Early Hospital Mortality Prediction using Vital Signals

Reza Sadeghi[a,*], **Tanvi Banerjee**[a], and **William Romine**[b]

[a]Department of Computer Science and Engineering, Kno.e.sis Research Center, Wright State University, Dayton, OH, USA

[b]Department of Biological Sciences, Wright State University, Dayton, OH, USA

Abstract

Early hospital mortality prediction is critical as intensivists strive to make efficient medical decisions about the severely ill patients staying in intensive care units (ICUs). As a result, various methods have been developed to address this problem based on clinical records. However, some of the laboratory test results are time-consuming and need to be processed. In this paper, we propose a novel method to predict mortality using features extracted from the heart signals of patients within the first hour of ICU admission. In order to predict the risk, quantitative features have been computed based on the heart rate signals of ICU patients suffering cardiovascular diseases. Each signal is described in terms of 12 statistical and signal-based features. The extracted features are fed into eight classifiers: decision tree, linear discriminant, logistic regression, support vector machine (SVM), random forest, boosted trees, Gaussian SVM, and K-nearest neighborhood (K-NN). To derive insight into the performance of the proposed method, several experiments have been conducted using the well-known clinical dataset named Medical Information Mart for Intensive Care III (MIMIC-III). The experimental results demonstrate the capability of the proposed method in terms of precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC). The decision tree classifier satisfies both accuracy and interpretability better than the other classifiers, producing an F1-score and AUC equal to 0.91 and 0.93, respectively. It indicates that heart rate signals can be used for predicting mortality in patients in the care units especially coronary care units (CCUs), achieving a comparable performance with existing predictions that rely on high dimensional features from clinical records which need to be processed and may contain missing information.

Keywords

intensive care; mortality prediction; statistical and signal-based features

*Corresponding author: sadeghi.2@wright.edu (Reza Sadeghi), tanvi.banerjee@wright.edu (Tanvi Banerjee), william.romine@wright.edu (William Romine).

1. Introduction

Intensive care unit (ICU) is a ward in hospital, where seriously ill patients are cared for by specially trained staff. Quick and accurate decisions for the patients are needed. As a result, a wide range of decision support systems have been deployed to aid intensivists for prioritizing the patients who have a high risk of mortality.

Most mortality prediction systems are considered as score-based models [1] [2] [3] [4] which appraise disease severity to predict an outcome. These models utilize patient demographics and physiological variables such as age, temperature, and heart rate collected within the initial 12 to 24 hours after ICU admission with the aim of assessing ICU performance. The score-based models employ certain features that sometimes are not available at ICU admission. Also, they make decisions according to a collection of data after at least first 12 hours of ICU admission. To enhance the proficiency, the customized models refine the score-based models for usage within specific conditions. For instance, [5] introduces a model to predict the risk of mortality due to cardiorespiratory arrest. Although these models provide adequate results, the ICU patients are varied and subjected to multiple diseases. Therefore, selecting the right model for a special patient who is immediately admitted to ICU is difficult. On the other hand, various studies [6][7][8][9] [10] express the superiority of data mining techniques over traditional score-based models. The data mining models have exerted different techniques such as random forest [6] [7], support vector machine [8], decision tree [9], and deep learning [10] [11] [12] [13]. Furthermore, some of the methods like [14] engage a pipeline of data mining techniques to predict the risk of mortality. These methods are organized based on certain clinical records which are collected in initial hours after ICU admission. However, laboratory test results need to be processed and many clinical records contain missing values [15]. While vital signals can provide numerous information which has been proven to possess strong relation with the mortality [16]. Therefore, vital signal fluctuations can provide high capability to predict the mortality risk more accurately and faster than clinical-based methods.

The main goal of this paper is to provide an early mortality prediction of patients based on their first hour after ICU admission according to their heart rate signals. Our study relies on the Medical Information Mart for Intensive Care III, MIMIC-III Waveform Database records [17]. We propose a method to extract both statistical and signal-based features from the heart signals and employ well-known classifiers such as logistic regression and decision tree to predict hospital mortality, i.e. death inside the hospital.

The rest of the paper is organized as follows: Section 2 presents a literature review on the related studies. Section 3 describes the proposed method in four subsections of data description, signal preprocessing, feature extraction, and classification. To evaluate the performance of the proposed method, Section 4 is allocated to the experiments and discussions. Finally, Section 5 summarizes the conclusion and future work.

2. Related Work

There is an increasing interest in addressing early hospital mortality prediction. The proposed systems can be categorized into three classes of score-based, customized, and data mining models.

Various score-based approaches such as acute physiology and chronic health evaluation (APACHE) [4], simplified acute physiology score (SAPS) [3], and quick sepsis-related organ failure assessment score (qSOFA) [2] have been proposed. APACHE score is the best-known and widely used in intensive cares [18]. The original APACHE score [19] employed 34 physiological measures from initial 24 hours after ICU admission to determine the chronic health status of the patients. [4] introduced the APACHE II scoring model including a reduction in the number of variables to 12 routine physiological measurements, along with the age of patients. Extending that, the APACHE III improved the effectiveness of mortality prediction by adding new variables such as race, length of stay in ICU, and prior place before ICU. APACHE IV also endeavored to enhance the over prediction problem of the APACHE III by adding new variables and using the weights utilized in APACHE III [20]. The traditional severity of illness score-based models commonly attempted to predict based on either specific age ranges, or information recorded within the first 24 hours of ICU admission [21]. Furthermore, they utilized features which are not always available at the time of ICU admission. For instance, the APACHE IV applied its analysis on over 100 variables like chronic health variables of AIDS, cirrhosis, hepatic failure, immunosuppression which may not be recorded at the time of admission.

The customized models make a decision according to the characteristics of either specific health problems such as cardiorespiratory arrest [5] and early severe sepsis [22], or specific geographical areas such as France [23] or Australia [24]. For instance, Le Gall and coworkers [23] customized the SAPS II model based on the French patients' characteristics. They used the logit of the original SAPS II model and computed the coefficients according to the data. Furthermore, they tried to expand the second version of SAPS by adding six variables (age, sex, length of hospital stay before ICU admission, and the patient's location before ICU) that are potentially associated with mortality. Although these models provide adequate results, most ICU patients are elderly people over 65 years [25] who are faced with multiple ailments. Also, selecting the right model is challenging due to the variety of patients who are immediately admitted to ICU. Moreover, the models for specific geographical areas are not extendable for other cases.

The third class of methods employ data mining techniques to forecast mortality. For instance, [6] devised a method based on random forest and the synthetic minority over-sampling technique. In another method, Venugopalan et. al [14] used a pipeline of logistic regression, neural network, and conditional random forest. The three categories of demographic, lab, and chart data such as gender, age, height, sodium, creatinine, and heart rate have been fed to logistic regression, neural network, and conditional random forest, respectively. These methods focus on using clinical records instead of waveform data while in practice, many clinical records such as laboratory test results need to be processed which could delay the clinical decision support process.

To address these issues, we propose a method for early mortality prediction of patients based on the first hour after IGU admission using heart rate signals. To the best of our knowledge, this paper is the first work which utilizes only heart signals for early mortality prediction using the MIMIC-III dataset. We describe each signal in terms of 12 statistical and signal-based features which are fed into multiple transparent and non-transparent classifiers.

3. Methodology

This section presents a novel method which utilizes statistical and signal-based features with the purpose of fast and accurate early hospital mortality prediction. Subsection 3.1 provides a review on the MIMIC-III clinical dataset while subsections 3.2 and 3.3 describe signal preprocessing and feature extraction, respectively. Ultimately, subsection 3.4 presents an overview on the descriptive classifiers employed to predict whether a patient survives or passes away based on the characteristics of their EGG signal.

3.1. Data Description

This study is conducted over the well-known MIMIC-III database comprising the records of 46520 patients who stayed in critical care units. Due to the de-identification process, there are only 10282 patients whose the clinical data in the MIMIC-III are associated with the related vital signals in the Matched Subset. As shown in the Figure 1, the age distributions of the whole MIMIC-III (without infants) and the Matched Subset are similar. Hence, the outcomes of the Matched Subset can be extended to the whole database. It is worth mentioning that due to the de-identification process, all the patients greater than or equal to 90 years of age are assigned to one group.

Also, the hospital wards for patients throughout their hospital stay have been reported via the transfers table in the clinical dataset. Indeed, it specifies which of the care units described in Table 1 have been allocated to each patient in a certain time. Since nearly 90 percent of patients in the Matched Subset suffer from cardiovascular diseases, we have focused on predicting the risk of mortality among patients who stayed in coronary care unit (GGU) in this study. CCU is an ICU that takes patients with cardiac conditions required continuous monitoring and treatment.

3.2. Signal Preprocessing

The recorded physiological signals are always accompanied with noise due to different recording systems. The MIMIC-III database is extracted from the CareVue and MetaVision clinical information systems provided by Philips and iMDSoft, respectively [17]. After extracting the data, we truncated the tails which contain only zeros or undefined values. Following this, we replaced the missing values with the previous known ones. Finally, the smoothed version of heart rate signal, $S'(t)$, was computed according to the moving average filter with one-hour windows size ρ in the form of Equation 1.

$$S'(t) = \begin{cases} \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} S(t) & \rho > = T > = 1 \\[2ex] \dfrac{1}{\rho}\displaystyle\sum_{t=T}^{T-\rho+1} S(t) & L - \rho > = T > = \rho \\[2ex] \dfrac{1}{T}\displaystyle\sum_{t=L-\rho+1}^{L} S(t) & L - \rho + 1 > = T > = L \end{cases} \quad (1)$$

where the original signal $S(t)$ contains L samples. On the other hand, the heart signals were recorded with different lengths and sampling rates. For instance, the sampling rate of the heart rate (HR) signals are varied from 1 to 0.17 Hz in MIMIC-III database. To avoid biased comparison among signals due to the different sampling rates and lengths, the anti-aliasing finite impulse response (FIR) low-pass filter [26] was performed over the low sampling rate signals. Indeed, a linear-phase FIR filter interpolates new samples to resample the signals with a lower sampling rate. For instance, as shown in Figure 2 the noise samples have been removed by applying the moving average over the original signal. Then, the oversampling method increases the frequency of the heart rate signal to $1\,Hz$, leading to increasing the number of samples from 9021 to 541310[5].

3.3. Feature Extraction

In order to predict the risk of mortality after the first hour of ICU admission, quantitative features have been computed based on the HR signals. Each signal is described in terms of 12 statistical and signal-based features which were extracted from the patient's ECG signal. The statistical features reveal useful information about the distributions of the processed data described in the subsection 3.2. Signal preprocessing. Maximum, minimum, and range can demonstrate the spectrum in which the distribution lies. The skewness indicates whether the distribution is symmetric or skewed. The kurtosis measures the thickness of the tails of the distribution and the standard deviation shows how the data samples scatter around the mean. Table 2 indicates the average of each feature for both passed away and living patients. The reported values indicate the capability of these features in segregating the two groups of patients based on the proposed statistical and signal-based features.

The signal-based features in this study fall into two different groups of averaged power and power spectral density [27]. The averaged power of a hnite discrete-time signal is defined as the mean of the signal's energy. The averaged power of a discrete-time signal $S[n]$ is computed as:

$$\bar{P} = \frac{E}{n_2 - n_1 + 1} = \frac{1}{n_2 - n_1 + 1}\sum_{n_1}^{n_2} S[n]^2 \quad (2)$$

where n_1 and n_2 are the first and last samples, respectively. The signal power is computed by taking the integral of the power spectral density (PSD) of a signal over the entire frequency space. The PSD is the Fourier transform of the biased estimate of the autocorrelation sequence. The PSD of the signal $S[n]$ with sampling rate ρ, in the interval T can be computed as follows:

$$\bar{P} = \frac{\Delta T}{N} \mid \sum_{n=0}^{N-1} S[n] e^{-i2\pi\rho} \mid \quad (3)$$

3.4. Classification

In the MIMIC III dataset, the number of patients who passed away inside the hospital is relatively small in comparison with the number of patients who survived, meaning the dataset is imbalanced. The ratio of physiological signals pointing to the passed away patient in contrast to those who survive is equal to 7.03. Thus, the early mortality prediction systems are faced with an imbalanced dataset. To handle this issue, a wide range of techniques such as resampling [6], cost sensitive classifiers [28], and one-class classifiers [29] [30] have been proposed. Resampling methods make no assumptions about the distribution of samples and therefore, they can be applicable to any classification problem. Also, they are less sensitive to outliers than other techniques. In this study, we utilize a resampling method called adaptive semi-unsupervised weighted oversampling (A-SUWO) [31] to balance the dataset.

The 10-fold cross-validation strategy was used to evaluate the performance of classifiers on the same dataset. In this way, samples are arbitrarily divided into ten disjoint sections. In ten iterations, nine folds shape a group of samples used to train classifiers. Furthermore, the remaining one is utilized to test the learning process. The mean of learning rates determines the performance of the methods in segregation of classes.

In this study, two categories of classifiers are examined: transparent or interpretable models, and non-transparent or black-box models. Transparent classifiers such as decision tree, linear discriminant, logistic regression, and support vector machine (SVM) using the linear kernel explain hidden clinical implications and integrate background knowledge into analysis. Also, they are not only easy to interpret and fast, but also need small memory in practice. On the other hand, non-transparent classifiers like random forest, K-NN, boosted tree, and Gaussian SVM are black-box methods which frequently provide adequate classification results. However, these non-transparent classifiers suffer from lack of easily-comprehensible descriptions for the relations between input and output variables.

4. Experiments and Results

In these experiments, a retrospective analysis on patients who stayed in CCU was performed using the information recorded in from the MIMIC-III Waveform Database Matched Subset. This dataset contains the records of 365 patients who passed away while staying at CCU and 2614 patients successfully discharged. As mentioned above, the effect of noise samples was reduced by smoothing the heart rate signals using the averaged smoothing hlter. Also,

resampling of low-sampled signals was used to have a fair comparison. Eventually, the combination of statistical and signal-based features after normalization was fed to several interpretable and non-transparent classifiers which are easy to interpret and statistically powerful, respectively.

Four transparent classifiers: decision tree, linear discriminant, logistic regression, and support vector machine (SVM) were examined. The decision tree was implemented based on a CART tree algorithm [32] with Gini's diversity index (*GDI*) as a split criterion. This splitting criterion is one of the most popular impurity measurements which not only performs similar to information gain in most cases [33], but also has lower computational complexity as a result of avoiding use of the logarithm. The Gini index in the form of Equation 4 is utilized to select the next feature at each node of the tree for splitting the data.

$$GDI = 1 - \sum_i (p(i))^2 \quad (4)$$

where $p(i)$ is the observed fraction of samples in the node, which are labeled as *i*. Therefore, the *GDI* equal to zero points out to a pure node which contains samples of one class. On the other hand, the *GDI* for binary classification is equal to 0.5 at most when a node contains samples of both classes with identical numbers. Furthermore, the linear SVM working based on dot product kernel is a simple linear classifier. As a result, this version of SVM is both easy to be interpreted and fast in prediction.

Regarding to the non-transparent classifiers, four black-box methods of random forest, boosted trees, Gaussian SVM, and K-nearest neighborhood (K-NN) are employed. The random forest and boosted trees utilize 60 decision tree learners according to the bootstrap aggregating [34] and adaptive boosting [35] ensemble methods, respectively. Moreover, the Gaussian SVM uses radial basis function kernel and K-NN exerts the K equal to 100. All the experiments are implemented in MATLAB 9.2.0.538062(*R*2017*a*) on the same machine with an Intel processor 2.50 GHz with 8 GB RAM

4.1. Results

The outputs of classifiers can be summarized in four groups: the patients who are truly diagnosed as passed away (TP), the people who are incorrectly labeled as passed away (FP), the records correctly detected as information belonging to survived patients (TN), and finally the ones incorrectly assigned as living patients (FN). These four groups can be aggregated in different ways.

Equation 5 indicates the precision metric as the fraction of patients who have been truly diagnosed as passed away over all the patients predicted as having passed away. Indeed, the larger number of patients incorrectly predicted as passed away leads to the lower precision for the classifier. Moreover, to see the ability of the classification method in predicting all passed-away patients, we utilize the recall metric presented in Equation 6. In other words,

this metric presents the fraction of the patients who are correctly predicted as passed-away over the whole number of passed-away patients.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

It is worth mentioning that all samples being assigned to positive group lead to high recall and low precision. Then, the harmonic average of precision and recall called F1-score is also considered. Indeed, F1-score described in Equation 7 calculates the quality of classification for both passed away and living patients, simultaneously.

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (7)$$

As shown in the Table 3, the decision tree outperforms all transparent classifiers which are easily interpretable and provide some clinical insights into the classification process. Also, the values for F1-score among the transparent classifiers demonstrate a big gap between the decision tree and the others. The F1-score of linear discriminant, linear SVM, and logistic regression is near to 0.71 while the decision tree results in 0.91. The linear discriminant assumes that different groups of data are generated based on different Gaussian distributions. However, the amounts of Skewness and Kurtosis of both passed away and surviving patients are not equal to zero (table 2) which indicates non-Gaussian distribution for the both groups of patients. This is the likely reason why the linear discriminant results in low performance. In addition, weak performance of the logistic regression and linear SVM may indicate that the data are not linearly separable. Furthermore, the performance of these supervised methods is similar to the results of the other empirical comparisons such as [36] describing that random forest can outperform other classifiers like SVM and K-NN in certain conditions.

From another point of view, all interpretable classifiers except the decision tree have lower recall (near 0.65) rather than their precision. However, the decision tree has both high precision and recall that shows not only most of the passed-away patients have been correctly recognized but also most of the predicted passed-away patients are correctly assigned to the correct category. As expected, most of the non-transparent classifiers achieve higher performance in comparison to the interpretable classifiers. In addition, random forest comprising several decision tree learners performs better than the other black boxe methods. The interesting point is that the decision tree exceeds many of the non-transparent classifiers including K-NN and boosted tree.

Decision support systems are required to be accurate and robust; however, they also should be interpretable, transparent, and capable of integrating clinical background knowledge into the analysis. Hence, we focus on transparent classifiers and scrutinize their performance in different thresholds. Figure 3 demonstrates that the decision tree outperforms the other transparent classifiers in terms of AUG. Furthermore, the linear SVM, logistic regression, and linear discriminant have similar performance even on different thresholds, which lie lower than the AUG of the decision tree.

Referring to the ROG curve of black-box methods plotted by solid lines in Figure 3, random forest has the best performance in comparison to Gaussian SVM, boosted trees, and K-NN. Moreover, the curves indicate that random forest and Gaussian SVM have a homogeneous ratio of true positive rate over false positive rate. Furthermore, the ROG curve of decision tree represents the outperformance of this transparent classifier over two black-box methods of K-NN and boosted trees.

The results reveal that the most non-transparent classifiers achieve higher discrimination power while they failed to provide adequate explanations about how the classification results are derived. On the other hand, the interpretable classifiers often attempt to create a decision boundary using the value of linear combination of the sample features. However, most real samples originate from a complex system such as human body. Hence, the decision tree may provide the best choice as a tradeoff between transparency and accuracy. The decision tree discovers knowledge which can be expressed in a readable form while its classification performance is comparable with other methods, even popular non-transparent classifiers.

4.2. Discussion

In order to interpret the decision tree qualitatively, Figure 4 illustrates the best trained structural model of this classifier gained in the experiments. The tree model hierarchically separates data according to the features leading to a more stable and pure tree. For instance, the left-most child of the decision tree displayed by green star contains records from class 1 (survived patients). The highlighted path shows records which satisfy the three rules shown in the graph. The first rule divides samples according to the amount of energy spectral density computed for each record. The samples with energy spectral density lower than −0.85 are passed to the decision Node 2 which provides a rule for the amount of Skewness of signals. Node 4 then filters the samples with value of Maximum less than −0.83 which will be assigned to the green star node.

The averaged power, the 11th feature, has been selected as the root of the tree with the highest Gini index. This result shows that using the averaged power features may be promising in early mortality predictions. To further scrutinize the effects of individual features in the decision tree, the estimate of predictor importance is computed. It sums up changes in the risk caused by splits on every independent variable and divides the total result by the number of branch nodes (the tree nodes without any children). Indeed, this sum is taken over the best splits found at each branch node. The importance of features according to this separation is computed as the difference between the risk for the parent node and the sum of risks for its children.

The risk of splitting for each node is composed of the impurity measurement and the node probability. As explained before, we employed the Gini index as the impurity measurement which has less computational complexity in comparison to the information gain. Also, node probability is defined as the number of records reaching the node, divided by the total number of records. Then, the risk of splitting for node x is computed as follows.

$$Risk(x) = GDI(x) \times Probability(x) \quad (8)$$

The estimate of predictor importance for a certain feature is directly associated with the *GDI* gap between the node corresponding to that feature and its children. This estimation assigns higher importance to features which lead to the largest number of pure children (i.e. terminal nodes). This estimation allots greater importance to the features which have influence on a larger portion of the records. As a result, the feature comprising the root node (in this case the Averaged Power from Figure 4) has higher probability than other features that define rules at lower levels. It allows the feature of the root node to be considered as one of the most important features.

The energy spectral density, averaged power, and range are found to be the most important features in the mortality prediction based on the heart rate signal (Figure 5). As described above, the averaged power is one of the most important features since it is placed as the root of the decision tree. However, the energy spectral density gained the highest score of importance in comparison to the other features. Hence, the nodes corresponding to the energy spectral density feature have higher amount of *GDI* compared to their children. As a matter of fact, this is a sign of high *GDI* gap between these nodes and their children.

The energy spectral density provides basic information about the power variation in frequency components comprising the original signal within a finite interval. Since the power spectral density employs Fourier transform to decompose original signals into a spectrum of frequencies, it can reflect the parasympathetic and sympathetic activities which are highly correlated to the fluctuation of frequency components of heart signals. It has been reported [37] that the high-frequency component reflects parasympathetic nervous activity, while the ratio of low-frequency over the high-frequency components reflects sympathetic nervous activity. Hence, a combination of frequency-domain (e.g. energy spectral density) and time-domain signal analysis (such as skewness) enables us to separate GGU patients who survive or pass away.

5. Conclusion and Future Work

Early hospital risk of mortality prediction in CCU units is critical due to the need for quick and accurate medical decisions. This paper proposes a new signal-based model for early mortality prediction, leveraging the benehts of statistical and signal-based features. Our method is a clinical decision support system which focuses on using only the heart rate signal instead of other health variables such physical state or presence of chronic diseases. Since such variables require laboratory test results which could delay the decision-making time or may not be available at the time of admission, our proposed method may give faster

feedback to healthcare professionals working in CCUs. We demonstrate the capability of using statistical and signal-based features, especially the energy-based features of heart rate signals, to distinguish between patients who survive or pass away in the CCU. Among the interpretable classifiers, the decision tree achieved the highest accuracy, allowing for both accurate and explainable outcomes.

In our future work, we plan to apply our proposed method over other intensive care units, incorporating multiple vital signals along with the heart rate signal as a means to better understand the cause of mortality. The study also can be extended to develop a framework using sensors, laboratory data, and information cached from intensivists and nurses' reports using knowledge graph [38] and text mining [39]. Another direction is to explore the effect of computing features from vital signals with different length of windows and using dynamic feature selections [40] [41]. Finally, we plan on creating a real-time mortality prediction system based on the variability of physiological signals [42] that can predict patient outcomes for early intervention.

Acknowledgments

References

[1]. Calvert J, Mao Q, Hoffman JL, Jay M, Desautels T, Mohamadlou H, Chettipally U, Das R, Using electronic health record collected clinical variables to predict medical intensive care unit mortality, Annals of Medicine and Surgery 11 (2016) 52–57. [PubMed: 27699003]

[2]. Simpson SQ, New sepsis criteria: a change we should not make, CHEST Journal 149 (5) (2016) 1117–1118.

[3]. Le Gall J-R, Lemeshow S, Saulnier F, A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study, Jama 270 (24) (1993) 2957–2963. [PubMed: 8254858]

[4]. Knaus WA, Draper EA, Wagner DP, Zimmerman JE, APACHE II: a severity of disease classification system., Critical care medicine 13 (10) (1985) 818–829. [PubMed: 3928249]

[5]. Dervishi A, Fuzzy risk stratification and risk assessment model for clinical monitoring in the ICU, Computers in Biology and Medicine.

[6]. Awad A, Bader-El-Den M, McNicholas J, Briggs J, Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach, International Journal of Medical Informatics.

[7]. Wojtusiak J, Elashkar E, Nia RM, C-Lace: Computational Model to Predict 30-Day Post-Hospitalization Mortality., in: HEALTHINF, 2017, pp. 169–177.

[8]. Ribas VJ, López JC, Ruiz-Sanmartín A, Ruiz-Rodríguez JC, Rello J, Wojdel A, Vellido A, Severe sepsis mortality prediction with relevance vector machines, in: Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE, IEEE, 2011, pp. 100–103.

[9]. Kim S, Kim W, Park RW, A comparison of intensive care unit mortality prediction models through the use of data mining techniques, Healthcare informatics research 17 (4) (2011) 232–243. [PubMed: 22259725]

[10]. Purushotham S, Meng C, Che Z, Liu Y, Benchmark of Deep Learning Models on Large Healthcare MIMIC Datasets, arXiv preprint arXiv:1710.08531.

[11]. Avati A, Jung K, Harman S, Downing L, Ng A, Shah NH, Improving palliative care with deep learning, arXiv preprint arXiv:1711.06402.

[12]. Beaulieu-Jones BK, Orzechowski P, Moore JH, Mapping patient trajectories using longitudinal extraction and deep learning in the MIMIC-III critical care database, bioRxiv (2017) 177428.

[13]. Song H, Rajan D, Thiagarajan JJ, Spanias A, Attend and Diagnose: Clinical Time Series Analysis using Attention Models, arXiv preprint arXiv:1711.03905.

[14]. Venugopalan J, Chanani N, Maher K, Wang MD, Combination of static and temporal data analysis to predict mortality and readmission in the intensive care, in: Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE, IEEE, 2017, pp. 2570–2573.

[15]. Yadav P, Steinbach M, Kumar V, Simon G, Mining Electronic Health Records (EHRs): A Survey, ACM Computing Surveys (CSUR) 50 (6) (2018) 85.

[16]. Zhang D, Shen X, Qi X, Resting heart rate and all-cause and cardiovascular mortality in the general population: a meta-analysis, Canadian Medical Association Journal (2015) cmaj–150535.

[17]. Johnson AE, Pollard TJ, Shen L, Lehman L.-w. H., Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG, MIMIC-III, a freely accessible critical care database, Scientific data 3.

[18]. Vincent J-L, Singer M, Critical care: advances and future perspectives, The Lancet 376 (9749) (2010) 1354–1361.

[19]. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE, APACHE-acute physiology and chronic health evaluation: a physiologically based classification system., Critical care medicine 9 (8) (1981) 591–597. [PubMed: 7261642]

[20]. Zimmerman JE, Kramer AA, McNair DS, Malila FM, Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients, Critical care medicine 34 (5) (2006) 1297–1310. [PubMed: 16540951]

[21]. Johnson AE, Pollard TJ, Mark RG, Reproducibility in critical care: a mortality prediction case study, in: Machine Learning for Healthcare Conference, 2017, pp. 361–376.

[22]. Le Gall J-R, Lemeshow S, Leleu G, Klar J, Huillard J, Rué M, Teres D, Artigas A, Customized probability models for early severe sepsis in adult intensive care patients, Jama 273 (8) (1995) 644–650. [PubMed: 7844875]

[23]. Le Gall JR, Neumann A, Hemery F, Bleriot JP, Fulgencio JP, Garrigues B, Gouzes C, Lepage E, Moine P, Villers D, Mortality prediction using SAPS II: an update for French intensive care units, Critical Care 9 (6) (2005) R645. [PubMed: 16280063]

[24]. Metnitz B, Schaden E, Moreno R, Le Gall J-R, Bauer P, Metnitz PG, Group AS, Austrian validation and customization of the SAPS 3 Admission Score, Intensive care medicine 35 (4) (2009) 616–622. [PubMed: 18846365]

[25]. Banerjee T, Peterson M, Oliver Q, Froehle A, Lawhorne L, Validating a Commercial Device for Continuous Activity Measurement in the Older Adult Population for Dementia Management, Smart Health.

[26]. Hentschel T, Henker M, Fettweis G, The digital front-end of software radio terminals, IEEE Personal communications 6 (4) (1999) 40–46.

[27]. Cohen H, Neumann L, Shore M, Amir M, Cassuto Y, Buskila D, Autonomic dysfunction in patients with fibromyalgia: application of power spectral analysis of heart rate variability, in: Seminars in arthritis and rheumatism, Vol. 29, Elsevier, 2000, pp. 217–227. [PubMed: 10707990]

[28]. Perry T, Bader-El-Den M, Cooper S, Imbalanced classification using genetically optimized cost sensitive classifiers, in: Evolutionary Computation (CEC), 2015 IEEE Congress on, IEEE, 2015, pp. 680–687.

[29]. Sadeghi R, Hamidzadeh J, Automatic support vector data description, Soft Computing 22 (1) (2018) 147–158.

[30]. Hamidzadeh J, Sadeghi R, Namaei N, Weighted support vector data description based on chaotic bat algorithm, Applied Soft Computing 60 (2017) 540–551.

[31]. Nekooeimehr I, Lai-Yuen SK, Adaptive semi-unsupervised weighted oversampling (A-SUWO) for imbalanced datasets, Expert Systems with Applications 46 (2016) 405–416.

[32]. Breiman L, Friedman J, Stone CJ, Olshen RA, Classification and regression trees, CRC press, 1984.

[33]. Raileanu LE, Stoffel K, Theoretical comparison between the gini index and information gain criteria, Annals of Mathematics and Artificial Intelligence 41 (1) (2004) 77–93.

[34]. Breiman L, Random Forests Machine Learning. 45: 5–32, View Article PubMed/NCBI Google Scholar.

[35]. Margineantu DD, Dietterich TG, Pruning adaptive boosting, in: ICML, Vol. 97, 1997, pp. 211–218.

[36]. Caruana R, Niculescu-Mizil A, An empirical comparison of supervised learning algorithms, in: Proceedings of the 23rd international conference on Machine learning, ACM, 2006, pp. 161–168.

[37]. Hasegawa M, Hayano A, Kawaguchi A, Yamanaka R, Assessment of autonomic nervous system function in nursing students using an autonomic reflex orthostatic test by heart rate spectral analysis, Biomedical reports 3 (6) (2015) 831–834. [PubMed: 26623025]

[38]. Shekarpour S, Marx E, Auer S, Sheth AP, RQUERY: Rewriting Natural Language Queries on Knowledge Graphs to Alleviate the Vocabulary Mismatch Problem., in: AAAI, 2017, pp. 3936–3943.

[39]. Allahyari M, Pouriyeh S, Assefi M, Safaei S, Trippe ED, Gutierrez JB, Kochut K, A brief survey of text mining: Classification, clustering and extraction techniques, arXiv preprint arXiv: 1707.02919.

[40]. Zabihimayvan M, Sadeghi R, Rude HN, Doran D, A soft computing approach for benign and malicious web robot detection, Expert Systems with Applications 87 (2017) 129–140.

[41]. Hamidzadeh J, Zabihimayvan M, Sadeghi R, Detection of Web site visitors based on fuzzy rough sets, Soft Computing 22 (7) (2018) 2175–2188.

[42]. Kaffashi F, Variability Analysis & Its Applications to Physiological Time Series Data, PhD Thesis, Case Western Reserve University (2007).

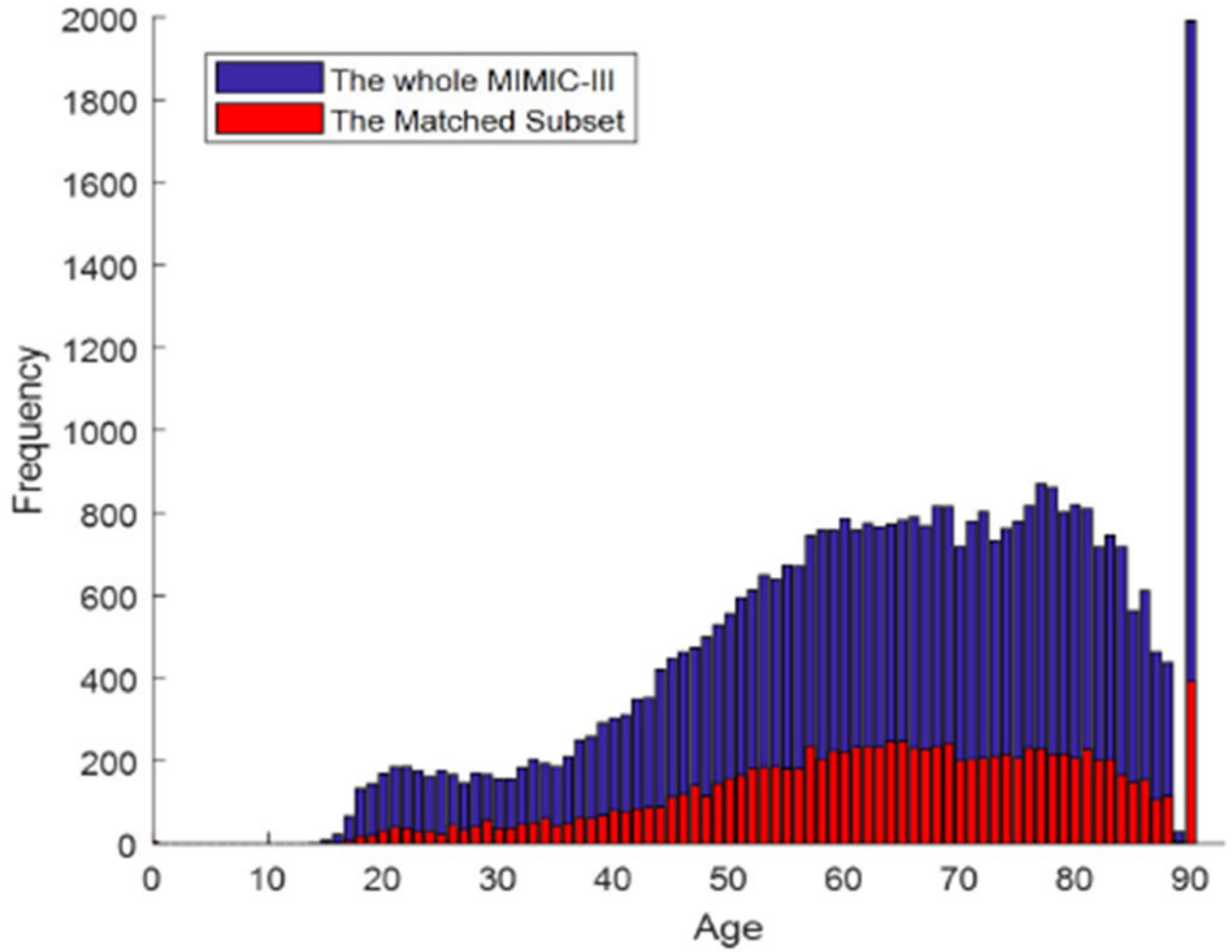**Figure 1:**
The age distribution over the Whole MIMIC-III (without infants) and the Matched Subset
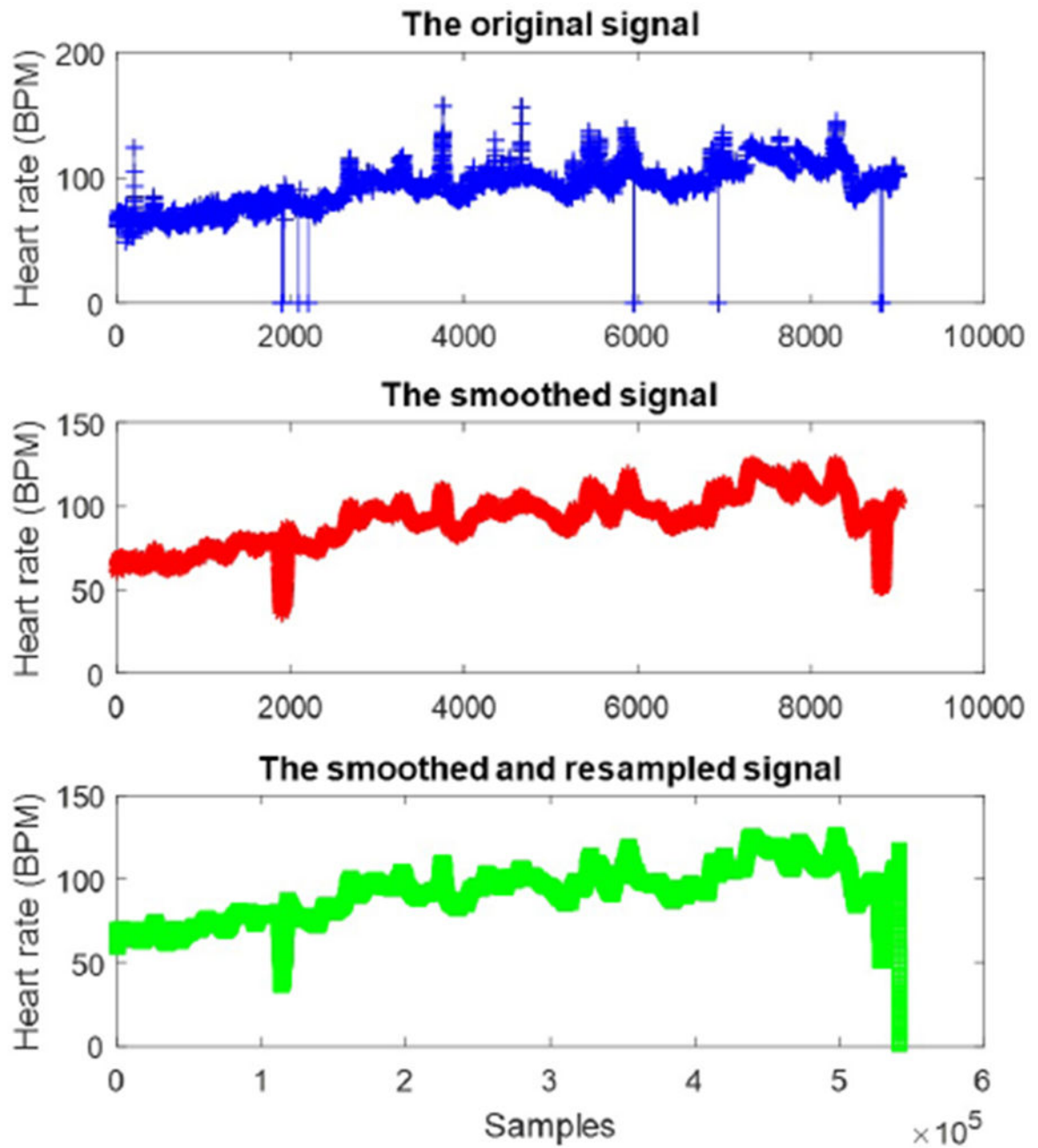
Figure 2:
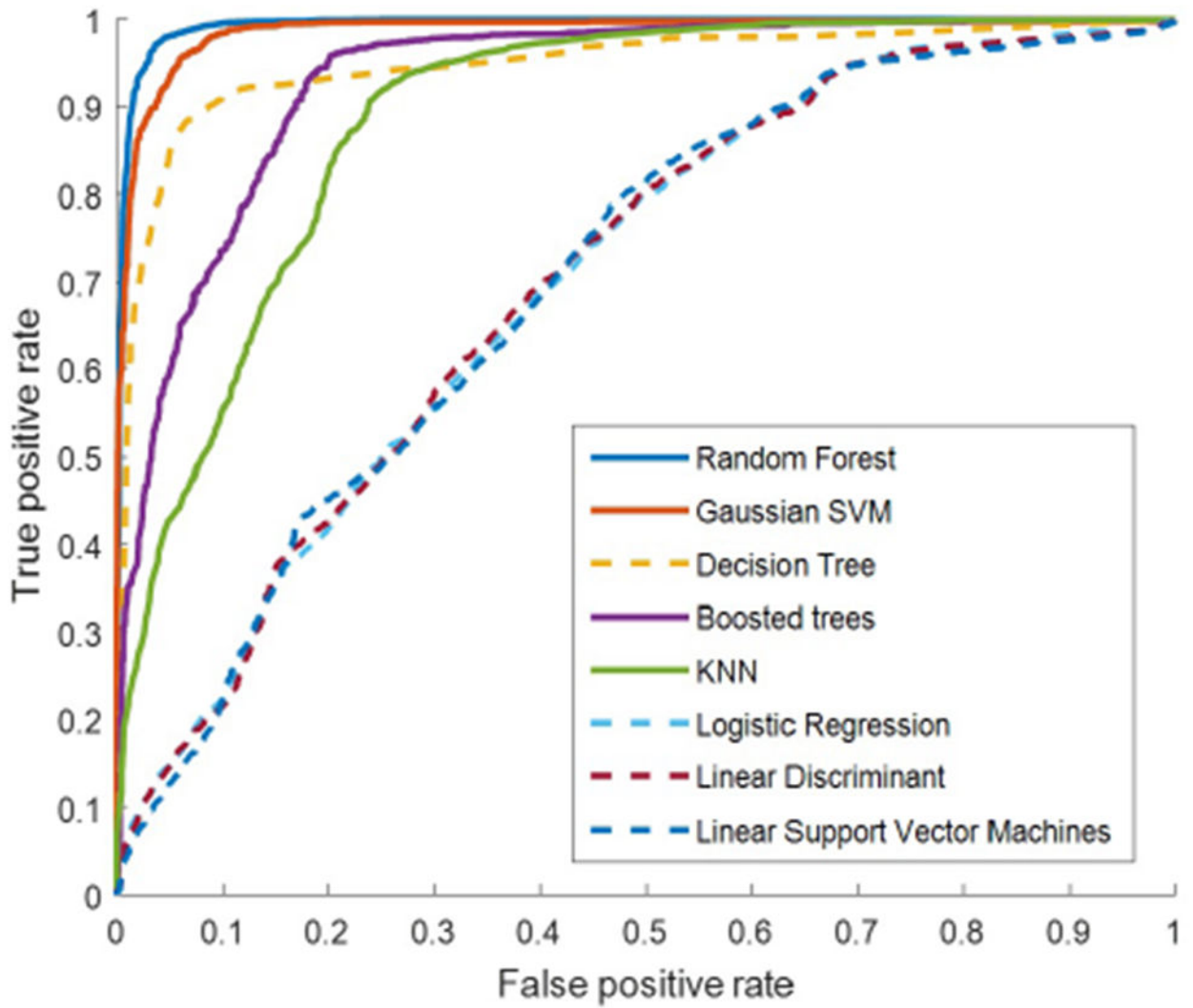The preprocessed heart rate signal of one survived patient from CCU

Figure 3:
The ROC curves of transparent and black-box classifiers shown by dashed and solid lines, respectively
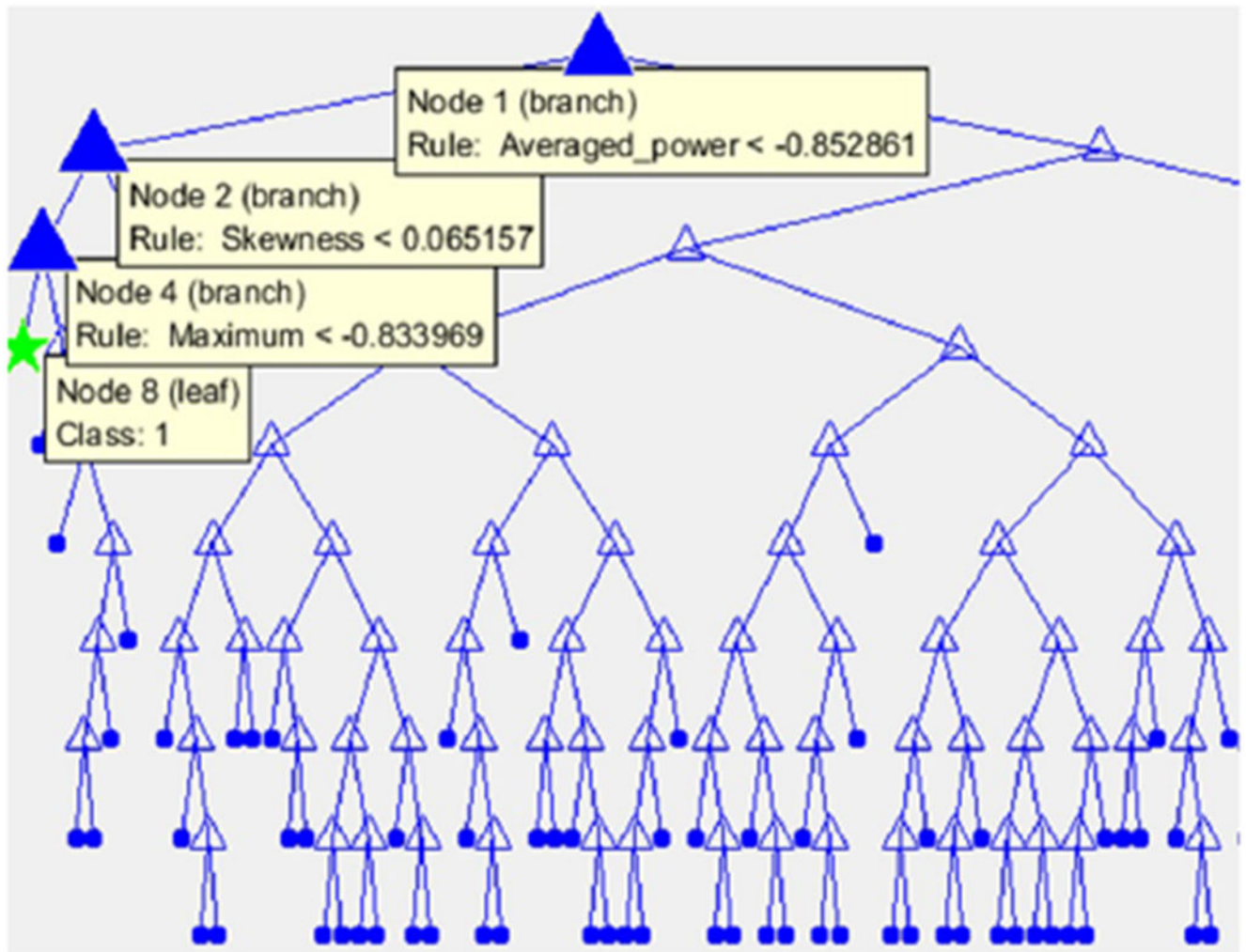
Figure 4:
The decision tree model comprises decision and leaf nodes represented by triangles and the dots, respectively. The green star shows the left-most leaf node of the model.
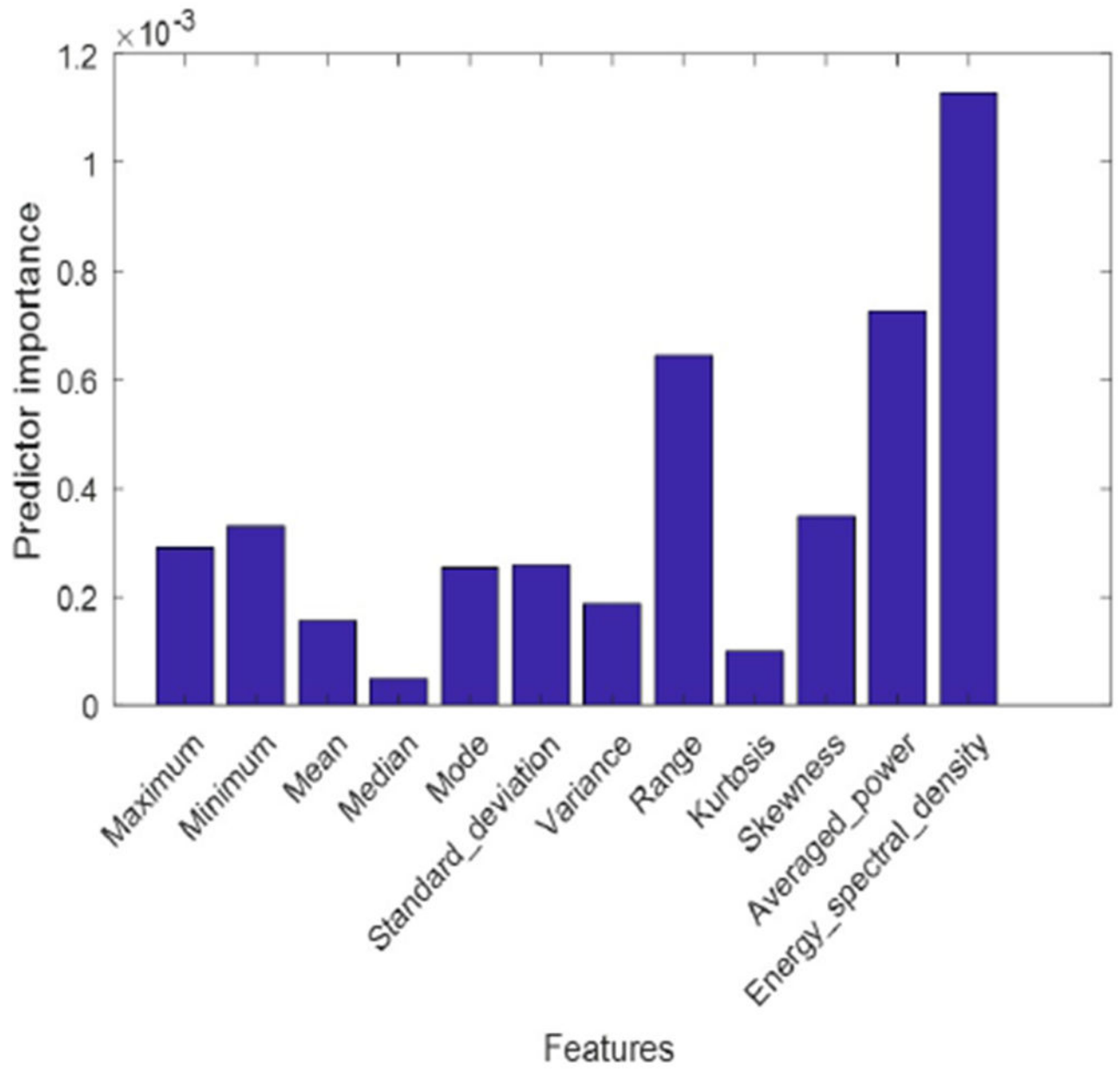
Figure 5:
Feature importance in the proposed model for mortality prediction based on heart rate signal

Table 1:

Care Units in MIMIC-III

Care unit	Description
CCU	Coronary care unit
CSRU	Cardiac surgery recovery unit
MICU	Medical intensive care unit
NICU	Neonatal intensive care unit
NWARD	Neonatal ward
SICU	Surgical intensive care unit
TSICU	Trauma/surgical intensive care unit

Table 2:

Descriptive Statistics for Statistical and Signal-Based Features

Column	Feature	Passed away patients	Alive patients
1	Maximum	97.82	90.92
2	Minimum	80.69	76.24
3	Mean	88.46	81.92
4	Median	88.45	81.81
5	Mode	85.25	79.98
6	Standard deviation	2.63	2.25
7	Variance	15.84	11.56
8	Range	17.13	14.68
9	Kurtosis	17.48	17.85
10	Skewness	0.83	1.02
11	Averaged power	8186.02	7045.04
12	Energy spectral density	5114.78	4420.38

Table 3:

Classification Results for CCU Mortality

Classifier	Precision	Recall	F1-score	Interpretability
Random forest	0.97	0.97	0.97	Hard
Gaussian SVM	0.95	0.96	0.96	Hard
Decision tree	0.90	0.92	0.91	Easy
Boosted trees	0.91	0.83	0.87	Hard
K-NN	0.80	0.85	0.82	Hard
Logistic regression	0.77	0.67	0.72	Easy
Linear discriminant	0.78	0.66	0.71	Easy
Linear SVM	0.80	0.63	0.70	Easy