medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

TITLE

Prediction of 30-Day Mortality for ICU Patients with Sepsis-3

Yu Z[1], Li H[1], Pishgar M[1]

1 Department of Industrial and Systems Engineering, University of Southern California (USC), 3650 McClintock Ave, Los Angeles, CA 90089, United States of America

Corresponding Author
Maryam Pishgar, PhD
Professor of Industrial and Systems Engineering
University of Southern California
Phone: (213) 740-4893
Postal address: Andrus Gerontology Center, 3715 McClintock Ave GER 240, Los Angeles, CA 90089
Email: pishgar@usc.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

ABSTRACT

Background

There has emerged an increasing demand for advanced methodologies aimed at augmenting our comprehension and prognostication of illnesses. This study is distinctly centered on tackling the complexity of Sepsis, an immediate bodily reaction to infection. Our objective is to refine the early identification and mortality forecasting for patients diagnosed under the Sepsis-3 criteria, with the overarching aim of enhancing the allocation of hospital resources.

Methods

In this study, we introduced a Machine Learning (ML) framework aimed at predicting the 30-day mortality rate among Intensive Care Unit (ICU) patients diagnosed with Sepsis-3. Leveraging the Medical Information Mart for Intensive Care III (MIMIC-III) database, we systematically identified eligible patients using advanced big data extraction tools such as Snowflake. Additionally, we employed decision tree models to ascertain the importance of various features and conducted entropy analyses across decision nodes to refine feature selection. Collaborating with esteemed clinical experts, we curated a list of 30 relevant features. Moreover, we used the Light Gradient Boosting Machine (LightGBM) model due to its gradient boosting architecture and computational efficiency.

Results

The study comprised a cohort of 9118 patients diagnosed with Sepsis-3. Through our meticulous preprocessing techniques, we observed a marked enhancement in both the Area Under the Curve (AUC) and accuracy metrics. The LightGBM model yielded an impressive AUC of 0.983, with a 95% confidence interval [0.980-0.990]. Moreover, it exhibited a commendable accuracy of 0.966 and an F1-score of 0.910. Notably, LightGBM showcased a substantial 6% enhancement over our best baseline model and a significant 14% enhancement over the best existing literature. These noteworthy advancements can be attributed to several factors: (I) the incorporation of a novel and pivotal feature in our model, Hospital Length of Stay (HOSP_LOS), which has not been included in previous literature; (II) the inherent strengths of LightGBM's gradient boosting architecture, enabling robust predictions even with high-dimensional data, while maintaining computational efficiency, as evidenced by its learning curve.

Conclusions

The introduced preprocessing methodology not only led to a substantial reduction in the number of relevant features compared to the best existing literature, thereby alleviating computational complexities, but also enabled the identification of a crucial feature previously ignored in existing literature. Through the integration of these pivotal features and meticulous parameter tuning, our proposed model achieved remarkable predictive power, with its learning curve demonstrating its capacity for generalization to unseen data. This underscores the potential of ML as indispensable tools in the dynamic environment of the ICU. Employing our model stands to streamline resource allocation within ICUs, offering clinicians greater efficiency and tailored interventions for patients afflicted with Sepsis-3.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Keywords: Sepsis-3 prediction, Machine Learning, Entropy Analysis, Gradient Boosting Machine Model

BACKGROUND

Sepsis [1], a life-threatening condition triggered by infection, often leads to organ failure and exhibits rapid, unpredictable progression [2]. In the United States alone, Sepsis affects approximately 1.7 million adults annually, resulting in around 270,000 deaths. Notably, recent research involving over 110,000 hospital admissions underscored a significant association between prolonged hospital stays and diminished survival rates, particularly for stays exceeding nine days. Globally, Sepsis accounted for nearly a fifth of all reported fatalities in 2017, with an estimated 11 million deaths out of nearly 49 million reported cases [3]. Given the severity of Sepsis's impact, understanding the factors contributing to elevated mortality rates among patients is imperative.

The understanding of Sepsis has evolved notably with the introduction of Sepsis-3 by the Third International Consensus Definitions for Sepsis and Septic Shock in 2016 [4]. This new paradigm, emphasizing a clearer correlation between infection and subsequent organ failure, calls for fresh avenues of research. It not only reshapes diagnostic and treatment approaches but also provides clinicians and researchers with a refined framework for identifying and analyzing Sepsis cases accurately. Familiarity with this contemporary approach is indispensable for the development of effective diagnostic and therapeutic strategies, empowering healthcare professionals to confront this formidable medical challenge more effectively.

Previously, methods for assessing Sepsis severity and mortality risk relied heavily on tools like the Simplified Acute Physiology Score-II (SAPS-II) [5], a severity-of-disease classification system primarily based on physiological data collected within the first 24 hours of ICU admission. However, the limitations of SAPS-II, particularly its susceptibility to missing data and rapid changes in patient condition post-admission, pose challenges in the dynamic ICU setting. Other methods, such as calculating a Sequential Organ Failure Assessment (SOFA) score [6], may suffer from subjectivity and interrater variability, leading to inconsistent results.

In addition to traditional scoring methods, conventional statistical models like Logistic Regression [7] have been widely used for outcome prediction in Sepsis. However, these models often struggle to capture the intricate, non-linear relationships inherent in medical data. Moreover, they rely on assumptions about data distribution that are rarely met in medical contexts, leading to suboptimal predictions. The inadequacy of these models underscores the need for more advanced analytical techniques.

In recent years, ML models have emerged as promising alternatives, particularly for handling high-dimensional and unnormalized data [8-13]. Due to its unique characteristics such as efficiency, accuracy, and the ability to handle large datasets, LightGBM [14] stands out. Leveraging those ensemble learning techniques, LightGBM sequentially builds decision trees to correct errors and improve predictive performance. Despite the increasing use of ML algorithms to predict mortality in ICU patients with Sepsis [15,16], none have yielded satisfactory results, potentially due to poor feature selection methodologies and inadequate parameter tuning.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

To address these challenges, our proposed model incorporates a novel decision tree-based entropy analysis [17] for feature selection, identifying significant factors for mortality prediction. This approach enhances computational efficiency and identifies hidden relationships in complex datasets, offering a more nuanced and precise approach to medical prediction.

Our study aims to demonstrate the clinical applicability of this feature engineering process and evaluate the predictive performance of proposed model in mortality prediction for Sepsis-3 patients. Additionally, the learning curve of our proposed model is plotted to validate its generalization and predictive accuracy. Our prediction model complies with the standards of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) initiative, guaranteeing thorough and transparent reporting [18,19].

METHODS

Data Availability

The Medical Information Mart for Intensive Care III (MIMIC-III) is a comprehensive dataset, available to the public via https://physionet.org/content/mimiciii/1.4/, which includes de-identified health information from more than 40,000 ICU admissions at the Beth Israel Deaconess Medical Center from 2001 to 2012 [20]. Created by the MIT Lab for Computational Physiology, MIMIC-III encompasses diverse data categories such as demographics, vital signs, laboratory test results, medications, and mortality outcomes. This extensive dataset enables multifaceted research in clinical informatics.

Patient Selection

We initially included patients who were classified as “Sepsis,” “severe Sepsis,” and “septic shock.” To exclude incomplete and repeated data, we have further narrowed data, as illustrate in Figure 1, adhering to specific inclusion criteria: (I) patients aged 18 years or older; (II) patients lacking demographic and lab test results and with fewer than 20% of features missing; (III) patients with SOFA scores.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

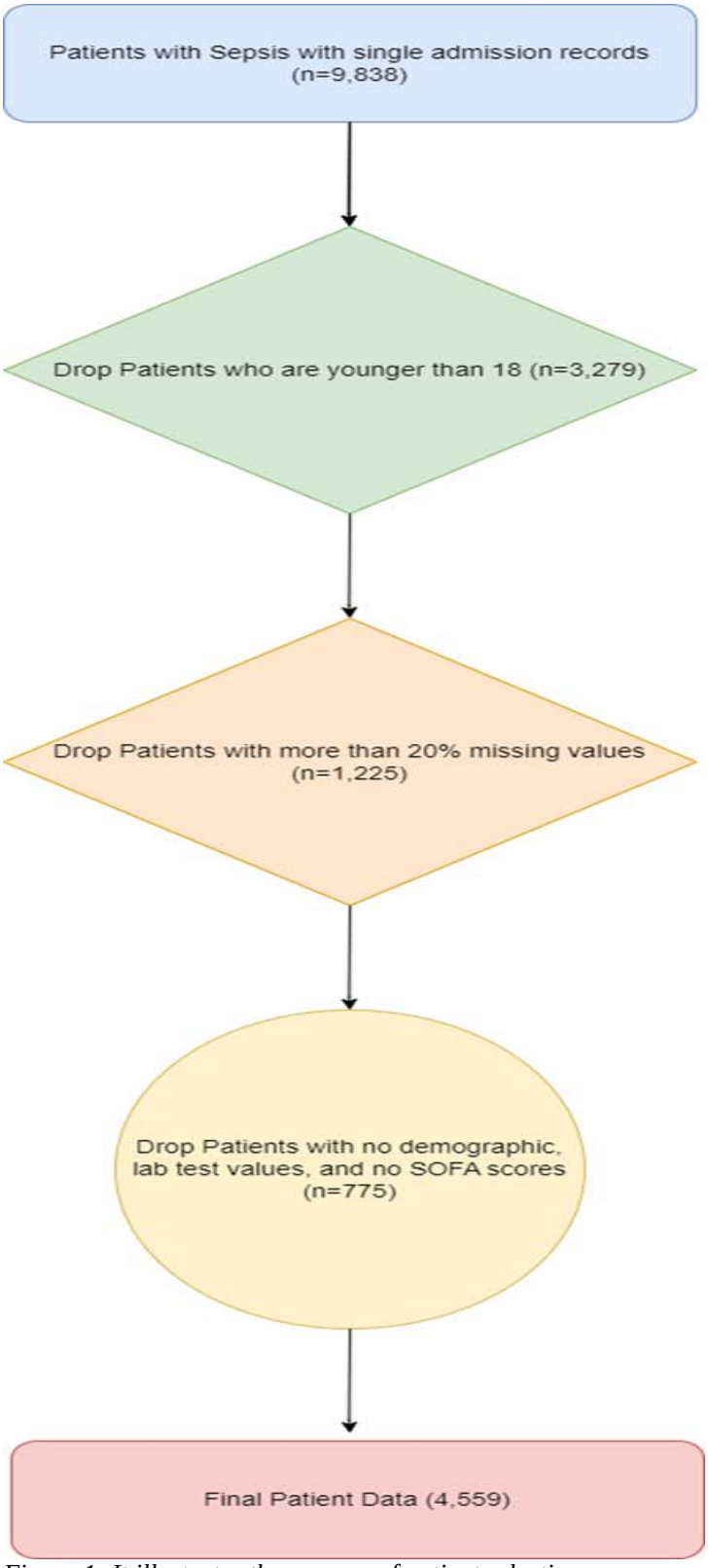

Figure 1: It illustrates the process of patient selection

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Feature Selection and Pre-processing

The feature selection process of the research unfolds in two distinct stages. Firstly, we employed entropy analysis using Decision Trees to discern and filter out the most significant features. Secondly, we sought the input of clinical medical experts to validate and refine the selection. Initially, drawing from existing literature and expert insights, demographic data including age, gender, ethnicity, weight, height, and body mass index (BMI), along with hospital and ICU lengths of stay and in-hospital mortality status, were extracted from initial ICU admission records. Vital signs such as heart rate (HR), mean arterial pressure (MAP), temperature (TEMP), respiratory rate (RR), and oxyhemoglobin saturation (SpO2) were recorded from the first 24 hours of ICU admission. Additionally, laboratory values encompassing blood routine examination, liver and kidney function, blood glucose, and arterial blood gas (ABG) measurements were abstracted. Given the high sampling frequency, maximum, minimum, and mean values were utilized to incorporate vital signs and related laboratory indicators effectively.

Subsequently, employing entropy analysis based on decision trees with a threshold of 30, we refined the feature set, resulting in 30 features selected for further analysis as shown in Table 1. Furthermore, owing to the constraints posed by the relatively modest final sample size (4,559), we implemented bootstrapping [21], a statistically robust resampling technique aimed at augmenting the volume and diversity of the original patient population.

Features Category	Features Table
Numerical Features	International Normalized Ratio_Max Anion Gap_Max Lactate Level_Max Potassium Level_Min White Blood Cell Count_Mean Blood Urea Nitrogen_Min Temperature_Min White Blood Cell Count_Min Platelet Count_Min Anion Gap_Min Platelet Count_Max Oxygen Saturation Level_Mean Urine Output Temperature_Max Blood Pressure_Min Blood Urea Nitrogen_Mean Systolic Blood Pressure_Mean Oxygen Saturation Level_Min Respiratory Rate_Mean Systolic Blood Pressure_Min Age Lactate Level_Mean Lactate Level_Min Temperature_Mean ICU Length of Stay Hospital Length of Stay Logistic Organ Dysfunction System
Categorical Features	Sodiumn Level_Max Elixhauser Comorbidity Index Sequential Organ Failure Assessment

Table 1: A summary of numerical and categorical features.

During the data preprocessing stage, we employed data imputation techniques. Leveraging random data from the dataset to populate missing values serves to mitigate biases and minimize information loss inherent in incomplete datasets [22]. Moreover, this approach

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

generates additional samples for model training, thereby enhancing its robustness. The Min-Max Scaler was employed to rescale numerical features, thereby normalizing them to a range of 0 to 1. This procedure plays a crucial role in ensuring that all numerical features contribute equally to the analysis, thereby mitigating biases resulting from features with larger scales. Categorical features underwent transformation using Label Encoder, which involves converting categorical labels into numerical codes, thereby enabling their integration into regression and ML models. These preprocessing techniques serve to standardize the dataset, a fundamental prerequisite for efficient model training and evaluation. Furthermore, they ensure that analyses accurately reflect the original measurements and categories present in the dataset. In conclusion, the complete process is demonstrated in Figure 2.

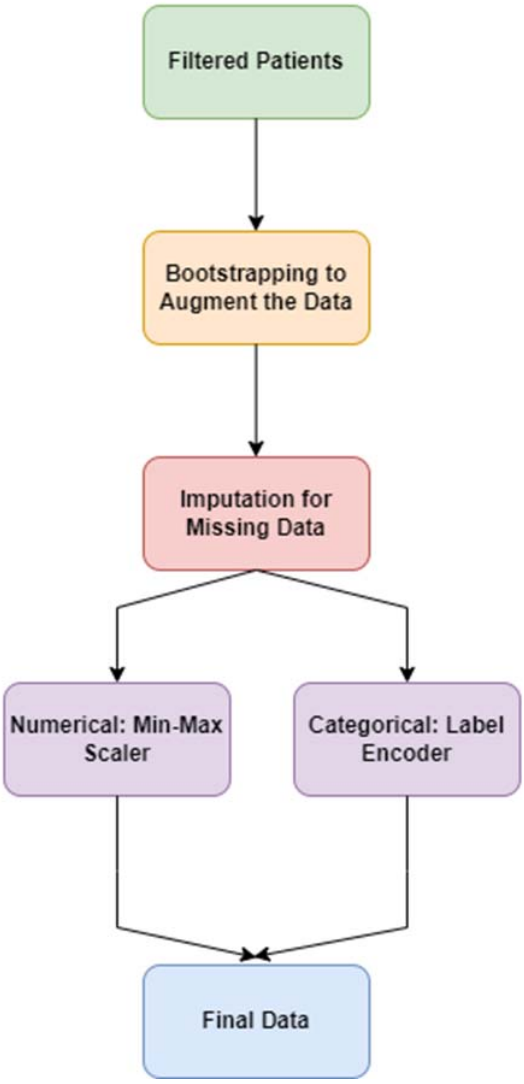

Figure 2: It presents the work flow of data preprocessing

Model Development and Optimization

Our final dataset encompasses 9,118 patients with 30 features. A train-test split is executed with an 80/20 ratio to facilitate model evaluation. To mitigate overfitting, we utilize Grid

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Search CV to identify the optimal combination of hyperparameters. We construct various ML algorithms, including Logistic Regression, LightGBM, CatBoost [23], Random Forest [24], K-Nearest Neighbors (KNN) [25], Support Vector Machine (SVM) [26], and Extra Gradient Boosting (XGBoost) [27]. The training process of these models included a grid search of model parameters. This search process aimed to find the best model which was determined based on the Area Under the Receiver Operating Characteristic (AUROC) scores of the cross-validation cohort. Accuracy and F1 scores are also computed for comparative analysis of model performance. Given the widespread adoption of AUC as an evaluation metric in existing literature, the selection of the proposed model is based on AUC on the cross-validation. LightGBM emerges as the top performer, consistent with our expectations of its superior performance compared to other models. Since we observed a notably high importance score for the feature ho HOSP_LOS during entropy analysis, we calculated the AUC scores of the proposed model with and without this feature. An overview of the methodologies employed is illustrated in Figure 3.

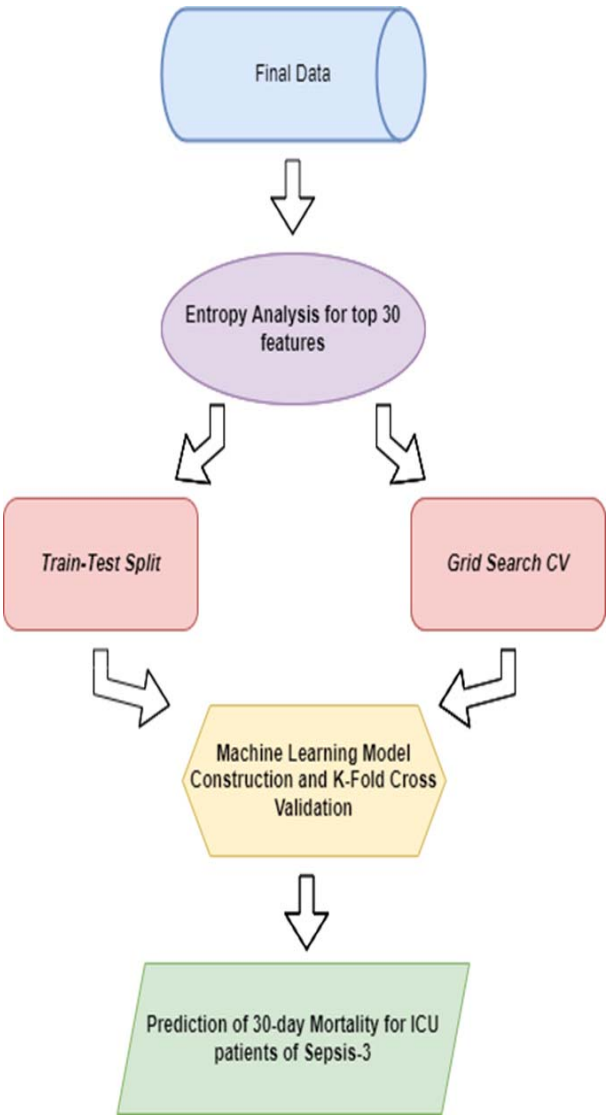

Figure 3: It elaborates our novel feature engineering methodologies and model construction process.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Statistical Analysis of Models

To validate the statistical robustness of our model results, we employed comprehensive statistical tests, utilizing diverse criteria to evaluate overall performance.

To ascertain whether these AUC scores were statistically different, we conducted the Mann-Whitney U Test (Wilcoxon Rank-Sum Test) [28]. Unlike the Student's t-test, the Mann-Whitney U Test does not require assumptions about the underlying dataset distribution, making it more suitable for our analysis. The null hypothesis posits that the AUC scores with and without HOSP_LOS are not statistically different, while the alternative hypothesis suggests AUC scores are significantly different.

Lastly, we conducted a statistical analysis on our train/validation dataset to compare their cumulative distributions. Utilizing the Kolmogorov-Smirnov test for its non-parametric nature, we made no assumptions about the specific distribution of the data [29]. This is particularly important as some features in our dataset may not adhere to a normal distribution. With a predetermined significance level of 0.05, our null hypothesis assumes no statistically significant difference between the test and validation sets.

Feature Impacts

To deepen our analysis, we utilized SHapley Additive exPlanations (SHAP) [30] analysis to evaluate feature importance and elucidate the decision-making mechanisms of the predictive models, particularly within the framework of random forests. This advanced technique quantifies the influence of each feature on the model's predictions, offering valuable insights into the reasoning behind specific predictions.

RESULTS

Cohort Characteristics Model Completion

Following our approach for feature selection and data augmentation for ICU patients with Sepsis-3 discussed before in this paper, our final dataset contained 9118 patients from MIMIC-III database. The selected cohort was then split into train/test cohorts randomly with a ratio of 80/20, which yielded a result of 7294 patients for train and 1824 patients for the test cohorts. Moreover, the training cohort was used to train the model, Grid Search CV and K-fold cross-validation were then used to identify the optimal parameters and to validate our model, and the testing cohort was used to evaluate the performance of our proposed model. Furthermore, the best model was chosen based on its AUC performance on the test set. The mortality for train and test were 19.5% and 20% respectively, out of 7294 patients in the training cohort, 1422 of them did not survive within 30-day, out of 1824 patients in the testing cohort, 365 of them did not survive. The Mann-Whitney U Test is performed for the AUC results of LightGBM with and without HOSP_LOS, and the P-value is 9.182598395744396e-89. Thus, our null hypothesis for The Mann-Whitney U Test was rejected, suggesting that AUC scores are significantly different. The Kolmogorov-Smirnov test for training and test data distribution is also performed. The null hypothesis is that data distributions from training and test are the same, and the alternative hypothesis is that test and training datasets come from different distributions. Significance level is pre-determined to be 0.05. The P-value for Hospital Length of Stay (HOSP_LOS) is 0.1062, suggesting there were

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

no significant differences between cohorts. In terms of Logistic Organ Dysfunction System (LODS), a scoring system used to assess the severity of organ dysfunction in critically ill patients which is a typical symptom for ICU patients with Sepsis-3, P value is 0.1648, showing that there were no statistically significant differences between cohorts either. However, P-value for age and P-value for average white blood cell counts (WBC_mean) did not exceed the significance threshold of 0.05, suggesting statistically significant difference in distribution between training and test cohorts. Details are illustrated in Table 2, which indicates no statistically differences between cohorts, except for age and WBC_mean. This ensures robust model generalization and performance across datasets.

Feature	p-value
age	0.0106
wbc_mean	0.0122
platelet_max	0.0501
urineoutput	0.0552
bun_min	0.0654
bun_mean	0.0672
platelet_min	0.0988
wbc_min	0.1046
hosp_los	0.1062
sodium_max	0.14
lods	0.1648
sofa	0.2113
sysbp_mean	0.2669
aniongap_min	0.2918
tempc_max	0.3161
lactate_min	0.3346
sodium_min	0.4262
tempc_mean	0.436
bicarbonate_min	0.436
lactate_mean	0.5951
icu_los	0.6063
inr_max	0.6949
lactate_max	0.7786
resprate_mean	0.8366
meanbp_min	0.8804
sysbp_min	0.8853
spo2_mean	0.9132
tempc_min	0.93
spo2_min	0.9766
elixhauser_hospital	0.9999

Table 2: It illustrates the P-values of train/test cumulative distributions for all the significant features.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Evaluation Metrics, Proposed and Baseline Models Performance

The summary of the results for both proposed and baseline models are shown in Table 3. The proposed approach resulted in the following metrics, AUC = 0.983, 95% CI = [0.980-0.990] as shown in Figure 4, accuracy score = 0.966, F1 score = 0.910, underscoring not only its accuracy but also its robustness in minimizing both Type I and Type II errors, thereby affirming its suitability for our predictive modeling tasks.

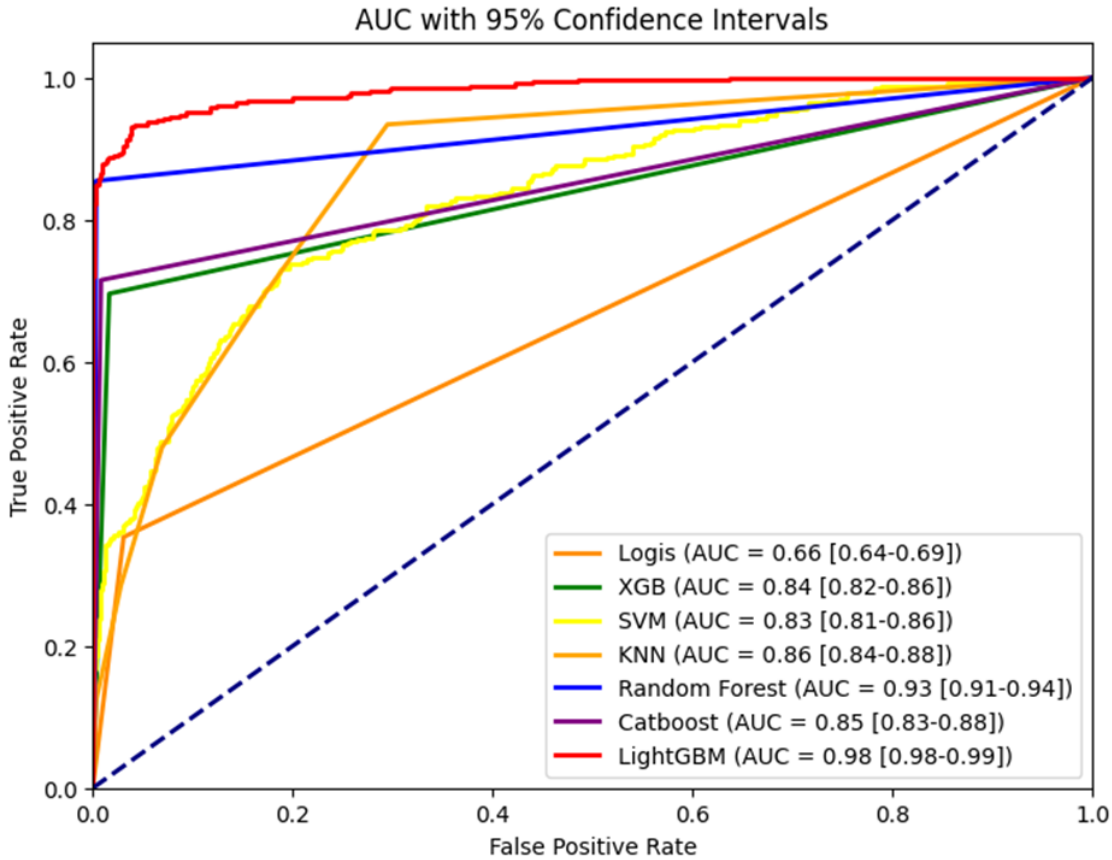

Figure 4: A comparison of the ROC AUCs for all Models

On the other hand, the baseline model development utilizing the MIMIC-III database, Random Forest proved to be the best baseline model. Random Forest resulted in the following metrics, AUC = 0.926, 95% CI = [0.910,0.940], accuracy score = 0.968, F1 score = 0.915. It can be observed that the results of the proposed approach are far better than the results of the best baseline model.

Shapley Value Analysis

Figure 5 illustrates the SHAP value results. Based on this figure, HOSP_LOS had the most significant impact on the prediction of mortality for ICU patients with Sepsis-3, followed by ICU Length of Stay (ICU_LOS) and age. Intriguingly, the relationship between HOSP_LOS and mortality appears to be negative, suggesting that extended hospital stays may correlate with lower mortality rates among ICU patients with Sepsis-3. Furthermore, the analysis

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

indicates a positive contribution of ICU_LOS to mortality prediction, as evidenced by the proliferation of red dots towards the right. This implies that prolonged ICU stays may elevate mortality risks for ICU patients with Sepsis-3. Moreover, age and the Elixhauser Comorbidity Index (ELIXHAUSER_HOSPITAL) exhibit positive associations with prediction accuracy. This aligns with intuition, as advanced age and greater comorbidity burden typically elevate patient vulnerability and subsequent mortality risks. Whereas Minimum of Lactate Level (LACTATE_MIN) and Average Blood Urea Nitrogen level (BUN_MEAN) were the least important features for the prediction of mortality.

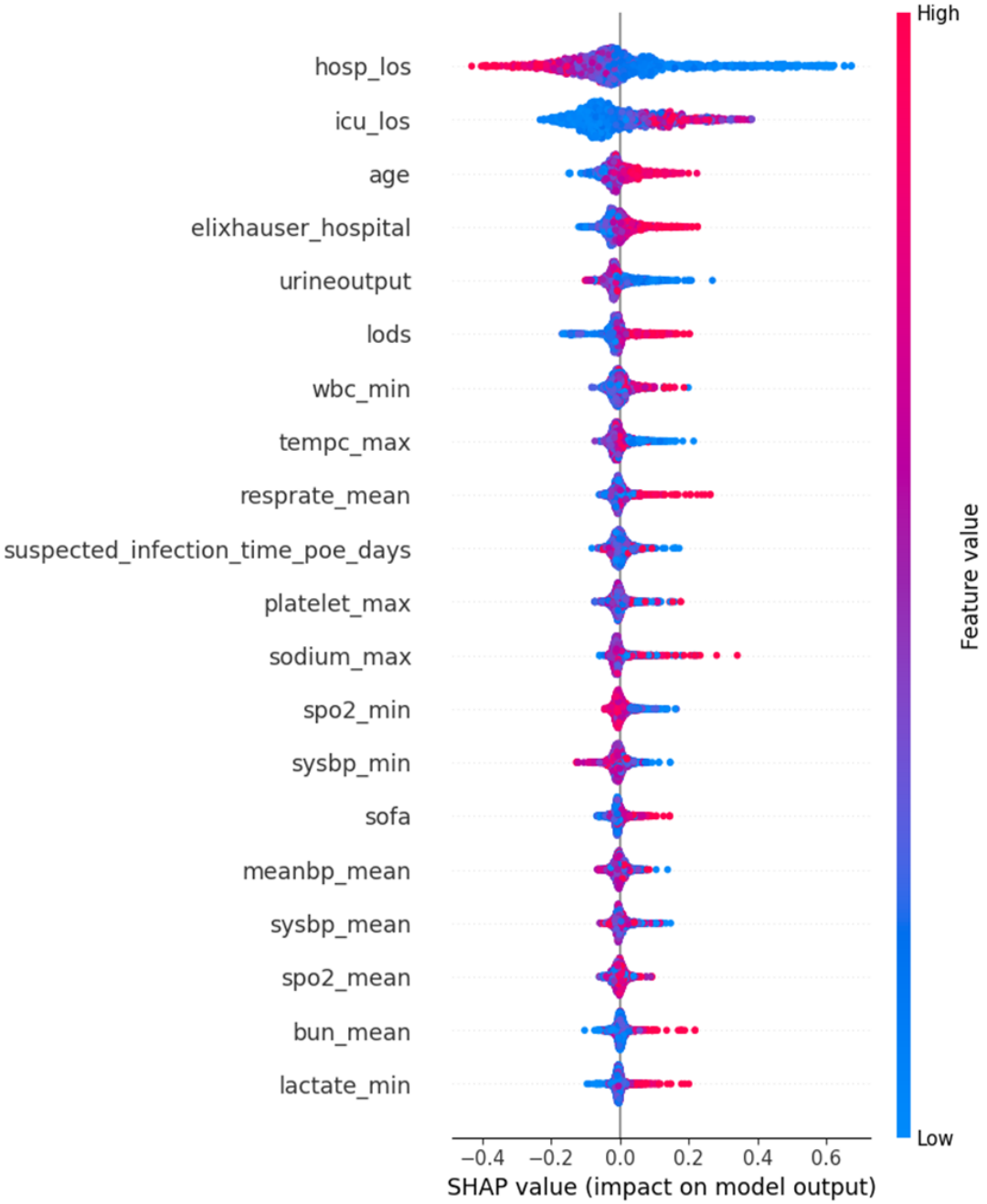

Figure 5: It illustrates the contribution of each feature to the model's prediction for a specific instance.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

LightGBM's Learning Curve Analysis

Additionally, to evaluate LightBGM's ability to handle unseen data, its learning curve has been plotted by software Python. Binary error, the proportion of incorrectly classified instances in the dataset, is decreasing as boosting rounds, sequential training of individual decision trees, rises, which is implying model's classification accuracy. The convergence of both the training and testing curves has been observed. As depicted in Figure 6, a decreasing training error accompanied by a decreasing testing error indicates that LightGBM is learning effectively without overfitting.

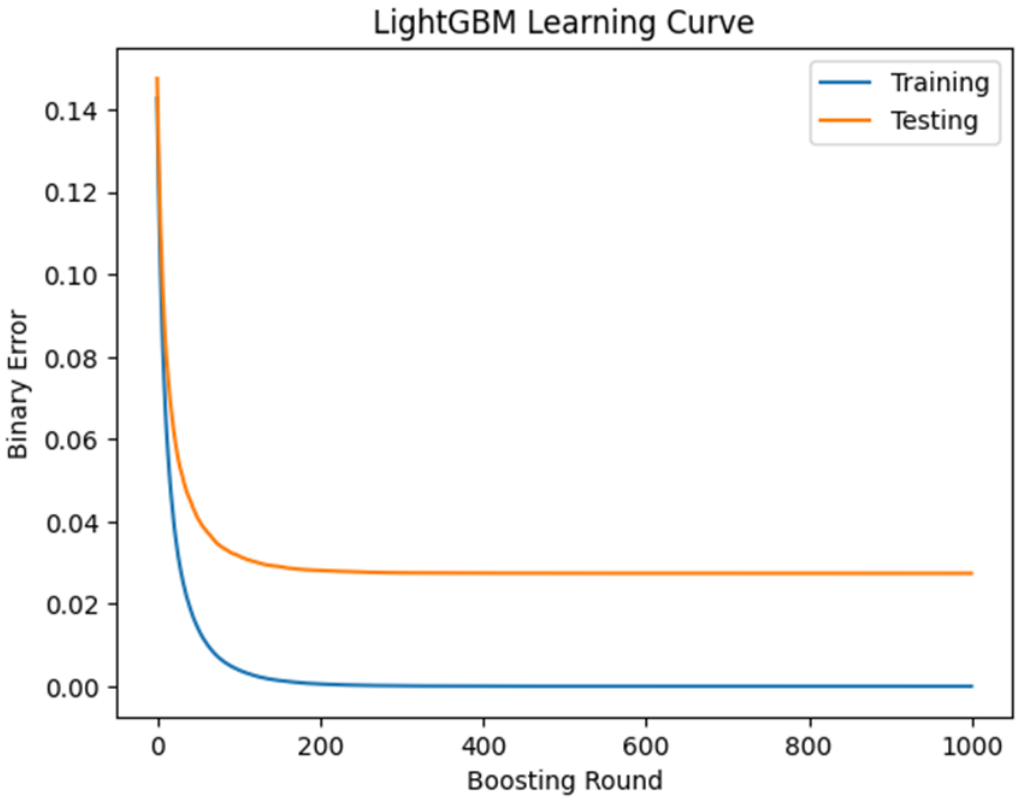

Figure 6: Error curve for training and validation data as tested on the LightGBM model.

DISCUSSION

Existing Model Compilation Summary

Several methods have been concurrently developed to predict mortality for ICU patients with Sepsis [13,16]. However, to the best of our knowledge there is only one research focusing on prediction of mortality for ICU patients of Sepsis-3 using MIMIC-III database [15]. The paper proposed XGBoost as the best model, which yielded in the following results: AUC = 0.857, 95% CI = [0.839-0.876].

In our study, novel data preprocessing techniques have been utilized. Bootstrapping, a statistical resampling method, is applied to augment the limited sample size, doubling our dataset's size to 9118 samples.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Our novel feature selection approach, entropy analysis using Decision-Tree, has allowed our proposed model to produce more accurate and robust results. Our approach has resulted in an almost 15% Improvements of AUC metric. Our proposed model, LightGBM, has distinct features such as Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), enabling it to be more computationally efficient.

In addition, our approach has also enabled us to find the most statistically significant factor in prediction of mortality, which is HOSP_LOS. Mann-Whitney U Test is also performed to prove that the absence of HOSP_LOS results in statistically different AUC score, and this feature is not included in the best existing literature [15].

Although the existing proposed methodology in the literature was successful in predicting 30-day mortality for ICU patients of Sepsis-3, it possessed several drawbacks. First, they have failed to consider more efficient feature engineering techniques such as Entropy Analysis, thus it ignored the most significant feature in predicting mortality which is HOSP_LOS. Secondly, the existing literature has a wide confidence internal for its AUC metric. A wide confidence interval can be a consequence of an inadequate sample size used for evaluation.

Our proposed approach had several advantages over prior research papers which are as follows: (a) Novel preprocessing and feature engineering techniques collectively augmented the predictive performance of our models, leading to substantial advancements over those of existing literatures. (b) Identification of a statistically significant feature in the prediction of mortality which was ignored by other research papers. (c) Choice of LightGBM and Grid Search CV. LightGBM is more suitable at handling large-scale and high dimensional datasets with better computational efficiency. All of those have led an enhancement of nearly 10% compared to the existing literature as shown in Table 3. This substantial increase underscores the efficacy of our preprocessing strategies and feature engineering in refining model performance and underscores the potential for significant advancements in mortality prediction within the medical domain.

	LightGBM	Random Forest	KNN	CatBoost	XGB	SVM	Logistic Regression
AUC	0.983	0.926	0.863	0.853	0.840	0.832	0.661
Accuracy Score	0.966	0.968	0.840	0.936	0.926	0.837	0.846
F1 Score	0.910	0.915	0.546	0.818	0.790	0.339	0.479

Table 3: A summary of various evaluation metrics for the models

Study Limitation

Our study has two main limitations. Firstly, we were unable to validate our model using external datasets due to lack of having access to comprehensive databases like MIMIC-III. The absence of external data hampers our ability to confirm the effectiveness of our proposed model. It opens room for future researchers to validate our models externally by having comprehensive data ready.

Secondly, the MIMIC-III database, which we used, is more than 10 years old, and it lacks historical information for many patients. This could introduce bias when selecting our

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

datasets. It opens room for newer database. To address this concern in future research, we aim to utilize newer databases to mitigate potential biases.

CONCLUSION

Utilizing bootstrapping, entropy analysis via Decision-Tree algorithms, data imputation strategies, and model optimization techniques not only enhances predictive model accuracy and robustness but also improves generalization to unseen data. This is evident from notable improvements in AUC and Accuracy metrics, outperforming existing methodologies. These advancements underscore the importance of feature selection for computational efficiency and highlight the potential of ML as a valuable tool in the dynamic ICU environment, where precise predictive models are crucial.

In future studies, the proposed approach could be validated using datasets from other healthcare systems to ensure its applicability and robustness across different populations and settings. Moreover, incorporating the comprehensive information available in the MIMIC-III dataset, such as clinical notes and images, as inputs for models holds promise for further research and exploration. These rich data sources could provide additional context and features, potentially leading to even more accurate and reliable predictive models. By leveraging diverse data types and sources, we aim to develop models that not only perform well in controlled settings but also demonstrate strong generalizability and reliability in real-world clinical environments.

Ultimately, through ongoing research efforts and the integration of varied datasets and advanced data types, we anticipate further enhancing the model's generalizability and performance. This continuous improvement is essential to developing reliable tools that can aid clinicians in making informed decisions, thereby improving patient outcomes in intensive care units and beyond.

ABBREVIATIONS

ML: Machine Learning; ICU: Intensive Care Unit; SAPS-II: Simplified Acute Physiology Score-II; SOFA: Sequential Organ Failure Assessment; AUROC: Average Area Under the Receiver Operating Characteristic Curve

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Acknowledgements

The authors extend their gratitude to the developers of MIMIC-III for offering a detailed and comprehensive public EHR dataset.

Funding

No funding was involved in this research.

Author information

Authors and Affiliations

Department of Industrial and Systems Engineering, University of Southern California (USC), Andrus Gerontology Center, 3715 McClintock Ave GER 240, Los Angeles, CA 90089
Yu Z, Li H, Pishgar M

Authors' contributions

Z.Y., M.P.: Involved in all aspects of this study. H.L.: Revision of the manuscript.

Declaration

Availability of data and materials

The MIMIC-III database which was used during the current study is publicly available. The Medical Information Mart for Intensive Care III (MIMIC-III) is a comprehensive dataset, available to the public via https://physionet.org/content/mimiciii/1.4/

Ethics approval and consent to participate

The dataset used to support the conclusions of this article is sourced from the Medical Information Mart for Intensive Care version III (MIMIC-III). As this database is public and de-identified, informed consent and Institutional Review Board approval were not required. All procedures followed the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

REFERENCES

[1] Fleischmann-Struzek C, Rudd K. Challenges of assessing the burden of sepsis. Med Klin Intensivmed Notfmed. 2023 Dec;118(Suppl 2):68-74. doi: 10.1007/s00063-023-01088-7.
[2] Evans T. Diagnosis and management of sepsis. Clin Med (Lond). 2018 Mar;18(2):146-149. doi: 10.7861/clinmedicine.18-2-146. PMID: 29626019; PMCID: PMC6303466.
[3] Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, Colombara DV, Ikuta KS, Kissoon N, Finfer S, Fleischmann-Struzek C, Machado FR, Reinhart KK, Rowan K, Seymour CW, Watson RS, West TE, Marinho F, Hay SI, Lozano R, Lopez AD, Angus DC, Murray CJL, Naghavi M. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet. 2020 Jan 18;395(10219):200-211. doi: 10.1016/S0140-6736(19)32989-7.
[4] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):801-10. doi: 10.1001/jama.2016.0287. PMID: 26903338; PMCID: PMC4968574.
[5] Kajdacsy-Balla Amaral AC, Andrade FM, Moreno R, Artigas A, Cantraine F, Vincent JL. Use of the sequential organ failure assessment score as a severity score. Intensive Care Med. 2005 Feb;31(2):243-9. doi: 10.1007/s00134-004-2528-6. Epub 2005 Jan 25. PMID: 15668764.
[6] Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. 1996 Jul;22(7):707-10. doi: 10.1007/BF01709751. PMID: 8844239.
[7] Ambrish G, Bharathi Ganesh, Anitha Ganesh, Chetana Srinivas, Dhanraj, Kiran Mensinkal, Logistic regression technique for prediction of cardiovascular disease, Global Transitions Proceedings, Volume 3, Issue 1, 2022, Pages 127-130, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2022.04.008.
[8] Pishgar M, Theis J, Del Rios M, Ardati A, Anahideh H, Darabi H. Prediction of unplanned 30-day readmission for ICU patients with heart failure. BMC Med Inform Decis Mak. 2022 May 2;22(1):117. doi: 10.1186/s12911-022-01857-y. PMID: 35501789; PMCID: PMC9063206.
[9] Pishgar M, Razo M, Theis J, Darabi H. Process mining model to predict mortality in paralytic ileus patients. 2021 International Conference on Cyber-Physical Social Intelligence (ICCSI). 2021 Dec 18;1–6. doi:10.1109/iccsi53130.2021.9736217.
[10] Liu Y, Xu X, Wang Y, Zhao Z, Pishgar M. Machine learning model utilization for mortality prediction in mechanically ventilated ICU patients. 2024 Mar 22; doi:10.1101/2024.03.20.24304653.
[11] Pishgar M, Harford S, Theis J, Galanter W, Rodríguez-Fernández JM, Chaisson LH, Zhang Y, Trotter A, Kochendorfer KM, Boppana A, Darabi H. A process mining- deep learning approach to predict survival in a cohort of hospitalized COVID-19 patients. BMC Med Inform Decis Mak. 2022 Jul 25;22(1):194. doi: 10.1186/s12911-022-01934-2. PMID: 35879715; PMCID: PMC9309593.
[12] Pishgar M, Harford S, Theis J, Galanter W, Rodríguez-Fernández JM, Chaisson LH, Zhang Y, Trotter A, Kochendorfer KM, Boppana A, Darabi H. A process mining- deep learning approach to predict survival in a cohort of hospitalized COVID-19 patients. BMC

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Med Inform Decis Mak. 2022 Jul 25;22(1):194. doi: 10.1186/s12911-022-01934-2. PMID: 35879715; PMCID: PMC9309593.
[13] Gao J, Lu Y, Domingo IR, Alaei K, Pishgar M. Predicting sepsis mortality using machine learning methods. 2024 Mar 16; doi:10.1101/2024.03.14.24304184.
[14] 1. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Neural Information Processing Systems [Internet]. 2017 Dec 4; Available from: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.
[15] Hou N, Li M, He L, Xie B, Wang L, Zhang R, Yu Y, Sun X, Pan Z, Wang K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. J Transl Med. 2020 Dec 7;18(1):462. doi: 10.1186/s12967-020-02620-5. PMID: 33287854; PMCID: PMC7720497.
[16] Su Y, Guo C, Zhou S, Li C, Ding N. Early predicting 30-day mortality in sepsis in MIMIC-III by an artificial neural networks model. Eur J Med Res. 2022 Dec 17;27(1):294. doi: 10.1186/s40001-022-00925-3. PMID: 36528689; PMCID: PMC9758460.
[17] Lee MY, Yang CS. Entropy-based feature extraction and decision tree induction for breast cancer diagnosis with standardized thermograph images. Comput Methods Programs Biomed. 2010 Dec;100(3):269-82. doi: 10.1016/j.cmpb.2010.04.014. PMID: 20537756.
[18] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015 Jan 6;162(1):W1-73. doi: 10.7326/M14-0698. PMID: 25560730.
[19] Amritphale A, Chatterjee R, Chatterjee S, Amritphale N, Rahnavard A, Awan GM, Omar B, Fonarow GC. Predictors of 30-Day Unplanned Readmission After Carotid Artery Stenting Using Artificial Intelligence. Adv Ther. 2021 Jun;38(6):2954-2972. doi: 10.1007/s12325-021-01709-7. Epub 2021 Apr 9. PMID: 33834355; PMCID: PMC8190015.
[20] Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Sci Data. 2016 May 24;3:160035. doi: 10.1038/sdata.2016.35. PMID: 27219127; PMCID: PMC4878278.
[21] Taleb I, Kassabi HT, Serhani MA, Dssouli R, Bouhaddioui C. Big Data Quality: A quality dimensions evaluation. 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld). 2016 Jul; doi:10.1109/uic-atc-scalcom-cbdcom-iop-smartworld.2016.0122.
[22] Kaiser J. Dealing with missing values in data. Journal of Systems Integration. 2014;42–51. doi:10.20470/jsi.v5i1.178.
[23] 1. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Neural Information Processing Systems [Internet]. 2017 Jun 28; Available from: https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf.
[24] Breiman L. Machine Learning. 2001;45(1):5–32. doi:10.1023/a:1010933404324.
[25] Kramer O. K-Nearest Neighbors. Dimensionality Reduction with Unsupervised Nearest Neighbors. 2013;13–23. doi:10.1007/978-3-642-38652-7_2.
[26] Adankon MM, Cheriet M. Support Vector Machine. Encyclopedia of Biometrics. 2015;1504–11. doi:10.1007/978-1-4899-7488-4_299.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.27.24308004; this version posted June 7, 2024. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

[27] Chen T, Guestrin C. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Aug 13; doi:10.1145/2939672.2939785.
[28] McKnight PE, Najab J. Mann – Whitney U Test. The Corsini Encyclopedia of Psychology. 2010 Jan 30; 1-1. Doi:10.1002/9780470479216.corpsy0524.
[29] Musakkir NA, Sunusi N, Thamrin SA. Stochastic model of the annual maximum rainfall series using probability distributions. Malaysian Journal of Fundamental and Applied Sciences. 2023 Oct 19;19(5):827–39. doi:10.11113/mjfas.v19n5.2945.
[30] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Neural Information Processing Systems. 2017 May 22.