



# A Supervised Approach for Patient-Specific ICU Mortality Prediction Using Feature Modeling

Gokul S. Krishnan<sup>(✉)</sup>  and S. Sowmya Kamath 

Department of Information Technology,  
National Institute of Technology Karnataka, Surathkal 575025, India  
gsk1692@gmail.com, sowmyakamath@nitk.edu.in

**Abstract.** Intensive Care Units (ICUs) are one of the most essential, but expensive healthcare services provided in hospitals. Modern monitoring machines in critical care units continuously generate huge amount of data, which can be used for intelligent decision-making. Prediction of mortality risk of patients is one such predictive analytics application, which can assist hospitals and healthcare personnel in making informed decisions. Traditional scoring systems currently in use are parametric scoring methods which often suffer from low accuracy. In this paper, an empirical study on the effect of feature selection on the feature set of traditional scoring methods for modeling an optimal feature set to represent each patient's profile along with a supervised learning approach for ICU mortality prediction have been presented. Experimental evaluation of the proposed approach in comparison to standard severity scores like SAPS-II, SOFA and OASIS showed that the proposed model outperformed them by a margin of 12–16% in terms of prediction accuracy.

**Keywords:** ICU mortality · Machine learning · Healthcare analytics · Clinical decision support systems

## 1 Introduction

Over the past decade, Clinical Decision Support Systems (CDSSs) have revolutionized healthcare management systems due to their ability to provide proactive decision-making capabilities. Hospital systems generate crucial clinical information which can be used for improving clinical care delivery and other healthcare services. The data generated for the patients are large, heterogeneous and temporal in nature. CDSSs can help in leveraging these heterogeneous data for supporting intelligent applications like patient profiling, disease prediction and risk identification, etc. The improved and faster decision-making afforded by this, specifically in the case of critical-care patients like those admitted in ICUs, can be very significant and often life saving.

ICU mortality risk prediction is a popular CDSS application in use in modern ICUs that can help doctors and other healthcare professionals plan for patient-specific emergency situations and intervene when required. Besides measuring

the severity of a patient's condition, ICU mortality prediction also plays a crucial role in the assessment and evaluation of the treatment and healthcare policies of a hospital. Moreover, the optimal allocation and effective utilization of limited resources in ICUs can be better planned. Hence, CDSSs have been an active area of research, leading to the development of several ICU severity scoring and mortality prediction models. These models can be categorized into two classes - prediction systems based on parametric scoring and those based on non-parametric methods like Data Mining and Machine Learning (ML).

Scoring based prediction models typically use the perceived relevance of distinct clinical measurements of an ICU patient, as per a model defined by a group of experts. The most popular severity model, Acute Physiology and Chronic Health Evaluation (APACHE) [9], also called Acute Physiological Score (APS), is a physiological classification/scoring system which uses a patient's clinical data measurements to compute mortality score. The Simplified Acute Physiological Score (SAPS) scoring system [4] uses 14 easily measured biological and clinical variables to predict the possibility of patient death in ICUs. The SAPS score compared favorably with APS and was considered advantageous as the latter was quite complex and time-consuming. Newer versions of SAPS, named SAPS-II [11] and SAPS-III [12] also use daily clinical data for calculating mortality risk of ICU patients. Another score, Sequential Organ Failure Assessment (SOFA) [16], tracks the extent of organ failure in ICU patients. It is very simple to calculate as it is based on only six variables (Respiration, Central Nervous System, Cardiovascular, Renal, Coagulation and Comorbidity). Oxford Acute Severity of Illness Score (OASIS) [6] uses a subset of APACHE-IV [17] variables along with elective surgery, age and length of stay prior to ICU admission to predict ICU mortality.

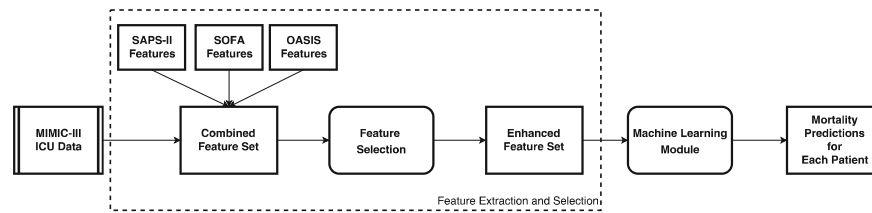
The second category of models include non-parametric methods using Data Mining and ML techniques for ICU mortality prediction. Dybowski et al. [3] developed one such system for predicting when to admit patients to the ICU and how long their stay would be, using Artificial Neural Networks (ANN) optimized by Genetic Algorithm. Ningaonkar et al. [13] proposed another ANN based model for ICU mortality prediction which performed better than the standard APACHE-III model. Celi et al. [2] proposed customized ML based mortality prediction models for various categories of ICU patients like patients suffering from acute kidney injury and other diseases to compare different ML models. Kim et al. [8] compared the predictive accuracy of ANN, SVM and decision trees with that of APACHE-III scoring systems to perform ICU mortality prediction, where the C5.0 decision tree algorithm outperformed ANN. The Super ICU Learner Algorithm (SICULA) [14] is based on a ML cascade that significantly outperformed SAPS-II and SOFA when applied to real-world hospital data. Calvert et al.'s model [1] predicted mortality risk within 12 h of patient's ICU admission, using Logistic Regression. The authors reported that system performed better against SAPS-II and SOFA.

Existing severity scores like APACHE, SAPS and SOFA are popular and used extensively in practice. But they compute mortality risk using selected clinical

features and based on human input. Even after several revisions over the years, recent studies have reported that the current version of SAPS and APACHE fall short in prediction accuracy [15], thus highlighting a significant scope for further improvement. In this paper, first, an empirical study on the effect of feature selection on feature variables of traditional severity scoring methods, to capture the latent relationships between various clinical factors that contribute to mortality risk of an ICU patient, is presented and then, an supervised learning approach for ICU mortality prediction is proposed. We consider features used by standard models, to which feature selection is applied for deriving the optimal feature set for each patient and then patient-specific mortality risk is predicted using ML classifiers. The rest of this paper is organized as follows: In Sect. 2, we describe the proposed approach for modeling patients using most relevant clinical features for improved accuracy. Section 3 presents the results of the experimental evaluation of the proposed methodology. We also benchmark its performance against that of existing severity scores on standard datasets using standard metrics, followed by conclusions and references.

## 2 Proposed Approach

The overall methodology adopted for the proposed empirical study is shown in Fig. 1.



**Fig. 1.** Process workflow

For the evaluation of the proposed methodology, the MIMIC-III<sup>1</sup> dataset [7] was used, which contains de-identified clinical data of more than 60,000 ICU patient admissions. During the first phase, all features or variables associated with the severity scores SAPS-II, SOFA and OASIS are extracted, all of which are considered initially as a feature set. In the next phase, a feature selection process was applied to the initial feature set to derive those most optimal features that best represent the characteristics of each individual patient. To measure the performance of this reduced feature set, a ML module was used for predicting patient-specific ICU mortality. The results obtained were compared with that of SAPS-II, SOFA and OASIS scoring systems. Being the most popularly used

<sup>1</sup> Medical Information Mart for Intensive Care, available online <https://mimic.physionet.org/>.

severity scores, SAPS-II and SOFA were chosen for the comparison, whereas OASIS was chosen as it is a recent score that is considered superior to even APACHE-III [10] as it uses lesser features to achieve similar performance [6].

The MIMIC-III dataset consists of de-identified health data of 46,520 critical care patients admitted to Beth Israel Deaconess Medical Center in Boston, USA, between 2001 and 2012. The data is quite extensive and hence, we designed a set of rules for selecting specific data. Using these rules, a subset of 32,622 patient records were selected for training and validation of the proposed approach. We list the defined cohort selection criteria below.

1. Only the records of adult patients were considered (age  $\geq 15$ ), as pediatric patients are treated with infants and adolescent-specific procedures.
2. To ensure that the prediction model had enough data to make predictions on, only those patient records, where the patient admitted in the ICU stayed there for at least one day (length of stay - los  $\geq 1$ ), were selected.

Another issue was the large number of missing values for clinical features in MIMIC-III. Handling missing data itself is an open research problem, and has been addressed by several researchers. Directly ignoring features with missing values adversely affected the number of patients selected for the study, hence we used a specific way for handling missing data. For the 32,622 patient records, those records with missing values were filled with the statistical median values of respective columns. Filling with median values ensures that the statistics of the model does not deviate from or be biased towards a particular clinical feature.

## 2.1 Feature Extraction and Selection

We implemented the SOFA, SAPS-II and OASIS scoring systems and generated the appropriate scores for each patient in the selected patient cohort. The feature set obtained after the feature extraction process is quite large as it essentially consists of all the features of SAPS-II, SOFA and OASIS. In addition to these, we also considered some new features like *patient demographics* (gender) and *first care unit* (type of ICU to which patient was first admitted to, e.g. surgical, medical, trauma, cardiac care etc). We also considered the ICD9<sup>2</sup> code of the first disease diagnosis for each patient. These two additional features, i.e., first care unit and ICD9 code, were considered on the basis that they can help determine the severity of the patient condition, which may contribute to the correct prediction of mortality risk. Elective surgeries are mostly less critical and hence were not considered as a feature and therefore ignored. With the inclusion of the standard severity score features and those we considered additionally, the final number of features obtained was 45.

To derive the most relevant features from this large set, we used a feature selection technique called Recursive Feature Elimination (RFE) [5]. RFE is a wrapper feature selection technique, which uses an estimator to repeatedly create a model based on a feature subset, and then prune features with low importance

<sup>2</sup> International Classification of Diseases, 9th Revision (ICD-9).



based on the accuracy of the estimation. The RFE algorithm is a 3-step process which is performed recursively. Firstly, the estimator is trained on the features by optimizing feature weights, next, the features are ranked based on the cost function, based on which finally, the features with the least rank are removed during each iteration. These steps are performed for a number of iterations till an optimal features set with the required number of features remain. We used the RFE algorithm with Logistic Regression (using L2 regularization and cross-entropy loss function) as the estimator, which uses the categorical variable (mortality prediction in this case) as the dependent variable. Logistic Regression being a statistical probabilistic prediction model based on features, works well for binary classification and hence is well-suited for selecting features that contribute most to the mortality prediction. Algorithm 1 depicts the process in detail.

---

**Algorithm 1** Optimal Feature Subset Selection using RFE

---

*Input:* The whole set of features and mortality labels and  $n$ , the required number of features to be selected.

*Output:* Optimal subset of  $n$  features.

- 1: Set the whole set of features as the current feature set
  - 2: **while** current feature set size  $> n$ , **do**
  - 3:   Train the Logistic Regression model with the set of features and corresponding labels, thereby tuning feature weights
  - 4:   Calculate cross-entropy loss function values of the Logistic Regression model for the current feature set
  - 5:   Rank the set of features based on cross-entropy loss values
  - 6:   Remove feature with the least rank from the current set
  - 7: The generated optimal subset of  $n$  features, can now be fed into a classifier model
- 

After the feature selection process, only 8 features were selected as the most relevant out of the original feature set of 45. These included scores that capture the correct functioning of the liver, renal output, cardiovascular activity, age score, blood urea nitrogen (BUN) score, sodium score, comorbidity score and first care unit. Interestingly, features were selected only from SAPS-II and SOFA and the additional variable, first care unit. Intuitively, this makes very good sense, as the first care unit does play an important role in determining the severity or mortality risk of the patients newly admitted. The score mostly says how critical the patient is, and hence contributes a lot of information towards predicting mortality. For predicting the mortality, a label called the expire flag is used, which is binary in nature (i.e., 0 for alive and 1 for expired). The final set of optimal features along with the labels are then fed into the ML module for obtaining patient-specific ICU mortality prediction.

## 2.2 Supervised Learning Process

The optimal feature set can now be used as a representation of patient-specific clinical data, based on which their personalized profile can be modeled. The

optimal feature set along with the associated mortality labels in MIMIC-III data were used for training a suite of machine learning classifiers (the ML module), for mortality risk predictions namely, Naive Bayes (NB) (Gaussian, Multinomial and Bernoulli), Support Vector Machine (SVM-linear kernel), Decision Tree (CART - Classification and Regression Trees) and Random Forest classifiers.

NB Classifiers are probability based classifiers which work based on application of Bayes' Theorem with an assumption that features are independent. SVM classifies data points into potential classes (alive and expired patients in this case) by representing them in vector space. The SVM algorithm constructs a hyperplane or a set of hyperplanes that divide the space so that data is classified into potential classes. The decision trees classifier employs a tree-like structure where the internal nodes represent the features and the leaf nodes represent the labels (the mortality labels - alive(0) or expired(1) in this case). Various branches represent various feature values leading to corresponding mortality labels, which are used to perform classification. Finally, the random forest classifier is an ensemble classifier which works by generating numerous decision trees the votes of which are used to predict. The mortality class which is voted or predicted by most decision trees decides the prediction of the random forest classifier.

The patient representations modeled as per the proposed feature extraction and selection approach are used as training data and are fed into the various classifiers for observing mortality prediction performance. We used 10-fold cross validation where we used 75% of the data for training and the remaining 25% was kept aside for testing/validation. We used standard metrics like average accuracy, average precision, average F-score and average Area Under Receiver Operating Characteristic Curve (AUROC) for evaluating the performance of the trained model. In Sect. 3, we discuss the results of these classification experiments.

### 3 Experimental Results

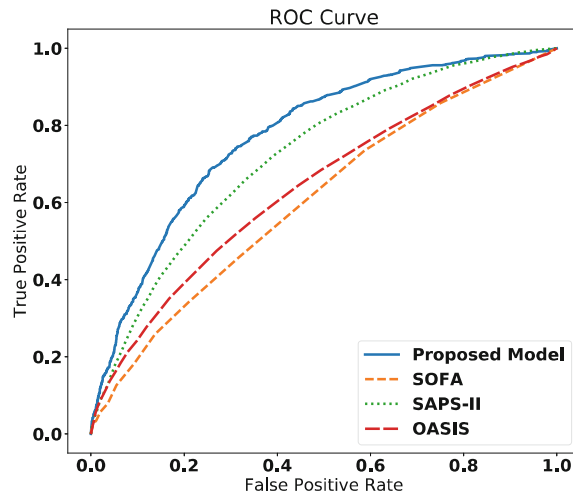
For validating the proposed approach, both the proposed model and the standard severity models (SAPS-II, SOFA and OASIS) were applied to the MIMIC-III subset of 32,622 patients. The generated SOFA, SAPS-II and OASIS scores of patients were used to predict ICU mortality [6, 14] and their performance was measured as well. The experiments were performed on a workstation running Ubuntu 17.04 with 3.5 GHz Intel Core i7 Processor, 16GB RAM and 2TB Hard Drive. The proposed prediction model was developed in Python and packages like Pandas, Scikit-Learn and Matplotlib were used for performing dataset operations, Machine Learning algorithms and plotting ROC curves.

Our observations on the performance of the various ML models and the standard severity scores (SAPS-II, SOFA and OASIS) on the MIMIC-III dataset are tabulated in Table 1. To the best of our knowledge, there is no published work yet on the application of these three severity scores on MIMIC-III. From the values of the various metrics, it can be seen that the Random Forest classifier performed best at an average accuracy of 0.71, average AUROC of 0.77, average precision of 0.71 and average F-score of 0.71, while SVM and Decision Tree classifiers were a close second. Random Forests being an ensemble classifier, predicts

a label based on the voting of multiple decision trees, hence performed the best. It can be observed that, among the standard severity scoring systems, SAPS-II outperformed SOFA and OASIS. However, the proposed ML based model with Random Forests classifier achieved significant improvement over SAPS-II in all metrics. A plot of the Receiver Operating Characteristic (ROC) curve is shown in Fig. 2, which highlights the superior performance of the proposed model over existing severity scores for MIMIC-III data.

**Table 1.** Benchmarking ML Module Classifiers against Standard severity scores SAPS-II, SOFA and OASIS

Classifier	Accuracy	AUROC	Precision	F-score
Gaussian NB	0.69	0.75	0.68	0.68
Multinomial NB	0.69	0.68	0.68	0.67
Bernoulli NB	0.68	0.71	0.69	0.68
Decision Tree	0.70	0.73	0.70	0.69
SVM	0.70	0.76	0.70	0.69
Random Forest	0.71	0.77	0.71	0.71
SAPS-II	0.63	0.72	0.65	0.57
SOFA	0.62	0.61	0.60	0.57
OASIS	0.61	0.64	0.67	0.50



**Fig. 2.** Area under ROC Curve for various models (The best performing model (Random Forest) taken as Proposed Model)

## 4 Conclusion and Future Work

In this paper, an empirical study on the effect of feature selection to select the optimal feature subset contributing to ICU mortality risk prediction, was discussed. Recursive Feature Elimination with Logistic Regression was used for deriving the optimal feature set from combined features of SAPS-II, SOFA and OASIS scoring models. This was used to train a suite of ML classifiers to predict patient-specific ICU mortality. Among the classifiers used, Random Forest achieved the best accuracy of 0.71 and average AUROC of 0.77. Its performance when compared to that of SAPS-II, SOFA and OASIS, was significantly higher by a factor of 12–16% in terms of accuracy. Benchmarking the proposed model against other standard scoring systems like APACHE III, APACHE IV and SAPS-3 will be explored next. Other feature engineering techniques can be explored with the possible inclusion of a better mathematical modeling for further improvements in the prediction model.

**Acknowledgements.** We gratefully acknowledge the use of the facilities at the Department of Information Technology, NITK Surathkal, funded by Govt. of India's DST-SERB Early Career Research Grant (ECR/2017/001056) to the second author.

## References

1. Calvert, J., et al.: Using ehr collected clinical variables to predict medical intensive care unit mortality. *Ann. Med. Surg.* **11**, 52–57 (2016)
2. Celi, L.A., et al.: A database-driven decision support system: customized mortality prediction. *J. Personalized Med.* **2**(4), 138–148 (2012)
3. Dybowski, R., et al.: Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet* **347**, 1146–1150 (1996)
4. Gall, L., et al.: A simplified acute physiology score for ICU patients. *Crit. Care Med.* **12**(11), 975–977 (1984)
5. Guyon, I., et al.: Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**(1), 389–422 (2002)
6. Johnson, A.E.W., et al.: A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy. *Crit. Care Med.* **41**(7), 1711–1718 (2013)
7. Johnson, A.E.W., et al.: MIMIC-III, a freely accessible critical care database. *Sci. Data* **3**, 160035 (2016)
8. Kim, S., et al.: A comparison of ICU mortality prediction models through the use of data mining techniques. *Healthc. Inform. Res.* **17**(4), 232–243 (2011)
9. Knaus, W.A., et al.: APACHE—a physiologically based classification system. *Crit. Care Med.* **9**(8), 591–597 (1981)
10. Knaus, W.A., et al.: The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults. *Chest* **100**(6), 1619–1636 (1991)
11. Le Gall, J., et al.: A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *Jama* **270**(24), 2957–2963 (1993)
12. Moreno, R.P., et al.: SAPS 3—from evaluation of the patient to evaluation of the intensive care unit. *Intensive Care Med.* **31**(10), 1345–1355 (2005)

13. Nimgaonkar, A., et al.: Prediction of mortality in an Indian ICU. *Intensive Care Med.* **30**(2), 248–253 (2004)
14. Pirracchio, R., et al.: Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir. Med.* **3**(1), 42–52 (2015)
15. Salluh, J.I.F., et al.: ICU severity of illness scores: APACHE, SAPS and MPM. *Curr. Opin. Crit. care* **20**(5), 557–565 (2014)
16. Vincent, J.L., et al.: The SOFA score to describe organ dysfunction/failure. *Intensive Care Med.* **22**(7), 707–710 (1996)
17. Zimmerman, J.E., et al.: Acute physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit. Care Med.* **34**(5), 1297–1310 (2006)