Predicting Mortality Rate based on Comprehensive Features of Intensive Care Unit Patients

Abstract—**Predictive analytics is an emerging area in healthcare to identify patient mortality through selected health metrics. As per the statistics, the survival rate of patients is found to be 89.76% while 10.24% was observed to be the mortality rate out of, 9111 ICU patients. Few research studies have attempted to use EHR (Electronic Health Record) data to identify the trends to improve the mortality rate. The EHR data along with the admission details of ICU patients are used to build the machine learning models. This paper proposes a hybrid prediction technique that uses Random Forest and XGBoost for predicting the mortality rate. This work found that the proposed method performed well in predicting the mortality rate while the class-imbalance problem was taken care. This was achieved by employing several class-imbalance techniques. Despite having high accuracy (89.72%), sensitivity (99.02%) and AUC(75.98%) of the XGBoost model, the specificity remained low due to class imbalance. The proposed method with the obtained results is discussed in the paper.**

Index Terms—**MIMIC-IV, Machine Learning, Intensive Care Unit, Mortality rate**

I. Introduction

With the increase in demand for healthcare policies, different healthcare service companies wanted a systematic approach to test the patients and, for the most part, determine their recovery, which is relatively significant. Therefore, they adopted the Electronic Health Record (EHR) system in a generally big way. Over two decades, the EHR data mainly has been used by healthcare service providers on a large scale. Generally, EHR data contains a detailed report of patients along with their historical data. Researchers attempted to solve the problem of identifying the trends in improving the mortality rate using data-driven approach [1], [2], [3], [4]. In addition to that, predictive analytics finds the mortality rate through pre-defined health metrics with the help of an EHR system. Even though the EHR is effective for some purposes, it doesn't provide a way to understand the possible treatment the patient needs through the historical data. This opens a problem of uncertainty in the medical facilities that a patient needs to go through. This scenario cannot be avoided by EHR because the models essentially use patients' demographic details. Due to this, Predictive analytics has become a new way of looking into the problem of mortality prediction through pre-defined parameters.

However, the uncertainty cannot essentially be addressed in the EHR system — most of the models built based on patient demographic details and, for all intents and purposes, other clinical tests, which shows that EHR consists of detailed patient details and their historical data.

There is a possibility that the majority of the Intensive Care Unit (ICU) patients will go under surgeries, which may vary accordingly. The medical conditions can be cardio, neuro, thoracic, vascular, reconstruction of the body involving plastic surgery, etc. [5]. Therefore, the surgeries will influence financial and social development, which should be a serious concern. Often, the patients are admitted on an emergency basis, which may lead to very acute medication in the ICU. According to a study, patients in the USA admitted to ICU 35% to 50% are admitted in an emergency mode whereas 18% to 25% with need different surgeries. Therefore, the statistics also show that the study of mortality rate is a critical problem. The present work focuses on predicting the mortality rate of ICU patients with various means. The main objective of this work is to build a predictive model that uses comprehensive data from various ICU patients. The data that has been used in this work is provided by Medical Information Mart for Intensive Care (MIMIC-IV version 1.0) [5].

The paper is organized as follows: The section II provides a brief overview on the prior work related to the problem. In section III, we discuss the methodology in detail. Section V gives a detailed analysis of the obtained results. Finally, VI will conclude the paper by giving some future insights.

II. Related Work

In [1] authors proposed a dynamic data clinical data mining in the context of electronic medical records (EMR). Initially, the patient EMR matches with the existing database to profile the diseases, then the system provides the feedback system to enhance the patient EMR. Once the profiling is completed, then clinical decision takes place. Eventually, the health records like hospital discharge, readmission, death, patient-reported outcome etc., going to share with inside and outside the medical agencies. Later, the entire data is updated to the global database.

Heart rate (HR) and blood pressure (BP) are regulated by an underlying control system, therefore time series of these vital signs exhibit substantial dynamical patterns of interaction in response to external perturbations and pathological states. In [2], employed a switching vector autoregressive framework to systematically learn and establish a set of vital sign time series dynamics that may be recurrent within the same patient and common across the cohort. Authors illustrate how these dynamical behaviors might be used to characterize a patient's physiological "state". For this task, they have used MIMIC-II database with 450 ICU patients and discovered cardiovascular dynamics are significantly associated with hospital mortality.

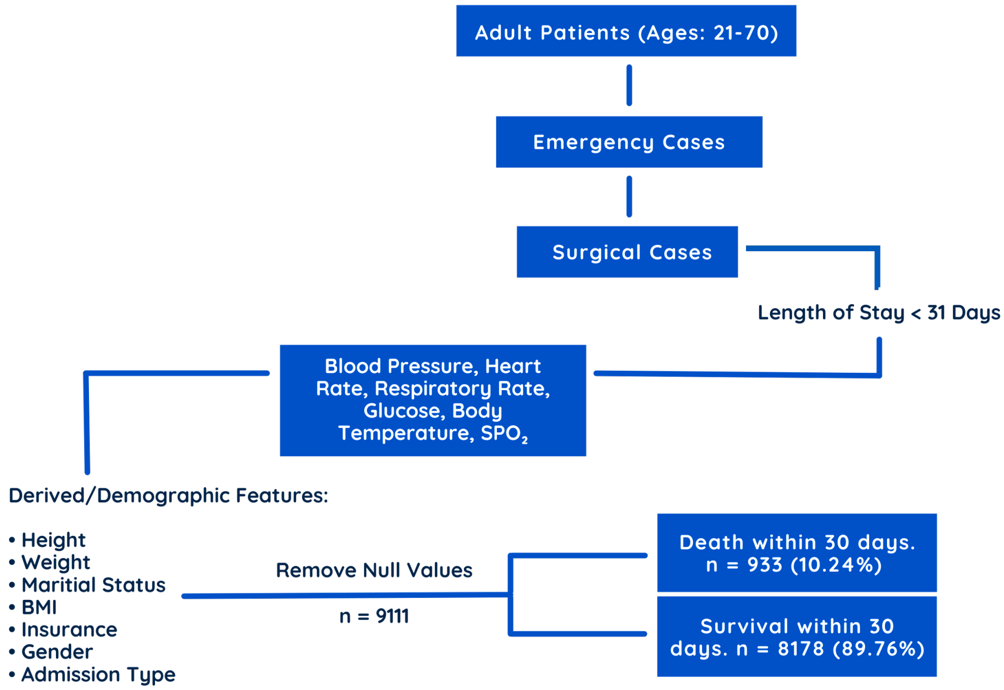

Fig. 1: Flowchart of Patient Profiling

Finally, the authors' calculate mortality risk scores from a 6-hour sliding window by stride of 1-hour, which were calculated as the likelihood of death of the ICU patients.

In [3] authors proposed prediction of patient mortality risk in ICU using randomForest. The main objective of the proposed method based on basic patient details and time series of vital signs and lab reports in the first 48 hours of ICU is about 4000 adult patients. The performance of their approach is based on the SAPS-1 scoring method and analyzing the risk predictions and alerting health agencies.

In [6], an ensemble method proposed to predict the mortality rate of patients in the ICU. In their approach, the features were selected by ICU domain experts. The entire data is divided into 6 modalities, depending on expert insights. Individual classifiers were selected for the prediction process based on the performance of the classifiers for each modality. We used the most popular and diverse classifiers in the literature, including linear discriminant analysis, decision tree (DT), multilayer perception, nearest neighbor method, and logistic regression (LR). The stacking ensemble classifier was then built and optimized based on the merging of these five classifier decisions. The framework was evaluated using 10,664 patients from the Medical Information Mart (MIMIC III) benchmark dataset for intensive care. Extensive experiments were performed using time series data of patients of various lengths to predict patient mortality. For each patient, the first 6, 12, and 24 hours of the first stay were tested. Experimental results showed an accuracy of approximately 94.4%, an F1 score of 93.7%, an accuracy of 96.4%, a recall of 91.1%, and a ROC curve of 93.3%.

In this work [7] authors, a deep learning framework is proposed to improve severity scoring methods of diabetes patients. Firstly, the historical data obtained from the health records based on the event logs, which are suitable for process mining. These event logs are then used to identify any hidden patterns that describe the past hospital encounters of patients. An adaptation of Decay Replay Mining is proposed to combine medical and demographic information with established severity scores to predict in-hospital mortality in diabetic patients in intensive care units. Significant overall performance upgrades are verified as compared to mounted threat severity scoring techniques and device mastering tactics, the use of the Medical Information Mart for Intensive Care III dataset.

A traversal pathology is the concept of sepsis, and it is the main reason leading to death of ICU patients reported in [8]. The main findings of this paper is to predict accurate and robust cause of ICU deaths based on APACHE II score and a total of 34 features used to evaluate the performance of their model using support vector machines.

The prediction of ICU patients mortality rate is based on the severity of illness scores proposed in [9]. The proposed approach used MIMIC-II dataset for 30 days of ICU stay of patients. The authors built a Deep Learning approach with self normalizing neural network that and compared the existing methods. The experimental results reported of their proposed approach based on AUC of 0.8445 for 30 days mortality.

In [10], the authors compared the popular two datasets EMR's and MIMIC-II to predict ICU patient mortality rate within a median of 72 hours. In their, firstly, the temporal features from EMRs and building a predictive model. Secondly, the MIMIC-II features extracted and determined the patient

TABLE I: Features of Mortality Rate used in this work

Feature	Description	Module
is_male	patient is male or female	Core
admission_age	age of a patient at the time of hospital	Core
marital_status	marital status	Core
insurance	type of insurance	Core
admission_type	type of Core (Emergency etc.)	Core
los_icu	length of ICU stay	ICU
service_type_SURG	is the current service type is surgery or other Hospital	Core
height	height of a patient in centimeters	Derived
weight	weight at the time of admission	Derived
heart_rate_mean	mean heart rate at the time of hospital	Derived
sbp_mean	mean sbp	Derived
dbp_mean	mean dbp	Derived
resp_rate_mean	mean respiratory rate	Derived
body_temperature_mean	mean body temperature	Derived
spo2_mean	mean spo_2	Derived
glucose_mean	mean glucose	Derived
hospstay_flag	patient admission ranking	Core

similarities and predicted the mortality rate. They analyzed and compare both approaches on the to predict patient mortality and find that the patient similarity approach. They found that these two models does not have any similarities and results in a less accurate model about AUC of 0.68 compared to the modelling approach 0.84.

III. Methodology

The steps involved in the approach can be summarized as follows:

1) Define and aggregate ICU patient data based on patient's admissions.
2) Extract comprehensive features from the historical data from Google BigQuery.
3) Build a machine learning model from the derived features accurately.
4) Validate test data from the built model.

In the below section, a detailed description of our methodology based on the proposed architecture. Firstly, patient profiling, feature extraction, statistical test, handling missing data, and system design.

A. Patient Profiling

The patient profiling with an emergency and type of surgery, patient age between 21 and 70 years old, and ICU stay length limited to less than are equal to 30 days included for this research. Figure 1, illustrates the patient profiling and a total of, 9111 patients. Out of, 9111 patients, around 10.24% and 89.76% of patient expired and survived in ICU respectively.

B. Feature Extraction

Using Google BigQuery SQL demographic features, vital signs, and derived features (height & weights) were extracted from the MIMIC-IV [1] dataset: Admissions, Patients, ICU stays and services. This research study focus mainly on the following features: demographic details' viz., in-hospital age of a patient, gender, ethnicity, weight, and height; vital signs' viz., average heart rate, mean systolic blood pressure (SBP), mean diastolic blood pressure (DBP), mean blood pressure, mean respiratory rate, mean body temperature, mean saturation pulse oxygen (SPO2), mean glucose. The main objective of the study is to determine the hospital mortality, based on the subset of feature that can bring a great impact of prediction can be accurate. The features extracted from the relational database by aggregating data from ICU stay at a given hospital. The aggregation data obtained from the Google BigQuery platform. Google BigQuery platform is distributed in nature that can process a billion records in a quicker manner. The below Table I includes the features extracted from the relational database.

The dataset contains a total number of patients around 400K, the total number of admissions around 600K, the patient non-survival rate during the hospital is 1.78%, and survival rate is 98.22%. The mortality rate is a pretty high class imbalance that is a challenging problem to solve.

C. Statistical Test

To understand the importance of features, χ^2 test is performed on Table I. χ^2 test is a statistical significance test that can calculate correlation between dependent and independent

[1]https://physionet.org/content/mimiciv/1.0/

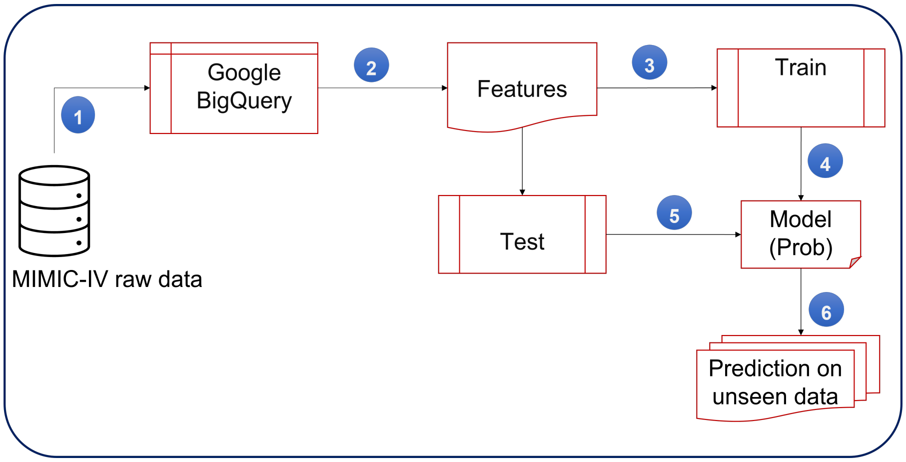

Fig. 2: Proposed workflow of ICU patient mortality prediction

of multivariables. The χ^2 is used to select relevant characterization of ICU patients of mortality. While selecting the features, if the target variable is independent of feature, can be removed. The final subset of features illustrated in Table II.

D. Handling Missing Data

It is very common to have a missing value in the data that must deal with it. It is due to human error, this study ignored any missing values in database that does not create any bias in our analysis.

E. System Design

This section discusses the initial approach to classify the mortality rate of ICU stay patients. Figure 2 describes about the proposed approach. Firstly, from Google BigQuery, SQL queries written to obtain the aggregated feature, which is shown in Table I. The data preprocessing, such as categorical data encoded as numerical and feature engineering on service type feature, has handled within SQL query. Secondly, the feature set exported as ".csv" from the SQL query, the dataset split into train and test for the machine learning task. Then, two machine learning models Naive Bayes and Random Forest models built. The model tested on unseen data to evaluate the model performance as accuracy, sensitivity, and specificity.

TABLE II: χ^2 test of independence of feature and its critical value

Feature	χ^2 **Critical value**
body_temparature_mean	0.181
spo_2	0.173
los_icu	0.167
rsp_rate_mean	0.165
sbp_mean	0.147
dbp_mean	0.114
heart_rate_mean	0.109

IV. Experimental Setup

In this section, we discuss the following :

1) Dataset
2) Implementation Details (libraries and Platform)

A. Dataset

The MIMIC-IV database version 1.0 [5] contains data from 2008 to 2019 that is collected from metavision bedside monitor of ICU patients. MIMIC-IV is a publicly available de-identified data of 382,278 subjects of ICU admissions. It consists of actual hospital patient stay data in a relational database admitted to a tertiary academic medical center in Boston, MA, USA. The dataset obtained from MIT after successful completion of Collaborative Institutional Training Initiative (CITI), total 15 courses under Sunny Buffalo suggested by CITI to obtain the access of MIMIC-IV maintained by research group "Physionet" [11] for the research purpose. It has comprehensive measures of each patient, classified into six different modules.

- **core:** Patient demography details as admissions and transfers
- **hosp:** Hospital details such a laboratory and electronic test reports
- **icu:** Intensive care unit level details
- **ed:** Data obtained from emergency department
- **cxr:** Data from analysis of patient chest x-rays
- **note:** de-identified clinical text data. It is not available for MIMIC-IV version.

V. Results and Analysis

A. Analysis of Machine Learning Models

The importance of variables chosen by each approach in the derivation group, assessed using multi-variate random forest and XGBoost analysis to uncover independent risk factors of in-hospital mortality. χ^2 test used to find variables that were strongly linked to in-hospital mortality. The potential for non-linearity between candidate continuous variables and in-hospital mortality rate examined.

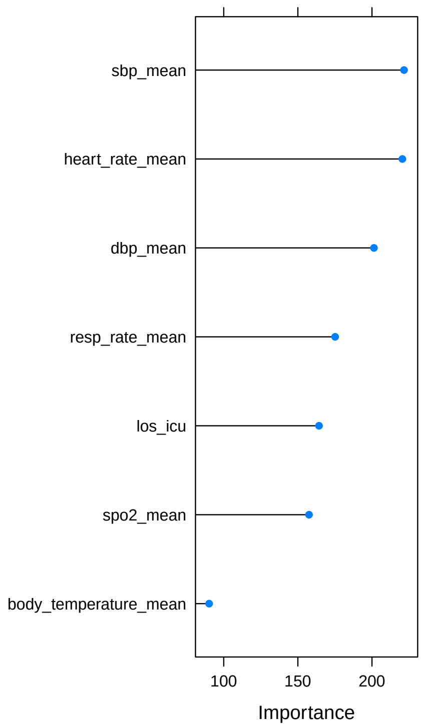

Fig. 3: Feature importance of both the models. X-axis indicates model score and y-axis selected feature

The process for developing the prediction model is described in Figure 2. Based on available research, expert knowledge, and availability in clinical practice. The predictor variables and summary statistics summarized in the following Table III. It shows, the selected features are more significant is based on the **p-value**. The objective is of this emergency cases and service type is surgery can be our hypothesis is measured. The most essential indicators for the mortality prediction model chosen from the derivation group using two different techniques. To begin, we utilized extreme gradient boosting (XGBoost) and Random Forest, supervised machine-learning and data-mining techniques which uses a meta-algorithm to build a powerful ensemble learner from weak learners like regression trees.

The tree topologies and leaf node weights make up the parameters of a regression tree. They are optimized sequentially using gradient techniques to minimize an objective function that consists of a fitting loss term and a regularization term. By using a weighted quantile sketch to approximate an optimization computation and designing a column block structure for parallel learning, XGBoost retrofits the tree-learning technique to handle sparse data. The XGBoost algorithm can show how much each predictor contributes, allowing you to pick the most relevant predictors, whereas the random forest algorithm generates reasonable predictions across a wide range of data while requiring little configuration. Further, we employed the feature importance that are effective to classify survival over

TABLE III: Descriptive statistics of Patient of ICU stay profiling less than are equal to 30 days using Multivariate Logistic regression

Features	**Total Patients (n=9111)**	**Survived (n=8178)**	**Died (n=8178)**	**P value (n=933)**
body_temperature_mean , mean	36.93	36.74	36.95	2.00E-16
spo2_mean , mean	97.02	96.24	97.11	1.23E-08
resp_rate_mean , mean	19.33	21.3	19.11	4.59E-16
los_icu , mean	4.55	6.64	4.311	2.00E-16
sbp_mean , mean	117.4	111.3	118.1	0.00217
dbp_mean , mean	65.85	62.38	66.25	3.02E-09
heart_rate_mean , mean	87.95	93.49	87.32	5.87E-09

non-survival rate, this can handle the class imbalance problem to some extent. The feature importance of both models in illustrated in Figure 3.

- Feature section technique
- Selecting non-linear model
- An ensemble learning approach can be selected, i.e., Voting or Stacking.
- Three metrics can be measured, i.e., accuracy, precision, and recall

B. Results and Discussions

There are a total of, 9111 patients based on the extraction criteria is show in 1 of the MIMIC-IV dataset. Patients without a record for any of the important features excluded. The patients randomly divided into test (n=6834, 75%) and train data (n=2277, 25%) group. There is small multicollinearity observed within the independent variables. This gives a reliable confidence interval that produce more reliable probabilities in terms of effect of the independent variables in a model. Therefore, the inference from the statistical analysis from the model can be more robust and reliable.

In this study, a Random Forest and XGBoost used to build the prediction models. The selection of these models is to handle the class imbalance well understood. As these are the ensemble models, these aggregate the prediction of each base model and results in a better prediction for the unseen data as mentioned in the previous section. The issue with the class imbalance dataset has high variance error than the bias. Despite the high accuracy (89.72%), sensitivity (99.02%), and AUC (75.98%) of the XGBoost model, the specificity remained low at 8.15% due to class imbalance. In [6], reported an AUC of 0.933 on class balanced dataset and in [7] reported an AUC of 0.826 of MIMMIC-III dataset. Several imbalance techniques like SMOTE (Majority Under Sampling or Minority Over Sampling) were applied to overcome this problem, the above table shows final results attained with these two classification models (random Forest, XGBoost) illustrated in Table IV.

In the model validation phase measured using area under the curve, the ensemble tree based ML algorithm (RF) and XGBoost is about 0.7592 and 0.7598 respectively. The AUC value has very marginal difference between Random Forest and XGBoost. However, in terms of computational speed, RF

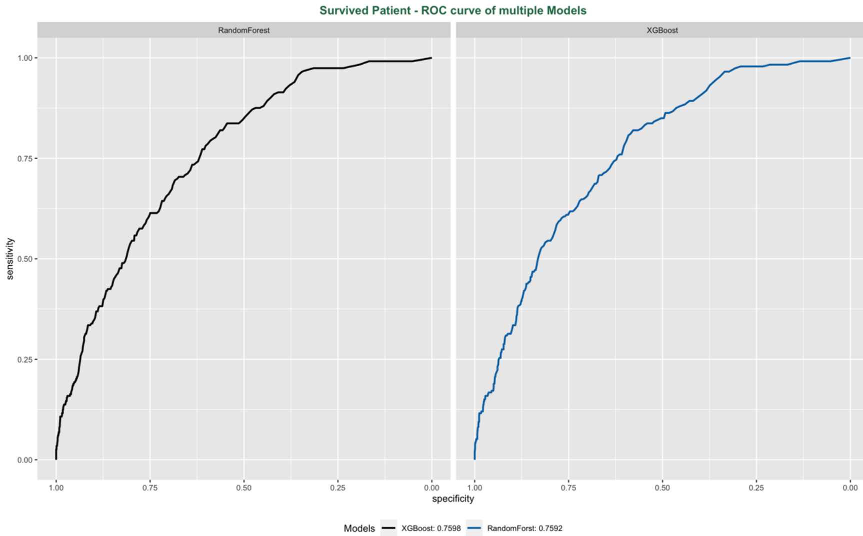

Fig. 4: The receiver operating characteristic (ROC) curves. Random Forest model (Left), area under curves (AUC) is 0.7592 c XGboost model (Right), AUC is 0.7598

TABLE IV: Test set Results

Model	**Accuracy %**	**Sensitivity %**	**Specificity %**
Random Forest	89.77	99.02	8.54
XGBoost	89.72	99.02	8.15

is faster than XGBoost due hyperparameter tuning would be more expensive.

VI. Future Work and Conclusions

This study shows a case of accuracy paradox, meaning high accuracy can sometimes be deceiving and alone is not a good metric for predictive models when classifying in predictive analytics. We believe with inclusion of more features from the chart events module from the database, these newly added features should be more significant enough should determine the mortality rate of a patient.

Our future work focuses on detailed study of fine-grained of each disease from the chart/lab events of MIMIC-IV. The chart of events had detailed information about the type of medication provided during the ICU stay. Further, the accuracy of class imbalance problem can be improving by introducing various methods like stratified. It is sampling of random split of training and testing set would be the limitation of this approach.

References

[1] L. A. Celi, A. J. Zimolzak, and D. J. Stone, "Dynamic clinical data mining: search engine-based decision support," *JMIR medical informatics*, vol. 2, no. 1, p. e13, 2014.

[2] H. L. Li-wei, R. P. Adams, L. Mayaud, G. B. Moody, A. Malhotra, R. G. Mark, and S. Nemati, "A physiological time series dynamics-based approach to patient monitoring and outcome prediction," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 1068–1076, 2014.

[3] S. Ghose, J. Mitra, S. Khanna, and J. Dowling, "An improved patient-specific mortality risk prediction in icu in a random forest classification framework," *Stud Health Technol Inform*, vol. 214, pp. 56–61, 2015.

[4] R. Pirracchio, M. L. Petersen, M. Carone, M. R. Rigon, S. Chevret, and M. J. van der Laan, "Mortality prediction in intensive care units with the super icu learner algorithm (sicula): a population-based study," *The Lancet Respiratory Medicine*, vol. 3, no. 1, pp. 42–52, 2015.

[5] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. A. Celi, and R. Mark, "MIMIC-IV version:1.0," 2021.

[6] N. El-Rashidy, S. El-Sappagh, T. Abuhmed, S. Abdelrazek, and H. M. El-Bakry, "Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model," *IEEE Access*, vol. 8, pp. 133 541–133 564, 2020.

[7] J. Theis, W. L. Galanter, A. D. Boyd, and H. Darabi, "Improving the in-hospital mortality prediction of diabetes icu patients using a process mining/deep learning architecture," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 388–399, 2022.

[8] V. J. Ribas, J. C. López, A. Ruiz-Sanmartín, J. C. Ruiz-Rodríguez, J. Rello, A. Wojdel, and A. Vellido, "Severe sepsis mortality prediction with relevance vector machines," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 100–103.

[9] M. A. H. Zahid and J. Lee, "Mortality prediction with self normalizing neural networks in intensive care unit patients," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2018, pp. 226–229.

[10] M. Hoogendoorn, A. el Hassouni, K. Mok, M. Ghassemi, and P. Szolovits, "Prediction using patient comparison vs. modeling: A case study for mortality prediction," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 2464–2467.

[11] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.