the risks and benefits attending intensive care admission. We hold that it is in the interests of patients, or their advocates, to be informed of a quantitative mortality risk, as early as possible, and preferably before committing to burdensome critical care interventions, whenever that is possible.

Similarly Pirracchio and others (2015) reported that Bayesian Additive Regression Trees (BART) is the best candidate when using transformed variables, while Random Forests outperformed all other candidates when using untransformed variables. Other authors achieved improved mortality prediction using a method based on SVMs (Citi and Barbieri 2012). Davoodi et al. Davoodi and Moradi (2018) proposed a Deep Rule-Based Fuzzy System (DRBFS) to develop an accurate in-hospital mortality prediction in the intensive care unit employing a large number of input variables. The method developed was evaluated against several common classifiers including naïve Bayes, decision trees, Gradient Boosting and Deep Belief Networks. The area under the receiver operating characteristics curve for NB, DT, GB, DBN and proposed method were 73.51%, 61.81%, 72.98%, 70.07% and 73.90% respectively.

Many studies show that customized models perform better than traditional scoring systems Awad and

others (2017). Lee and Maslove (2017) conducted a retrospective analysis using data from the MIMIC-II database; the study concluded that customized models trained on ICU-specific data provided better mortality prediction than traditional SAPS scoring using the same predictor variables. However, ICU is a very complex environment where patients may suffer from more than one condition, which makes it difficult to specify which customized model to use. Therefore, there is a need for general mortality prediction models, which is the focus of this study.

CHALLENGES IN ICU DATA

There are a number of challenges due to the characteristics of typically available ICU data: (1) attribute selection; (2) missing values in data; and (3) the class imbalance problem. In this section, we will briefly discuss each challenge and the details of how we addressed each of these challenges will be discussed further in the paper.

1. Attribute Selection: It is often difficult to decide which attributes in a dataset should be used to construct the model. Therefore, one of the core stages is to select the appropriate attributes; several manual and automatic methods are used to select attributes.
2. Missing values: Not all medical variables/tests are measured for all patients within the first few hours of admission, therefore (for each patient) there may be some missing data. Missing values can be handled either by ignoring those records from the dataset that are not complete, or by filling in missing values by a number of techniques Berry and Linoff (1997).
3. Class imbalance: Class imbalance is a major problem in EMPICU, because the number of patients who die inside the hospital is relatively tiny in comparison with the number who survive. Techniques for dealing with class imbalanced datasets include modifying the dataset (re-sampling) (Berry and Linoff 1997), making the classifier 'cost sensitive' (Perry and others 2015) or a hybrid method that combines both.

TIME-SERIES ANALYSIS FOR MORTALITY PREDICTION USING DM TECHNIQUES

The target of this section is to realize how early is it to predict hospital mortality, considering the impact of missing measurements in early hours of ICU admission. This research performs experimental investigation on ICU patient data using data mining classification techniques to predict mortality. Earlier studies Luo and others (2016) have defined early as the first 12 hours of admission; others have defined it as 24, 48 or 72 hours after admissionLuo and others (2016); Celi and others (2012); Pirracchio and others (2015); Ribas and others (2011); Kim and others (2011); Delen and others (2005); Crawford and others (2000); Le Gall and others (1984); Knaus and others (1985); Lemeshow and others (1993); Vincent and others (1996). These assumptions triggered work done in this research to perform a time-series analysis for mortality prediction over the first 48 hours of ICU admission to try and define how early enough is it to effectively predict mortality in the ICU. The algorithms are evaluated on the PhysioNet/CinC Challenge 2012 Citi and Barbieri (2012) dataset. The study considered 4000 subjects with single ICU stays whose age at ICU admission was 16 years or over in Medical ICU (MICU), Surgical ICU (SICU), Coronary ICU (CICU) or Cardiac Surgery ICU (CSICU), and whose initial ICU stay was at least 48 hours long admitted. The data used for the challenge consisted of 5 general descriptors including age, gender, height, ICU type and initial weight. The remaining variables are 36 time-series (measurements of vital signs and laboratory results) from the first 48 hours of the first available ICU stay of a patient's admission, published previously in Citi and Barbieri (2012).

We employ the RF, PART and BN algorithms. Random Forest is an ensemble learning method for classification that operates by constructing multiple decision trees at training time and outputs the class that gets the majority vote of the individual trees. PART uses partial decision trees (feature subset selection) to generate the decision list shown in the output. Only the final decision list is used in classification. It produces rules from pruned partial decision trees. A Bayesian Network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph Berry and Linoff (1997).

The primary outcome was hospital mortality. Performance measures were calculated using cross-validated area under the receiver operating characteristic curve to minimize bias. All experiments were done using Weka (version 3.7.13; University of Waikato, Hamilton, New Zealand)Hall and others (2009). The results noted in table 3 are AUROC of the average of 10 runs, each run is 10-fold cross-validated. The results are presented in detail in the following subsections.

Experiment Setting

This section presents the results for the top performing DM algorithms - RF, BN and PART. It is important to note here that we have also evaluated a larger set of algorithms, such as Decision Trees - J48 (DT), Support Vector Machines (SVM) and JRip, however they were outperformed by the reported methods. Random Forest is one of the most accurate learning algorithms available. For many datasets, it produces a highly accurate classifier. It runs efficiently on large databases and it has an effective method for estimating missing data and maintaining accuracy when a large proportion of the data is missing. PART uses partial decision trees to generate the decision list shown in the output. Only the final decision list is used in classification. Bayesian Networks are an increasingly popular methods for modelling uncertain and complex domains, such

as medical diagnoses and the evaluation of scientific evidences. They provide a natural way to handle missing data, allow combination of data with domain knowledge, facilitate learning about causal relationships between variables and provide a method for avoiding over-fitting of data Berry and Linoff (1997).

Methods - A total of 4000 ICU patients and 37 time-series variables were selected from every hour over the first 48-hours of a patient's admission for modelling.

We evaluated each of the three data mining algorithms on each of the six versions of the dataset.

1. original datasets (original),
2. datasets after modified by applying the Synthetic Minority Oversampling Technique (SMOTE) Chawla and others (2002), an oversampling technique that involves increasing the size of the minority class with the insertion of synthetic data (original+smote),
3. datasets after replacing missing values with the mean (rep1) to handle the issue of missing values
4. datasets after replacing missing values with mean and then applying SMOTE (rep1+smote),
5. datasets after replacing missing values using the EMImputation algorithm (rep2),
6. datasets after replacing missing values using EMImputation algorithm and applying SMOTE (rep2+smote).

Results

The performance of RF, PART and BN for the six versions of the dataset are displayed in table 1 for simplicity. The performance of the three algorithms on all the original 48 datasets is displayed in Figure 1 (graph a).

Performance Analysis Table 1 shows the performance of the three machine-learning algorithms (at 0.05 confidence level) in predicting hospital mortality among this patient cohort. Results were obtained on the original, original+smote, rep1, rep1+smote, rep2 and rep2+smote datasets as shown in column 1 of table 1. Among the six experiment categories, RF performed best, followed by BN then PR. The most effective RF performance model was obtained on the rep1 with (AUROC = 0.83 $\pm$ 0.03) at hour 48, followed by the original, rep1+smote and rep2 datasets with (AUROC = 0.82 $\pm$ 0.03) at hour 40, then rep2+smote with (AUROC = 0.82 $\pm$ 0.03) at hour 48.

As shown in figure 1 (graph a), there is a dramatic increase in available measurements shown at hour 6 of ICU admission making it a suitable point for in-depth analysis of our proposed framework. We compared the performance of RF, PART and BN on patient data after 6 hours of ICU admission with the performance of SOFA and SAPS scores on patient data after 24 hours of ICU admission to figure out whether our proposed early mortality prediction framework (EMPICU) is relatively effective or not. Due to limited number of figures, we only display figure 4, which displays the performance of all algorithms against SAPS and SOFA scores on only one dataset setting (original dataset). The figure shows that our models after 6 hours of admission outperformed the main scoring systems used in intensive care medicine (APACHE, SAPS-I and SOFA) after 24 hour of admission. The best performing classifier is RF, followed by BN, then PART. As represented on the graph of figure 4, The RF model outperforms the main scoring systems (APACHE-II, SAPS-I and SOFA) both in terms of mortality prediction performance (AUROC) and in terms of time (i.e. early prediction; higher prediction performance at 6 hours after admission compared to that of the scoring systems at 24 hours after admission). Table 2 displays the AUROC and the standard deviation of the best performing model RF at 6 hours after admission and the scoring systems at 24 hours after admission.

Missing Values Analysis We also analyzed the missing values over the 48-hour time interval. Results displayed in Figure 1 (graph a) shows the percentage of available measurements during the first 48 hours of ICU admission. As noted on the graph, a dramatic increase in available measurements is shown at hour 6 of ICU admission and no major increase between hour 24 and hour 48. In addition, Figure 1 (graph b) displays the percentage of missing measurements of all attributes and vital signs attributes during the first 48 hour of ICU admission. As noted on the graph, respiratory rate (RespRate) has the highest percentage of missing values, followed by invasive systolic arterial blood pressure (SysABP) then partial pressure of arterial oxygen (PaO2), while heart rate (HR), Glasgow coma scale (GCS) and temperature (Temp) have the lowest percentage of missing values, while Creatinine is in the middle.

A FRAMEWORK FOR EARLY ICU MORTALITY PREDICTION

In this section, we present the general framework for dealing with early ICU mortality prediction. Figure 2 illustrates how we handle early hospital mortality prediction for ICU patients in this study. From the previous time-series analysis, it is clear that there are a number of challenges in ICU data. The framework addresses three of these: (1) attribute selection; (2) missing values in data; and (3) the class imbalance problem. In this section, we investigate different attribute selections, different methods of handling missing values and class imbalance problem. The focus of the framework in this section is early mortality prediction for ICU patients; by early we mean the first few hours of admission (i.e. 6 hours). We particularly selected the first 6 hours, as it is clear from Figure 1 (graph a) that the percentage of missing measurements significantly increase at the 6-hour threshold. In addition, after consulting several intensivists and considering gaps in literature it appeared that analysing patient data at the 6-hour threshold is a sensible time point, balancing the need for information early in the admission against data adequacy.

In this study, we used the MIMIC II Saeed and others (2011) database for analysis and modelling. In

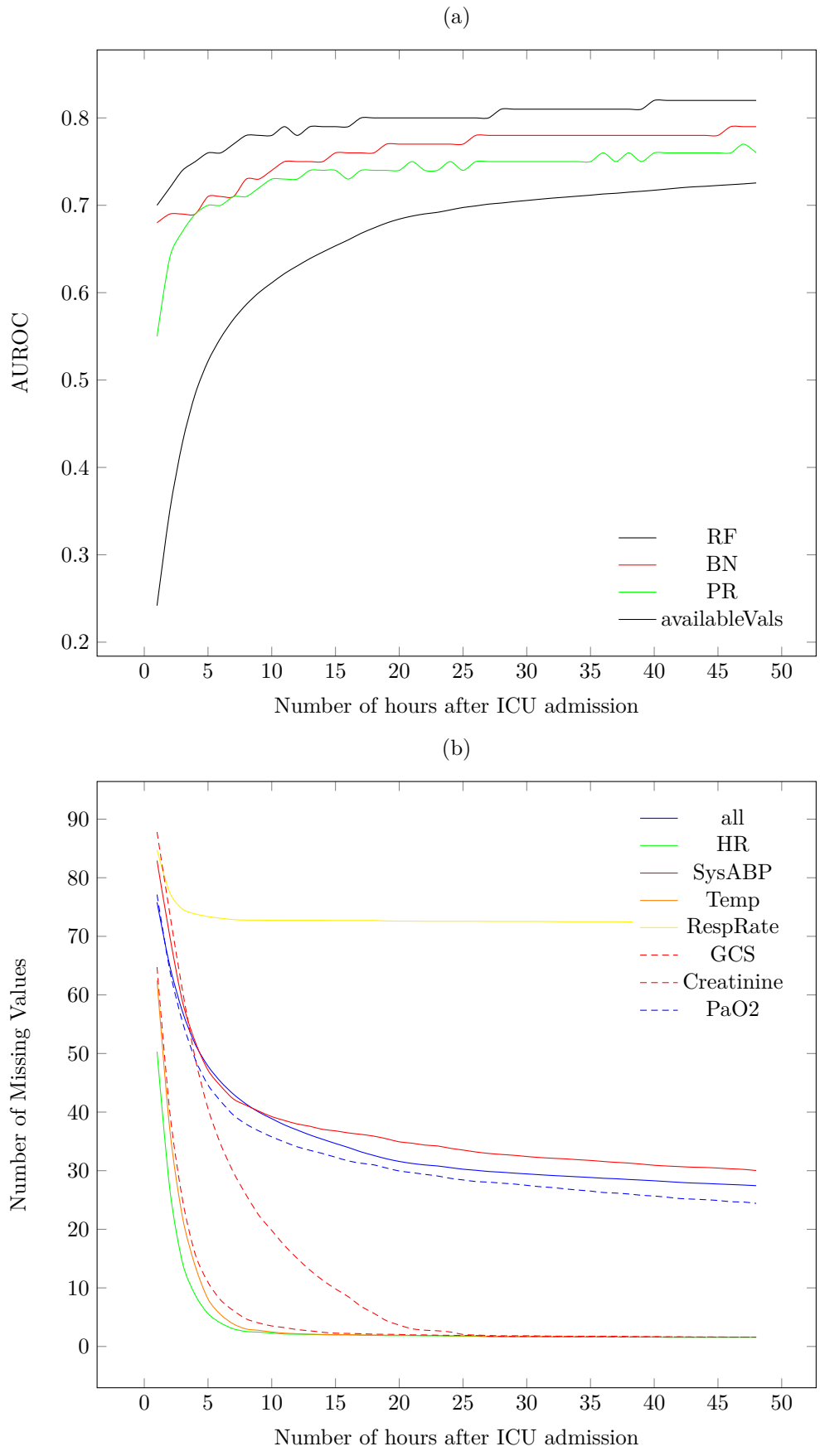

Figure 1. Graph (a) shows the performance of all algorithms on the Yes class (patients at risk of dying inside the hospital) per hour during the first 48 hours of ICU admission together with the percentage of available measurements during the first 48 hours of ICU admission. Graph (b) shows the percentage of missing values of all attributes and vital signs attributes during the first 48 hours of ICU admission

Table 1. Performance of early mortality prediction models developed using 10-fold cross validated RF, PR, and BN in the different experiment settings 2, 4, 6, 8, 10 and 12, 24 and 48 hours after ICU admission, measured with AUROC.

Physionet Dataset	Hours	RF	PART	BN
original	2	0.72 ±0.03	0.64 ±0.04	0.69 ±0.04
original	4	0.75 ±0.03	0.69 ±0.04	0.69 ±0.04
original	6	0.76 ± 0.03	0.70 ±0.04	0.71 ±0.04
original	8	0.78 ± 0.03	0.71 ±0.04	0.73 ±0.03
original	12	0.78 ±0.03	0.73 ±0.03	0.75 ±0.03
original	24	0.80 ±0.03	0.75 ±0.03	0.77 ±0.03
original	48	0.82 ±0.03	0.76 ±0.03	0.79 ±0.03
original+smote	2	0.70 ±0.04	0.65 ±0.04	0.67 ±0.04
original+smote	4	0.73 ±0.03	0.68 ±0.04	0.70 ±0.04
original+smote	6	0.75 ±0.03	0.68 ±0.04	0.71 ±0.04
original+smote	8	0.76 ±0.03	0.70 ±0.04	0.73 ±0.04
original+smote	12	0.77 ±0.03	0.71 ±0.04	0.74 ±0.04
original+smote	24	0.79 ±0.03	0.73 ±0.04	0.76 ±0.03
original+smote	48	0.81 ±0.03	0.75 ±0.04	0.77 ±0.03
Rep1	2	0.73 ±0.03	0.62 ±0.05	0.70 ±0.04
Rep1	4	0.76 ±0.03	0.62 ±0.05	0.71 ±0.04
Rep1	6	0.77 ±0.03	0.62 ±0.05	0.73 ±0.04
Rep1	8	0.78 ±0.03	0.61 ±0.06	0.74 ±0.04
Rep1	12	0.79 ±0.03	0.62 ±0.05	0.75 ±0.04
Rep1	24	0.81 ±0.03	0.64 ±0.05	0.77 ±0.03
Rep1	48	0.83 ±0.03	0.64 ±0.05	0.77 ±0.03
Rep1+smote	2	0.72 ±0.04	0.61 ±0.04	0.68 ±0.04
Rep1+smote	4	0.76 ±0.03	0.63 ±0.04	0.70 ±0.04
Rep1+smote	6	0.77 ±0.03	0.63 ±0.05	0.71 ±0.04
Rep1+smote	8	0.78 ±0.03	0.64 ±0.05	0.71 ±0.04
Rep1+smote	12	0.78 ±0.03	0.63 ±0.05	0.72 ±0.04
Rep1+smote	24	0.80 ±0.03	0.65 ±0.05	0.75 ±0.03
Rep1+smote	48	0.82 ±0.03	0.66 ±0.05	0.76 ±0.03
Rep2	2	0.71 ±0.04	0.63 ±0.05	0.68 ±0.04
Rep2	4	0.73 ±0.03	0.64 ±0.04	0.71 ±0.04
Rep2	6	0.75 ±0.03	0.64 ±0.04	0.72 ±0.04
Rep2	8	0.76 ±0.03	0.65 ±0.04	0.73 ±0.04
Rep2	12	0.78 ±0.02	0.65 ±0.04	0.75 ±0.03
Rep2	48	0.82 ±0.02	0.68 ±0.05	0.78 ±0.03
Rep2+smote	2	0.71 ±0.03	0.62 ±0.04	0.69 ±0.04
Rep2+smote	4	0.74 ±0.03	0.63 ±0.04	0.71 ±0.04
Rep2+smote	6	0.76 ±0.03	0.64 ±0.05	0.73 ±0.03
Rep2+smote	8	0.76 ±0.03	0.63 ±0.05	0.74 ±0.03
Rep2+smote	12	0.78 ±0.03	0.64 ±0.04	0.76 ±0.03
Rep2+smote	24	0.80 ±0.03	0.66 ±0.04	0.78 ±0.03
Rep2+smote	48	0.82 ±0.03	0.67 ±0.04	0.78 ±0.03

Table 2. Displays the AUROC and the standard deviation of the best performing model RF at 6 hours after admission and the scoring systems at 24 hours after admission.

Scoring System	AUROC	St. Deviation
RF at 6 hours	0.82	0.04
SAPS at 24 hour	0.650	0.012
APACHE at 24 hours	0.650	0.017
SOFA at 24 hours	0.623	0.013

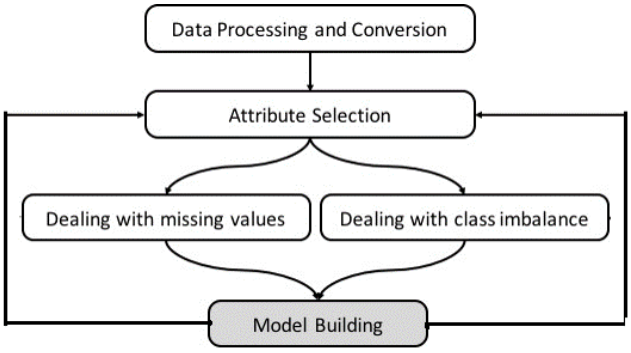

Figure 2. Proposed framework of an early mortality prediction model in the ICU

preparing the data for use, an extensive examination of data variables was conducted, which meant making a variety of choices and assumptions. Only patients with a single ICU stay at the age of 16 years old and above in Medical ICU (MICU), Surgical ICU (SICU) or Cardiac Surgery ICU (CSRU) are considered in the analysis; this cohort included 11,722 patients. Also patient mortality is defined as death inside the hospital.

The structure of data in the MIMIC II database had to undergo some initial preprocessing and conversion in order to prepare it for use in this study as shown on figure 2. We initially combined patient chart and lab test variables in one relational database in order to facilitate variable extraction. Data extraction was conducted in two stages as shown in figure 3. In the first stage, we extracted all variables for the entire patient ICU stay, then in the second stage we filtered the variables based on the required time window, which is the first 6 hours of a patient's admission. The final

process of attribute selection was based on 3 main criteria: (1) attribute coverage (measured for above 10% of the patients), (2) expertise of ICU consultants and (3) proposed variables from previous literature as shown on figure 3. We calculated both the coverage of each chart attribute and lab-test for patients within the first 6 hours to select those variables/ tests with high coverage. We only ignored attributes with coverage below 10%. This explains why some common variables in the literature might not be included in this study as they had low coverage in the first 6 hours of admission. In addition to the initial statistical experiments on the chart attributes and lab tests, direct consultation with subject matter experts in intensive care medicine, data proposed in previous work and data mining algorithms were also considered in attribute selection.

The following section discusses thoroughly which attributes are considered in this study. Finally, we extracted variable values, whether maximum and/or minimum values of each variable for each patient ICU stay within the specific required time window (6 hours after admission). It is important to note that there are several methods for selecting variable values. Each variable may have more than one value within the specified time window. For example, heart rate may have been measured 7 times within the first few hours of admission. In this case, the minimum and maximum heart rate values within the specified time window are both considered, as very low or very high heart rate values indicate severity. On the other hand, there are some variables that are one direction, such as Glasgow Coma Scale, in which only the minimum value of the variable indicates severity of illness; that's why the maximum value of the variable is ignored from our attribute selection. However, in the case of respiratory rate, for instance, only the maximum value is considered as it indicates a more critical patient condition than low respiratory rate. In summary, we used three strategies for value selection of the attributes: (1) minimum value, (2) maximum value or (3) minimum and maximum values.

Following attribute selection, we used two methods to handle missing values in data: (1) replacing missing values with the mean (Rep1) and (2) replacing missing values using EMImputation. The Synthetic Minority Oversampling Technique (SMOTE) Chawla and others (2002) was used to handle the issue of class imbalance. SMOTE is one of the most effective and widely used oversampling technique that was used by several work in literature Mi (2013); Bader-El-Den (2014); Lusa and others (2013); Wang and others (2006); Bader-El-Den and others (2018); Mohasseb and others (2018) to effectively handle the class imbalance problem. SMOTE increases the number of patient records who die inside the hospital (minority class) by inserting synthetic patient records. We employed RF, PART and BN algorithms to build the models. Building the models was done iteratively using different attribute selections as shown in figure 2, discussed thoroughly in the following section.

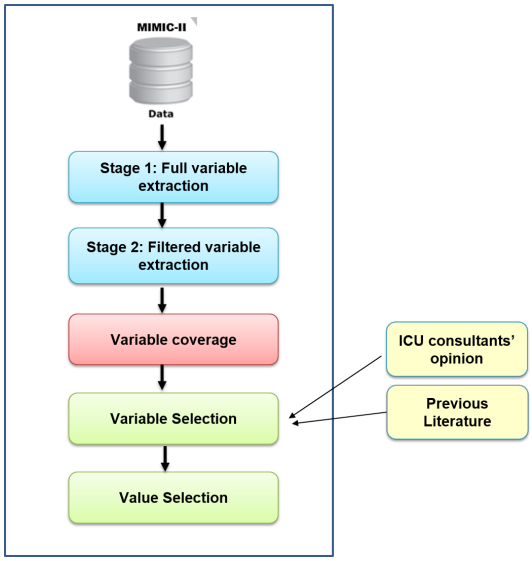

Figure 3. Data Extraction Process.

Selected Attributes

We selected 33 chart attributes and 25 lab-tests from the initially identified attributes with high coverage. Attributes with higher coverage were considered, resulting in a total of 20 unique variables (age, temperature, heart rate, respiratory rate, systolic blood pressure, arterial blood oxygen, Glasgow Coma Scale, creatinine, fractional inspired oxygen, serum urea nitrogen, potassium, sodium, hematocrit, White Blood Cells, blood clotting - INR, platelets count, bilirubin, AIDS, metastatic cancer and type of admission); 29 if we count maximums and minimums.

Results

This section presents the results for the top performing EMPICU DM models - EMPICU-RF, EMPICU-Decision Trees (DT), EMPICU-Naive Bayes (NB) and EMPICU-PART. It is important to note here that we have also evaluated a larger set of algorithms, such as Support Vector Machines (SVM) and JRip, however they were outperformed by the reported methods. Decision Trees are extremely fast at classifying unknown records. They are quite robust in the presence of noise. They also provide a clear indication of which fields are most important for prediction. The Naive Bayes algorithm affords fast, highly scalable model building and scoring. It scales linearly with the number of predictors and rows Hill and others (2006).

We conducted three different experiments. In the first experiment, we used all original 20 attributes and in the second we used only vital signs (age in addition to temperature, heart rate, respiratory rate, systolic blood pressure, arterial blood oxygen, Glasgow Coma Scale and creatinine). We evaluated each of the four data mining algorithms on each of six versions of the dataset mentioned earlier.

In the third experiment, we used filtered top 10 attributes that provide the highest information gain (IG) (i.e. those variables that contribute to better classification); we eliminated records missing any of the

Table 3. ranks top obtained results displayed in AUROC ± standard deviation for models developed using the most effective DM model - EMPICU-RF

	Experiment	AUROC
VS Attributes	Rep1+Smote	0.90 ± 0.01
Top 10 Attributes	Original	0.89 ± 0.02
Top 10 Attributes	Original+Smote	0.89 ± 0.02
Top 10 Attributes	Filter	0.87 ± 0.03
Top 10 Attributes	Filter+Smote	0.87 ± 0.03
All Attributes	Rep1+Smote	0.85 ± 0.02
All Attributes	Rep1	0.85 ± 0.01
All Attributes	Rep2	0.84 ± 0.02
All Attributes	Rep2+Smote	0.84 ± 0.02

top 10 attributes. The InfoGainAttributeEval algorithm in Weka Hall and others (2009) evaluates the worth of an attribute by measuring the information gain with respect to the class. We used four versions of the dataset:

1. dataset with eliminated records and the 20 unique variables (original),
2. dataset with eliminated records and the 20 unique variables and then applying smote (original+smote),
3. dataset with eliminated records and the top filtered ranked variables only (filtered top 10), and
4. dataset with eliminated records and the top filtered ranked variables only and then applying smote (filter+smote).

All experiments were done using Weka (version 3.7.13; University of Waikato, Hamilton, New Zealand). The results are noted in AUROC of the average of 10 runs, each run is 10-fold cross-validated. Table 3 ranks the experiments that showed the best performance (highest AUROC) using the best performing model, RF.

RESULTS' DISCUSSION

By referring to Figure 1 (graph a) , when comparing the performance of all algorithms on the Yes class (patients at risk of dying inside the hospital) per hour during the first 48 hours of ICU admission, we find that there is an abrupt improvement in performance at the 6th hour of ICU admission, after which the increase in performance is relatively smaller till the 48th hour of ICU admission. Similarly, the percentage of available values in the dataset increases dramatically at the 6th hour of ICU admission and continues to increase gradually till the 48th hour of ICU admission. In general, as time proceeds, the performance of the RF, BN and PART models increases as shown in Figure 1 (graph a) and table 1. As displayed in table 1, in general, both replacing the missing values with mean (rep1) and replacing missing values with EMImpuation (rep2) gave almost similar performance results. In addition, SMOTE oversampling technique hasn't enhanced the classification performance.

On the other hand, as shown in table 3, when comparing the performance of all three experiments in the proposed early mortality prediction models - EMPICU, in general applying the SMOTE oversampling technique significantly enhances the classification performance. Both replacing the missing values with mean (rep1) and replacing missing values with EMImpuation (rep2) gave almost similar performance results. In addition, we also find that when using the vital signs and filtered top10 attributes the prediction performance is better than when using all 20 unique attributes. In general in the filtered top10 experiment categories, the models developed with the original attributes (without any filtering) performed better than those with filtering. In the experiments without filtering, top 10 (original) and (original+smote) performed best (AUROC = 0.89 ± 0.02). As for the filtered experiments, top 10 (filter) and (filter+smote) also performed best (AUROC = 0.87 ± 0.03).

CONCLUSION

The Intensive Care Unit is an information rich environment, uniquely suited to data analysis. Several scoring systems and data mining methods have been developed to predict clinical deterioration and mortality in the ICU. However, most of these methods are designed for prediction after one or more days of admission. To our knowledge, there have been no definitive studies comparing mortality prediction per hour during the first 48 hours of a patient's admission in order to define to clinicians when is the ideal time for ICU data analysis. This paper aims to draw attention of the medical and data science communities to the importance of time-series analysis in the ICU taking into consideration the challenge of missing values in early patient data. The work in this research evaluated a wide range of data mining methods on 4000 patients (from MIMIC II database). We acknowledge the specific findings are particular to this database, but the methodology we have used is transferable. We intend to validate this work on the MIMIC-III Johnson and others (2016) database, which was released in August 2015, one year after this research project has started.

From a data mining perspective, the best performing model in this study is the EMPICU-RF, followed by EMPICU-BN and EMPICU-PART. In all experiments, EMPICU-RF performed significantly better than EMPICU-BN and EMPICU-PART(at a 5% confidence level). As mentioned earlier, other algorithms were tested, such DTs, SVM and JRip, however their performance was relatively poor. This finding supports work conducted by Ramon et al. Ramon and others (2007) which reported that AUROCs of a DT yielded smaller areas compared to a RF (DT, 65%; first order RF, 81%).

Our results shows that:

1. There is a sharp improvement in performance at the 6th hour of ICU admission, after which the increase in performance is relatively smaller till the 48th hour of ICU admission.
2. The percentage of missing values in the dataset drops dramatically at the 6th hour of ICU

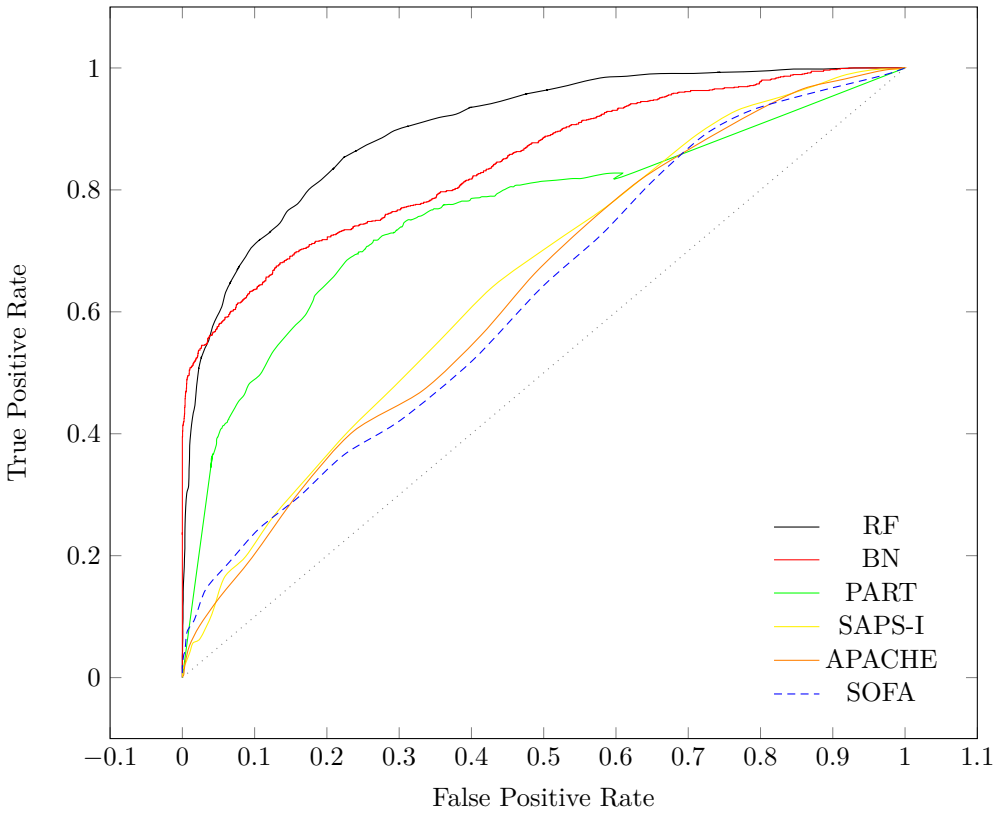

Figure 4. The performance of RF, BN and PART after 6 hours of ICU admission compared to SAPS-I, APACHE and SOFA on Rep1+smote dataset after 24 hours of ICU admission on the Yes class (patients at risk of dying inside the hospital)

admission and continues to decrease gradually till the 48th hour of ICU admission.

3. The discrimination power of the machine learning classification methods after 6 hours of admission outperformed the main scoring systems used in intensive care medicine (APACHE, SAPS-I and SOFA) after 24 hour of admission. The best performing classifier was RF, followed by BN, then PART on different experimental settings.
4. Both replacing the missing values with mean (rep1) and replacing missing values with EMImpuation (rep2) gave almost similar performance results.
5. SMOTE oversampling technique did not enhance the classification performance when the dataset was 4000 patients only, while it did enhance the classification performance with the larger dataset of 11,722 patients.

For clinicians, this research draws attention to the problem of missing values in variables over time in order to emphasize on the importance of collecting certain measurements early on; this will influence the predictive performance of mortality prediction models. Whilst we fully acknowledge that we have not developed a usable clinical tool in this work, we have shown that there exists rich information signal early in a critical care admission, which can provide guidance about likely individual outcome. We have shown this on a database with incomplete data. It is our view that this signal may in future be further strengthened by refinements to the methodology, which we have used, in order to assist both clinicians and patients in early outcome prediction.

Funding

A PhD scholarship from the Arab Academy for Science and Technology, Egypt.

References

Awad, Aya, Bader-El-Den, Mohamed and McNicholas, James. (2017a). Patient length of stay and mortality prediction: A survey. Health services management research 30(2), 105–120.

Awad, Aya, Bader-El-Den, Mohamed, McNicholas, James and Briggs, Jim. (2017b). Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. International Journal of Medical Informatics.

Bader-El-Den, Mohamed. (2014). Self-adaptive heterogeneous random forest. In: 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA). IEEE. pp. 640–646.

Bader-El-Den, Mohammed, Teitei, Eleman and Perry, Todd. (2018). Biased random forest for dealing with the class imbalance problem. IEEE transactions on neural networks and learning systems.

Berry, Michael J and Linoff, Gordon. (1997). Data mining techniques: for marketing, sales, and customer support. John Wiley & Sons, Inc.

Calvert, Jacob, Mao, Qingqing, Hoffman, Jana L, Jay, Melissa, Desautels, Thomas, Mohamadlou, Hamid, Chettipally, Uli and Das, Ritankar. (2016). Using electronic health record collected clinical variables to predict medical intensive care unit mortality. Annals of Medicine and Surgery 11, 52–57.

Celi, Leo Anthony, Galvin, Sean, Davidzon, Guido, Lee, Joon, Scott, Daniel and Mark, Roger. (2012). A database-driven decision support system: Customized mortality prediction. Journal of personalized medicine 2(4), 138–148.

Chawla, Nitesh V., Bowyer, Kevin W., Hall, Lawrence O. and Kegelmeyer, W. Philip. (2002). Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research, 321–357.

Citi, Luca and Barbieri, Riccardo. (2012). Physionet 2012 challenge: predicting mortality of icu patients using a cascaded svm-glm paradigm. In: Computing in Cardiology (CinC), 2012. IEEE. pp. 257–260.

Crawford, E David, Batuello, Joseph T, Snow, Peter, Gamito, Eduard J, McLeod, David G, Partin, Alan W, Stone, Nelson, Montie, James, Stock, Richard and Lynch, John. (2000). The use of artificial intelligence technology to predict lymph node spread in men with clinically localized prostate carcinoma. Cancer 88(9), 2105–2109.

Davoodi, Raheleh and Moradi, Mohammad Hassan. (2018). Mortality prediction in intensive care units (icus) using a deep rule-based fuzzy classifier. Journal of biomedical informatics 79, 48–59.

Delen, Dursun, Walker, Glenn and Kadam, Amit. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine 34(2), 113–127.

Gilani, Mahryar Taghavi, Razavi, Majid and Azad, Azadeh Mokhtari. (2014). A comparison of simplified acute physiology score ii, acute physiology and chronic health evaluation ii and acute physiology and chronic health evaluation iii scoring system in predicting mortality and length of stay at surgical intensive care unit. Nigerian Medical Journal 55(2), 144.

Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter and Witten, Ian H. (2009). The weka data mining software: an update. ACM SIGKDD explorations newsletter 11(1), 10–18.

Higgins, Thomas L, Teres, Daniel, Copes, Wayne S, Nathanson, Brian H, Stark, Maureen and Kramer, Andrew A. (2007). Assessing contemporary intensive care unit outcome: an updated mortality probability admission model (mpm0-iii). Critical care medicine 35(3), 827–835.

Hill, Thomas, Lewicki, Pawel and Lewicki, Paweł. (2006). Statistics: methods and applications: a comprehensive reference for science, industry, and data mining. StatSoft, Inc.

Hoogendoorn, Mark, el Hassouni, Ali, Mok, Kwongyen, Ghassemi, Marzyeh and Szolovits, Peter. (2016). Prediction using patient comparison vs. modeling: A case study for mortality prediction. In: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE. pp. 2464–2467.

Johnson, Alistair EW, Pollard, Tom J, Shen, Lu, Lehman, Li-wei H, Feng, Mengling, Ghassemi, Mohammad, Moody, Benjamin, Szolovits, Peter, Celi, Leo Anthony and Mark, Roger G. (2016). Mimic-iii, a freely accessible critical care database. Scientific data 3.

Kim, Sujin, Kim, Woojae and Park, Rae Woong. (2011). A comparison of intensive care unit mortality prediction models through the use of data mining techniques. Healthcare informatics research 17(4), 232–243.

Knaus, William A, Draper, Elizabeth A, Wagner, Douglas P and Zimmerman, Jack E. (1985). Apache ii: a severity of disease classification system. Critical care medicine 13(10), 818–829.

Knaus, William A, Wagner, Douglas P, Draper, Elizabeth A, Zimmerman, Jack E, Bergner, Marilyn, Bastos, Paulo G, Sirio, Carl A, Murphy, Donald J, Lotring, Ted and Damiano, Anne. (1991). The apache iii prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. Chest Journal 100(6), 1619–1636.

Kramer, Andrew A, Higgins, Thomas L and Zimmerman, Jack E. (2014). Comparison of the mortality probability admission model iii, national quality forum, and acute physiology and chronic health evaluation iv hospital mortality models: implications for national benchmarking. Critical care medicine 42(3), 544–553.

Le Gall, Jean-Roger, Lemeshow, Stanley and Saulnier, Fabienne. (1993). A new simplified acute physiology score (saps ii) based on a european/north american multicenter study. Jama 270(24), 2957–2963.

Le Gall, Jean-Roger, Loirat, Philippe, Alperovitch, Annick, Glaser, Paul, Granthil, Claude, Mathieu, Daniel, Mercier, Philippe, Thomas, Remi and Villers, Daniel. (1984). A simplified acute physiology score for icu patients. Critical care medicine 12(11), 975–977.

Le Gall, Jean R, Neumann, Anke, Hemery, François, Bleriot, Jean P, Fulgencio, Jean P, Garrigues, Bernard, Gouzes, Christian, Lepage, Eric, Moine, Pierre and Villers, Daniel. (2005). Mortality prediction using saps ii: an update for french intensive care units. Critical Care 9(6), R645.

Lee, Joon and Maslove, David M. (2017). Customization of a severity of illness score using local electronic medical record data. Journal of intensive care medicine 32(1), 38–47.

Lemeshow, Stanley, Teres, Daniel, Klar, Janelle, Avrunin, Jill Spitz, Gehlbach, Stephen H and Rapoport, John. (1993). Mortality probability models (mpm ii) based on an international cohort of intensive care unit patients. Jama 270(20), 2478–2486.

Lemeshow, Stanley, Teres, Daniel, Pastides, Harris, Avrunin, Jill Spitz and Steingrub, Jay S. (1985). A method for predicting survival and mortality of icu patients using objectively derived weights. Critical care medicine 13(7), 519–525.

Luo, Yuan, Xin, Yu, Joshi, Rohit, Celi, Leo and Szolovits, Peter. (2016). Predicting icu mortality risk by grouping temporal trends from a multivariate panel of physiologic measurements. In: Thirtieth AAAI Conference on

Artificial Intelligence.

Lusa, Lara and others. (2013). Smote for high-dimensional class-imbalanced data. BMC bioinformatics 14(1), 106.

Metnitz, Barbara, Schaden, Eva, Moreno, Rui, Le Gall, Jean-Roger, Bauer, Peter, Metnitz, Philipp GH and Group, ASDI Study. (2009). Austrian validation and customization of the saps 3 admission score. Intensive care medicine 35(4), 616–622.

Mi, Ying. (2013). Imbalanced classification based on active learning smote. Research Journal of Applied Science Engineering and Technology 5, 944–949.

Mohasseb, Alaa, Bader-El-Den, Mohamed and Cocea, Mihaela. (2018). Question categorization and classification using grammar based approach. Information Processing & Management 54(6), 1228–1243.

Moreno, Rui P, Metnitz, Philipp GH, Almeida, Eduardo, Jordan, Barbara, Bauer, Peter, Campos, Ricardo Abizanda, Iapichino, Gaetano, Edbrooke, David, Capuzzo, Maurizia and Le Gall, Jean-Roger. (2005). Saps 3—from evaluation of the patient to evaluation of the intensive care unit. part 2: Development of a prognostic model for hospital mortality at icu admission. Intensive care medicine 31(10), 1345–1355.

Perry, Todd, Bader-El-Den, Mohamed and Cooper, Steven. (2015). Imbalanced classification using genetically optimized cost sensitive classifiers. In: Evolutionary Computation (CEC), 2015 IEEE Congress on. IEEE. pp. 680–687.

Pirracchio, Romain, Petersen, Maya L, Carone, Marco, Rigon, Matthieu Resche, Chevret, Sylvie and van der Laan, Mark J. (2015). Mortality prediction in intensive care units with the super icu learner algorithm (sicula): a population-based study. The Lancet Respiratory Medicine 3(1), 42–52.

Poole, Daniele, Rossi, Carlotta, Latronico, Nicola, Rossi, Giancarlo, Finazzi, Stefano and Bertolini, Guido. (2012). Comparison between saps ii and saps 3 in predicting hospital mortality in a cohort of 103 italian icus. is new always better? Intensive care medicine 38(8), 1280–1288.

Ramon, Jan, Fierens, Daan, Güiza, Fabián, Meyfroidt, Geert, Blockeel, Hendrik, Bruynooghe, Maurice and Van Den Berghe, Greet. (2007). Mining data from intensive care patients. Advanced Engineering Informatics 21(3), 243–256.

Ribas, Vicent J, López, Jesús Caballero, Ruiz-Sanmartín, Adolf, Ruiz-Rodríguez, Juan Carlos, Rello, Jordi, Wojdel, Anna and Vellido, Alfredo. (2011). Severe sepsis mortality prediction with relevance vector machines. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE. pp. 100–103.

Rosenberg, Andrew L. (2002). Recent innovations in intensive care unit risk-prediction models. Current opinion in critical care 8(4), 321–330.

Sadeghi, Reza, Banerjee, Tanvi and Romine, William. (2018). Early hospital mortality prediction using vital signals. arXiv preprint arXiv:1803.06589.

Saeed, Mohammed, Villarroel, Mauricio, Reisner, Andrew T, Clifford, Gari, Lehman, Li-Wei, Moody, George, Heldt, Thomas, Kyaw, Tin H, Moody, Benjamin and Mark, Roger G. (2011). Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database. Critical care medicine 39(5), 952.

Vincent, Jean-Louis, De Mendonça, Arnaldo, Cantraine, Francis, Moreno, Rui, Takala, Jukka, Suter, Peter M, Sprung, Charles L, Colardyn, Francis and Blecher, Serge. (1998). Use of the sofa score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Critical care medicine 26(11), 1793–1800.

Vincent, J-L, Moreno, Rui, Takala, Jukka, Willatts, Sheila, De Mendonça, Arnaldo, Bruining, Hajo, Reinhart, CK, Suter, PeterM and Thijs, LG. (1996). The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive care medicine 22(7), 707–710.

Vincent, Jean-Louis and Singer, Mervyn. (2010). Critical care: advances and future perspectives. The Lancet 376(9749), 1354–1361.

Wang, Juanjuan, Xu, Mantao, Wang, Hui and Zhang, Jiwu. (2006). Classification of imbalanced data by using the smote algorithm and locally linear embedding. In: Signal Processing, 2006 8th International Conference on, Volume 3. IEEE.

Yakovlev, Alexey, Metsker, Oleg, Kovalchuk, Sergey and Bologova, Ekaterina. (2018). Prediction of in-hospital mortality and length of stay in acute coronary syndrome patients using machine-learning methods. Journal of the American College of Cardiology 71(11), A242.

Zimmerman, Jack E, Kramer, Andrew A, McNair, Douglas S and Malila, Fern M. (2006). Acute physiology and chronic health evaluation (apache) iv: hospital mortality assessment for today's critically ill patients. Critical care medicine 34(5), 1297–1310.