Çanakkale Onsekiz Mart University
Journal of Advanced Research in Natural and Applied Sciences
dergipark.org.tr/tr/pub/jarnas
Open Access

e-ISSN: 2757-5195 2024, Vol. 10, Issue 4, Pages: 819-832 doi.org/10.28979/jarnas.1533962

Machine Learning and Medical Data: Predicting ICU Mortality and Re-admission Risks

Ulya Bayram[1], Runia Roy[2]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Çanakkale Onsekiz Mart University, Çanakkale, Türkiye
[2]Neural Stem Cell Institute, New York, USA

Article Info
Received: 15 Aug 2024
Accepted: 12 Dec 2024
Published: 31 Dec 2024
Research Article

Abstract— Intensive care units (ICUs) are divisions where critically ill patients are treated by medical experts. The unmet and vital need for automated clinical decision-making mechanisms is critical to maneuvering the large influx of patients. This became more apparent after the COVID-19 pandemic. Existing studies focus on determining the probability of patients dying in the ICUs and prioritizing patients in dire need. Only a few studies have calculated the patient's probability of returning to the ICUs after discharge. These studies reduce the problem into a binary task of predicting mortality or re-admission only. However, this is unrealistic since both outcomes are highly possible for each patient. In this interdisciplinary study, two main contributions are proposed for the automated clinical decision-making state-of-the-art: (1) using the real-life data collected from thousands of ICU patients by healthcare professionals, three possibilities (recovery, mortality, and returning to the intensive care unit within 30 days) are predicted for patients in intensive care instead of just one possibility. (2) A novel feature extraction approach is proposed by the biomedical expert in our team. Four machine learning algorithms are applied to the finalized feature set to understand the difference between the binary and the multi-class classification problems. Obtained results reach 78% success, proving the possibility of developing better clinical decision-making mechanisms for ICUs.

Keywords — *Clinical decision making, machine learning, intensive care units, mortality prediction, re-admission prediction*

1. Introduction

It is well recognized that many of the mortality cases in intensive care units (ICUs) were preventable if and only if the deteriorating decline of the patient could have been noticed at the right time [1, 2]. However, there are too few healthcare professionals with ICU expertise in the hospitals. Furthermore, since these professionals work for long hours, it becomes difficult to track the progress of every patient 24/7, non-stop. To overcome this challenge and reduce potentially dire consequences, healthcare professionals developed scoring systems specifically for the ICU. These systems involve healthcare professionals' manually noting the patient's condition obtained through measurements of vital signs and laboratory test results under certain categories, where the professional is required to access the

[1]ulya.bayram@comu.edu.tr (Corresponding Author); [2]rroy7@binghamton.edu

state of the patient and assign the correct score for each category, then obtaining a mathematical score by summing the noted scores from all categories, named the Acute Physiology and Chronic Health Evaluation (APACHE) scoring method. Based on 12 physiological parameters and the Glasgow coma score (GSC), the total score is 71. Any patient with scoring greater than 25 is considered to be at a higher risk for dying [3]. In real ICUs, this scoring is used for clinical decision-making - prioritizing care, intervention needs, monitoring trends over time, and resource and staff allocation [4].

Although still used today, methods like manual APACHE scoring have numerous drawbacks. It is difficult for healthcare employees under an intense workload to frequently visit every ICU patient, assign accurate scores, and repeat this process every 24 hrs or as the patients' conditions alter. Moreover, a statistical study found that scoring methods can make incorrect decisions depending on the patient's ethnicity [5]. Another study reports that these manual methods often calculate the patient's risk scores higher than the actual states [6], which means the patients received higher mortality scores, causing a shift of attention and workload from the actual high-risk patients towards less risky ones. This results in more frequent visits to less critical patients, thus wasting precious time that could have been dedicated to another, more risky patient. Such an error in the ICU can cause preventable deaths. This resource allocation and management issue in the ICUs became especially evident in early 2020 during the COVID-19 pandemic.

The in-hospital mortality rates in the United States published by the Centers for Disease Control and Prevention (CDC) [7] shows that in the United States alone, there was an unprecedented increase in in-hospital mortality rates in 2020, 2021, and 2022, while the rates started lowering in 2023 (Figure 1). Studies indicate that these mortality rate increases are not because of the coronavirus alone but are also due to the overcrowded ICUs, the inadequacy of the number of healthcare professionals, and their inability to allocate attention and time to the excessive number of ICU patients [8]. These rates only highlight the urgent need for the real-life use of automated clinical decision-making systems in this technology-driven century.

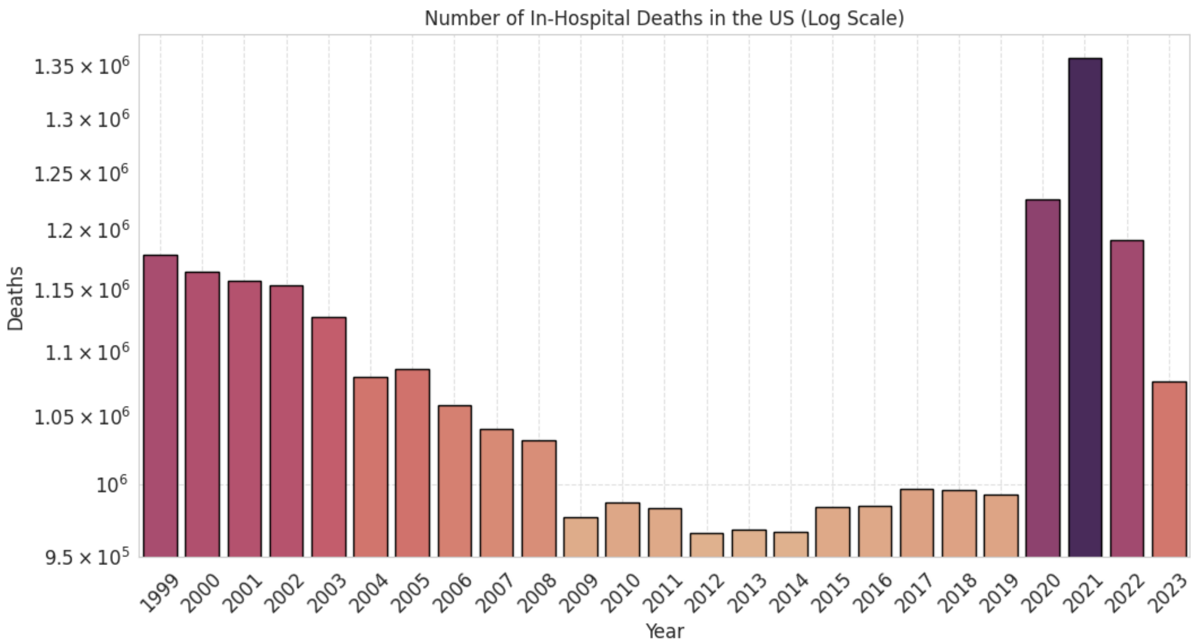


Figure 1. Annual hospital mortality rates in the United States of America.

To progress towards developing automated clinical decision-making systems, a team of researchers from interdisciplinary fields conducted various studies focusing on predicting specific pressing problems that can cause critical patients suffering from diseases such as circulatory failure, sepsis (blood poisoning), or kidney failure [9–12] in ICUs to lose their lives. Meanwhile, some studies showed that even small errors in dosages can be fatal for ICU patients [13, 14]. For example, if the circulatory failure is correctly predicted and intervened immediately, many lives can be saved. The machine learning methods these studies used to span from Logistic Regression (LR) to Support Vector Machines (SVM), LightGBM (LGB), and XGBoost (XGB). In addition to these specific predictive studies, researchers also attempted to model the general risk of ICU patients' mortality. For example, a study modeled the probability of patient deaths within 48 hours of ICU admission using a deep learning model called Long Short-Term Memory (LSTM) on doctor's notes [15]. These studies reached up to 80% success rates in predicting mortality in ICU.

Potential dire outcomes in an ICU environment are not restricted by patient mortality. Another critical problem is discharging ICU patients early. When an ICU patient is discharged early, before completing the treatment, or without following their progress a little longer, the condition of the patient can worsen without the care or attention of professionals, which might cause death to the patient. Alternatively, suppose the patient's condition is not as dire but has worsened after discharge. In that case, the patient will need to be re-admitted to the ICU; certain mandatory and expensive measurements will have to be repeated, and the treatments will need to be restarted in addition to handling the patient's worsened condition. This situation causes high financial losses and excessive use of medical resources [16]. To overcome this predicament and prevent follow-up and subsequent re-admissions, a few studies attempted to predict the likelihood of returning to the ICU within 30 days after getting discharged [17–19]. They used machine learning methods including LR, Naive Bayes, Random Forest (RF), SVM, Convolutional Neural Networks (CNN), and LSTMs and reported up to 75% success rates in predicting re-admission risk.

Thanks to the developments in artificial intelligence (AI) technology and the digitization of healthcare data, developing automated clinical decision-making systems is possible. However, existing studies show that there is still much more to achieve before the real-life use of these systems. First, they mainly rely on a binary classification task. Most studies focus on the easier task of mortality prediction, ignoring the re-admission risk. The same is true for the vice versa; re-admission prediction studies do not consider mortality prediction. Yet, a real-life clinical decision-making system for ICUs should be able to predict both risks at the same time. Second, existing studies do not conduct feature engineering, which is essential for obtaining explainable predictive systems. In this study, both issues are addressed.

The proposed study – to develop an improved and standardized clinical decision-making system for use in the ICU – introduces a novel feature extraction approach informed by biomedical expertise. Following a thorough feature engineering process, the study addresses the challenging multi-class classification problem of predicting mortality, re-admission, and survival risks together. Furthermore, the study accounts for data imbalance, which is how these cases are distributed in real life. Also, rather than deep learning techniques, the study used conventional machine learning methods and a few complex methods for several reasons. Considering the real-life application, lightweight methods such as traditional machine learning are preferable to heavy, data-demanding models. Plus, traditional methods perform as well as more complex methods in some scenarios [20, 21]. By focusing on feature engineering and leveraging domain expertise, the study maximizes the utility of conventional methods, proving them highly effective for this application.

In the next sections, the dataset and the selected methods are explained. Then, experiments and the obtained results are demonstrated, and discussions are conducted on the findings. Finally, conclusions are provided.

2. Materials and Methods

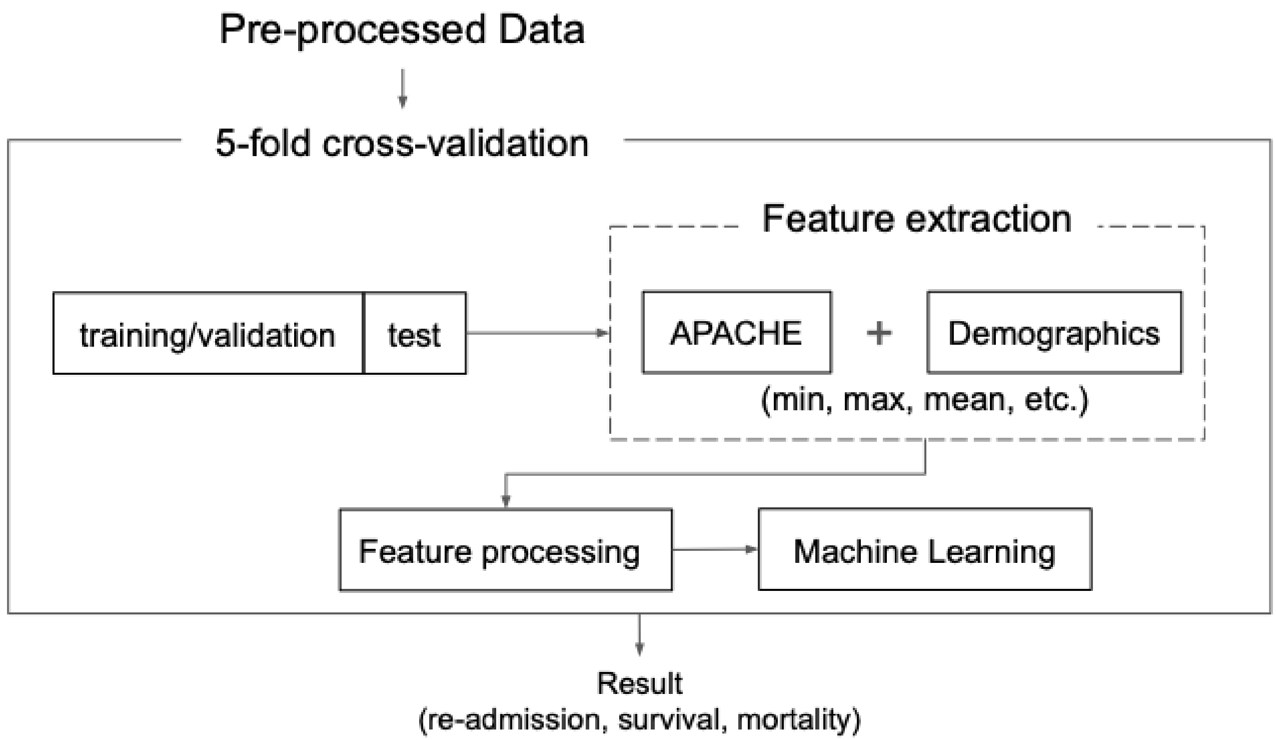


Figure 2. Work-flow of the present study.

The main workflow of the current study is present in Figure 2. Details of each step are explained in the following sections.

2.1. Data

This study uses the publicly available and popular Medical Information Mart for Intensive Care-III (MIMIC-III) dataset. This dataset contains various types of patient measurements collected from the ICU patients of Beth Israel Deaconess Medical Center in Boston, Massachusetts, between the years of 2001 and 2012 [22–24]. It contains information on each ICU patient, including demographics such as age, gender, and marital status. The entries were de-identified to protect the patients' identities, and information such as date of admission or date of birth was coded as a dummy number in the dataset.

For this study, patients outside the age range of 18 and 75 are excluded to develop a model for adults. After selecting only the adult patients, the number decreased from 47805 to 34969. This choice is made considering the possibility that older and younger age groups might need different clinical decision-making systems [25, 26]. There were also many erroneous data within the dataset. The most important problem among the whole dataset is the number of missing data. Some patients lack demographic information such as gender, some measurements were never obtained for some patients such as height or weight. Another critical issue is the lack of mortality and discharge dates for some patients altogether. After removing all these errors, the dataset contains 4253 ICU mortality, 2806 ICU re-admissions, and 30057 survivors who were discharged and did not get re-admitted to the ICU within 30 days after the discharge. Then, the complete dataset is divided into training and test sets following the common 80-20% division rate within a 5-fold cross-validation framework. Within each fold, to fine-tune the parameters of the complex machine learning methods, the training set of each fold is split into training and validation sets with the 80-20% division.

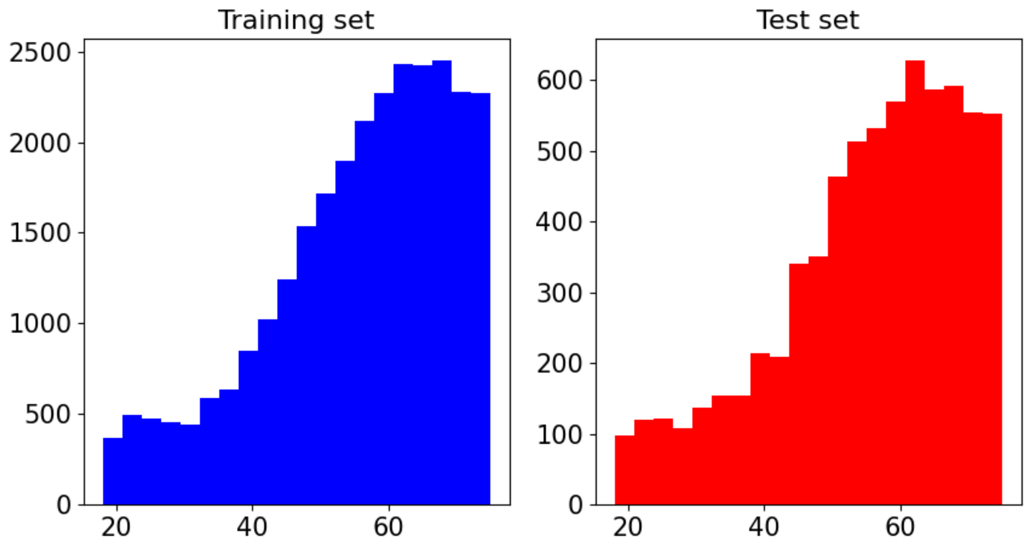


Figure 3. Age distributions in the training and test sets.

The age distributions of the remaining adult patients in the training and test sets are present in Figure 3. Distributions are similar between training and test sets, which shows the success of random sampling when dividing the dataset into these sets. Another visual finding from the age distributions is the higher number of elderly patients in the dataset compared to fewer young adults. Therefore, the number of deaths and re-admissions will be determined more heavily based on the conditions most older adults have in the ICUs.

Table 1. Sample size of the MIMIC-III dataset used in the experiments. (0: re-admission, 1: survival, 2: mortality)

	Re-admission (0)	Survival (1)	Mortality (2)
Training	2311	24006	3375
Test	495	6051	878

Another distribution obtained from the data is in Table 1. The table shows the high imbalance present in the training and test sets. Label 1 represents the dominant class, which is the recovery class. The remaining classes, 0 and 2, represent the re-admission and mortality classes, respectively. The number of patients in these two classes is similar. Therefore, the machine learning experiments will need to focus on not letting the dominant class (survival) overpower the decision-making process so that the system will predict the two rare classes: mortality and re-admission.

The distributions of the remaining demographics from the dataset are not shown in this study due to the large number of NULL values. For the rest of the study, to handle the NULL values, they are filled with the averages of the columns as was done in the previous literature that used the MIMIC-III dataset [27,28].

2.2. Feature Extraction

The MIMIC-III dataset has 26 tables full of different data that can be used as features for training machine learning models. However, selecting the correct set of features is key to achieving good performance from machine learning [29,30]. The biomedical collaborator brought knowledge and hard work to accomplish this good performance to the feature engineering task. Influenced by the aforementioned scoring method actively used in the ICUs in the United States, which requires healthcare professionals to manually score various conditions of the patients and develop a final mortality risk score, known as APACHE scoring, is selected as a valid basis. Leveraging this manual scoring approach, the following measurements are selected as the initial set of features:

i. Age (years) computed from the dataset after subtracting the date of birth from the date of admission, both de-identified.

ii. Temperature (C) of the patient.

iii. Mean arterial pressure (mmHg) of the patient.

iv. pH measurement.

v. Heart rate (beats per min).

vi. Respiratory rate (breaths per minute).

vii. Serum sodium (mEq/L), potassium (mEq/L), creatinine (mg/dL) measurements.

viii. Hematocrit.

ix. WBC (cells/ul).

x. Glasgow-coma-scale points.

xi. A - a gradient (if $FiO_2 \geq 0.5$) (mmHg)

xii. PaO_2 (if $FiO_2 < 0.5$) (mmHg)

xiii. History of organ insufficiency.

xiv. History of immunocompromise.

In addition to the above APACHE features, available demographics are included in the features such as gender, marital status, height, weight, etc. It is necessary to note that except for the constant features such as age or gender, the other measurement values are prone to change over the patient's progress during the ICU stay. Thus, for each patient, there are many measurements for most of the above features. To best express the range of measurements per patient, this study uses the range information, including minimum, maximum, and mean. Hence, in total, 165 features per patient are obtained.

2.3. Machine Learning Methods

Reducing the numerical difference between real-life patient feature values is essential before applying machine learning methods.

To overcome this problem, scaling the feature values in the data collection has become a standard approach. Scaling is achieved through the following:

$$x_i = \frac{x_i - \nu}{\rho}$$

where x_i represents the feature's value i, ν is the mean value of the feature column in the training set, and ρ shows the standard deviation of the same column. Through this operation, feature values present in the dataset are scaled to a smaller version of itself. Within the same features, their mathematical relations are preserved, meanwhile between the features, high differences are scaled down.

Machine learning methods have been popular in the biomedical research domain for many years [31]. To achieve comparability with the ICU mortality and re-admission prediction studies summarized in the introduction section, the same machine learning methods commonly used in interdisciplinary ICU-related research are selected: LR, XGB, and LGBM. In addition, RF is also selected.

LR is considered the baseline approach since it has the simplest methodology within its algorithmic structure. LR returns the correct classification result by combining selected attributes with a linear mathematical formula [20]. This method focuses on separating data from one class at a time in the most successful way. L2 was used as the loss function. This simple approach becomes the baseline.

The other approach, RF, is an ensemble method that constructs multiple decision trees and averages their decisions into one decision. As a more complex variant of an RF method, another approach used in the experiments is XGB. It is perhaps the most popular method among the ensemble methods in recent years. It is an optimized distributed gradient boosting method designed to be highly efficient, flexible, and portable. It applies machine learning algorithms under the Gradient Boosting framework. XGB provides a parallel tree-boosting solution that solves many data science problems quickly and accurately [32]. Unfortunately, its disadvantage is that it is highly parametric, meaning that it becomes difficult to achieve a good performance if the correct parameters are not selected.

Next, the LGBM method is included in the experiments. It combines multiple decision trees, each focusing on improving the predictions of the previous ones within a gradient-boosting framework, just like the XGB method. However, unlike XGB, this method creates histograms for each feature and uses them to approximate the best-split point [33]. LGBM has other algorithmic differences, such as allowing leaf-wise growth rather than depth-wise growth in the trees.

2.4. Optimization

Despite their success in performing accurate predictions, one disadvantage shared by most machine learning methods is the number of parameters they need to be tuned. It is unrealistic to expect to find the perfect parameters for each model. However, with the help of methods such as a grid or a random search, it is possible to search over a set of possible parameters and find an optimal combination. In the present study, a library named Optuna is used to find optimal parameters for each method [34].

For each method, ten trials are conducted within every fold of the 5-fold cross-validation framework, where each trial tried a different parameter subset. The Optuna library moves towards the parameters that returned high performance in the previous trials, thus ensuring an optimal parameter set. The highest average macro F1 scores are selected and used in the tests as the performance criteria. Also, since all classification tasks have high-class imbalance problems, the class weights are provided inside all machine learning methods, considering the class imbalance.

3. Results and Discussion

As stated earlier, studies in the literature mainly focus on the binary classification task of detecting mortality risk among ICU patients. A few studies follow the same binary classification task for detecting the re-admission risk. However, no studies have conducted a multi-class classification task and provided a deep comparison with the binary tasks. To complete this lack of comparative information, this study performs three distinct experiments:

i. Conducting a binary classification task for mortality risk detection.

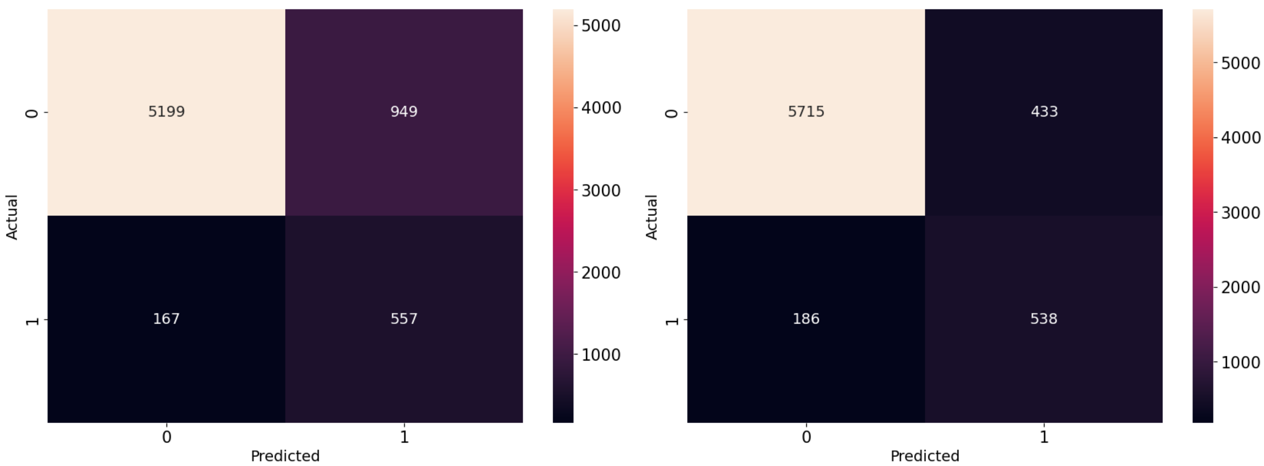
ii. Conducting a binary classification task for re-admission detection.

iii. Conducting a multi-class classification task to identify the likelihood of an ICU patient dying, recovering during the current ICU stay, or getting re-admitted after discharge.

For comparability, the same four machine learning methods and the same set of 165 features (APACHE and demographics) are used in the above experiments. Considering the data imbalance in the test set, among the possible performance scoring methods, Area under the curve (AUC) is selected for its ability to provide fair scoring [35], and its popularity as a scoring method in the mortality and re-admission prediction tasks as seen in Table 3. Meanwhile, macro F1 and recall (sensitivity) scores are also utilized to overcome possible overestimations AUC may return. Thanks to the macro setting of the F1 and recall metrics, class imbalance in the test set is handled fairly.

Table 2. AUC, macro F1, and macro recall scores of the proposed novel feature space used with four machine learning methods.

Goal	Metric	LR	RF	LGBM	XGB
Mortality	AUC	89.40	91.51	**92.43**	91.51
	F1	70.13	80.39	79.17	**82.65**
	Recall	80.75	78.78	**83.63**	79.36
Re-admission (30 days)	AUC	**65.03**	64.67	61.77	58.31
	F1	48.81	53.12	**53.15**	49.11
	Recall	**60.52**	52.75	54.71	50.58
Multiclass	AUC	75.54	**77.94**	76.98	76.50
	F1	48.92	53.49	**54.19**	54.03
	Recall	57.33	52.68	**58.26**	51.77



(a) LR (b) LGBM

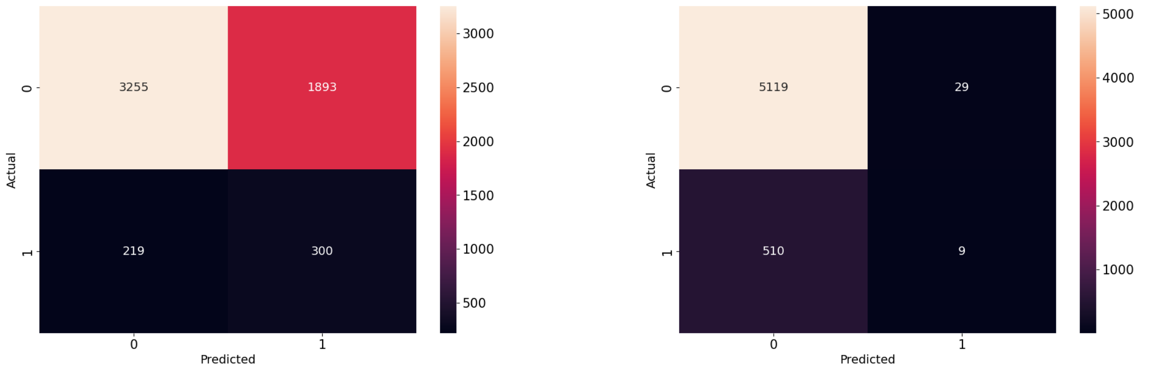
Figure 4. Confusion matrices of the mortality prediction task obtained from the baseline and the best-performing methods.

According to the results presented in Table 2, the mortality prediction task returned AUC scores around 90% with each method. Meanwhile, the highest F1 score is from the XGB method, and the highest sensitivity belongs to the LGBM. Compared to the mortality prediction results in the literature in Table 3, mortality prediction performance in the current study outperformed the past. Because the studies in the literature focused on specific disease mortality, which is more straightforward than a general mortality prediction, the current results are significant. This finding shows that including all the APACHE metrics as features, computing their statistical changes over the ICU stay of the patients through min, max, mean, and median operations, among others, provide more robust outcomes for mortality prediction. Also, by including the demographics in the feature set, mortality prediction can be achieved with better performance. This is particularly interesting as it points to the genetic variability and risk factors of the patients and shows the importance of considering demographics for clinical decision-making. Another finding is regarding the choice of methods. Figure 4 highlights that the choice of machine learning method does not affect the mortality prediction task thanks to the novel feature space introduced in the current study.

Table 3. Some prediction objectives and scores from the literature that used the MIMIC-III dataset, excluding the clinical notes. Each re-admission study performs a 30-day prediction.

Objective	Features	Method	Score	Study
Mortality (heart failure)	Patient measures (overlap with APACHE), demographics	LR, XGB	84.16 AUC	[28]
Mortality (ventilated)	Patient measures (overlap with APACHE), demographics, history	KNN, LR, DT, RF, XGB, ANN	82.1 AUC	[27]
Mortality (pancreatitis)	Patient measures (overlap with APACHE), demographics	LR, Ranson, ANN	76.9 AUC	[36]
Re-admission	17 APACHE measures, demographics, ICD-9 embeddings	LSTM	74.2 recall, 79.1 AUC	[17]
Re-admission	Patient measures (overlap with APACHE), demographics, ICD-9 embeddings	LR, RF, SVM, ANN	65 accuracy, 60 AUC	[19]
Re-admission	Patient measures (overlap with APACHE), demographics, ICD-9 code	LR, RF, XGB	37 F1, 75 AUC	[37]

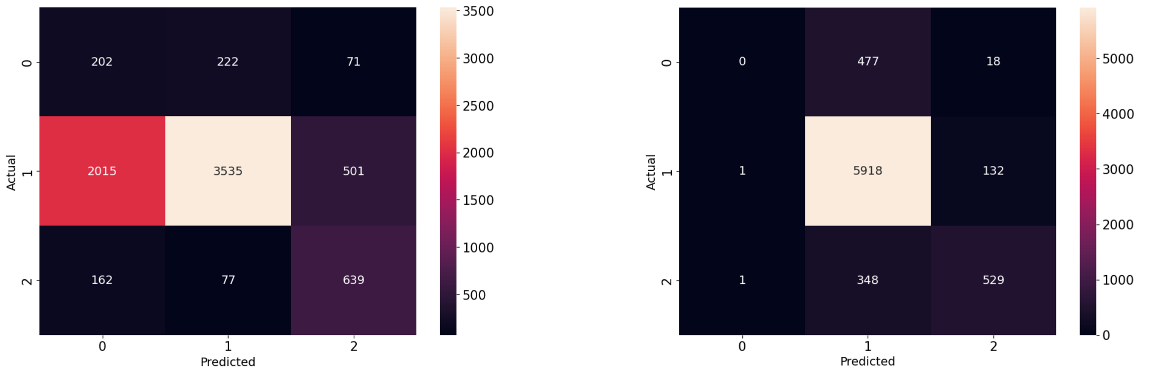
Re-admission prediction within the 30 days after discharge proves to be more complex than the mortality prediction according to the results in Table 2, which is validated by the scores in the literature Table 3. F1 score of 37% obtained in the literature [37] proves the difficulty of re-admission prediction even in a binary classification setting. In the current study, LR returned the highest AUC score of 65% for re-admission. Compared to the readmission AUC scores in Table 3, this score falls in the middle of the range in the literature. The F1 rate of 53% reported in the current study is higher than that of the literature. In Figure 5, two confusion matrices show that the LR method misclassified the survived patients as re-admission, and the XGB method did the opposite and misclassified the re-admissions as survived. One must decide based on the trade-off between high false positives and false negatives. Creating a model that returns high false positives would cause too much money and attention to be given to patients who are well, and high false negatives would cause patients who will be re-admitted within 30 days to be released early, which also causes a waste of money and resources. One conclusion from these findings is that, although APACHE features and demographics are successful at identifying mortality, they are not enough to predict ICU re-admission. Therefore, the current feature space needs to be enriched further for a higher re-admission prediction performance.



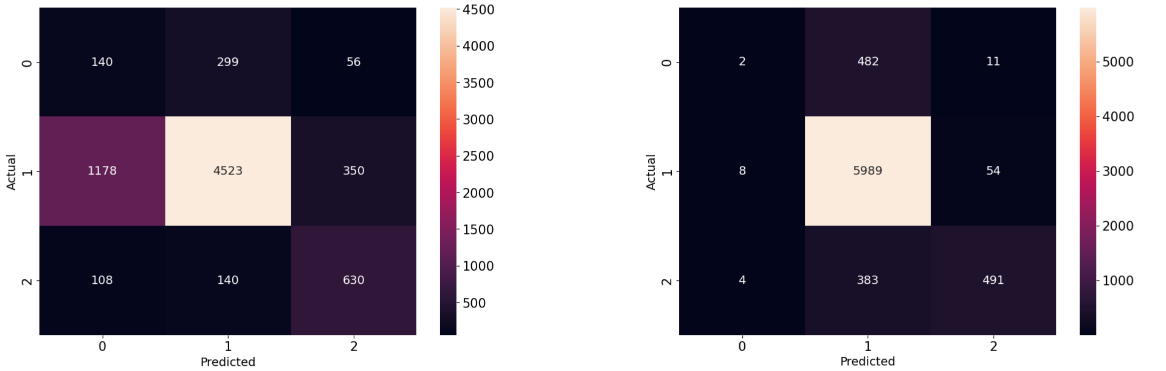
(a) LR **(b)** XGB

Figure 5. Confusion matrices of the 30 days re-admission prediction task obtained from the baseline and the worst performing methods

The last experiment tackles the challenge of performing multi-class classification to determine if a patient in the ICU will die, recover, or get re-admitted 30 days after getting discharged from the ICU. Table 2 shows that each method returned around 77% AUC for multi-class classification. Meanwhile, the F1 and recall scores show a distinction between the results. The confusion matrices in Figure 6 display the similarities and differences between the selected methods. The matrices of the RF and XGB methods appear too similar. For example, both methods failed to capture re-admissions, which is explainable considering the similarities between their algorithms - using decision trees within. The remaining two methods, LR and LGBM, returned more acceptable confusion matrices by successfully capturing re-admission cases. Among the two, LGBM has the best performance. For a 3-class classification problem, an F1 score of 54% is acceptably good compared to the random chance of 33%. LGBM's capability to capture complex non-linear relationships between a mix of numerical and categorical variables proves it to be better than the remaining models in the current context.



(a) LR **(b)** RF



(c) LGBM **(d)** XGB

Figure 6. Confusion matrices of the three-class classification framework obtained from the baseline and the best-performing methods (Re-admission=0, survival=1, mortality=2)

4. Conclusion

The study carried out in this paper showed the simplicity of mortality prediction tasks in the ICU setting, using the APACHE measures together with the demographics, which explains the reasoning behind the vast number of available studies that focus only on mortality prediction. In parallel, the study showed the difficulty of the 30-day re-admission prediction task. The state-of-the-art machine learning methods that win various imbalanced data classification competitions failed the re-admission prediction task, while the simple baseline method outperformed the rest. This finding shows the difficulty of the re-admission prediction task and the failure of the APACHE and demographics data in providing predictive power. Because real-life expert healthcare providers signed off on the discharge forms of these patients during the data collection process, it is possible to conclude that machine learning methods still performed better than the actual human experts, with a 65% success rate. Yet, there is still room for improvement. Furthermore, the multi-class classification performance showed 77% success at differentiating recovery, mortality, and returning to the intensive care unit in 30 days, which are highly acceptable results to be used in real-life automated clinical decision-making systems. The proposed system with a simple baseline approach can help healthcare professionals save more lives and reduce the risk of early discharges without adding features - since APACHE features and demographics are mandatory to collect already. Finally, our dataset was limited to MIMIC III, which contains the medical history and data of all patients admitted to Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2001 and 2012. Thus, while predicting re-admission, our study could not consider the possibility of the patients being re-admitted to another hospital.

The overloaded ICUs during the COVID-19 pandemic before vaccines were available showed the unmet need for automated clinical decision-making systems. While the proposed system has demonstrated acceptable performance, in the rework, more features that are also mandatory to collect in an ICU setting will be included in the experiments to increase the re-admission prediction performance. Future work will also consist of evaluating different real-life ICU data collections.

Author Contributions

The first author performed coding, running the experiments, and wrote the manuscript. The second author analyzed the raw data, guided the data science, and wrote the manuscript. All authors read and approved the final version of the paper.

Conflicts of Interest

All the authors declare no conflict of interest.

Ethical Review and Approval

No approval from the Board of Ethics is required.

Acknowledgment

This work was supported by the Office of Scientific Research Projects Coordination at Çanakkale Onsekiz Mart University, Grant number: FHD-2021-3737. Neural Stem Cell Institute is not responsible for the statements and results of the paper.

References

[1] K. Cooley-Rieders, K. Zheng, *Physician documentation matters. Using natural language processing to predict mortality in sepsis*, Intelligence-Based Medicine 5 (2021) 100028 7 pages.

[2] Y.-W. Lin, Y. Zhou, F. Faghri, M. J. Shaw, R. H. Campbell, *Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory*, Plos One 14 (7) (2019) e0218942 22 pages.

[3] T.C. Sağlık Bakanlığı, Yoğun Bakım Skorlama, `https://shgmkalitedb.saglik.gov.tr/TR,9024/yogun-bakim-skorlama.html`, Accessed: 2024-07-28.

[4] H. Mumtaz, M. K. Ejaz, M. Tayyab, L. I. Vohra, S. Sapkota, M. Hasan, M. Saqib, *APACHE scoring as an indicator of mortality rate in ICU patients: A cohort study*, Annals of Medicine and Surgery 85 (3) (2023) 416–421.

[5] R. Sarkar, C. Martin, H. Mattie, J. W. Gichoya, D. J. Stone, L. A. Celi, *Performance of intensive care unit severity scoring systems across different ethnicities in the USA: A retrospective observational study*, The Lancet Digital Health 3 (4) (2021) e241–e249.

[6] H. Yang, L. Kuang, F. Xia, *Multimodal temporal-clinical note network for mortality prediction*, Journal of Biomedical Semantics 12 (2021) 1–14.

[7] Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, Provisional Mortality on CDC WONDER Online Database, `http://wonder.cdc.gov/mcd-icd10-provisional.html`, Accessed: 2024-07-24.

[8] B. G. Pijls, S. Jolani, A. Atherley, R. T. Derckx, J. I. Dijkstra, G. H. Franssen, S. Hendriks, A. Richters, A. Venemans-Jellema, S. Zalpuri, M. P. Zeegers, *Demographic risk factors for COVID-19 infection, severity, ICU admission and death: A meta-analysis of 59 studies*, BMJ Open 11 (1) (2021) e044640 10 pages.

[9] S. L. Hyland, M. Faltys, M. Hüser, X. Lyu, T. Gumbsch, C. Esteban, C. Bock, M. Horn, M. Moor, B. Rieck, M. Zimmerman, D. Bodenham, K. Borgwardt, G. Rätsch, T. M. Merz, *Early prediction of circulatory failure in the intensive care unit using machine learning*, Nature Medicine 26 (3) (2020) 364–373.

[10] A. Mohammed, F. Van Wyk, L. K. Chinthala, A. Khojandi, R. L. Davis, C. M. Coopersmith, R. Kamaleswaran, *Temporal differential expression of physiomarkers predicts sepsis in critically ill adults*, Shock 56 (1) (2021) 58–64.

[11] J. Liu, J. Wu, S. Liu, M. Li, K. Hu, K. Li, *Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model*, Plos One 16 (2) (2021) e0246306 11 pages.

[12] J. Wang, L. Zhou, Y. Zhang, H. Zhang, Y. Xie, Z. Chen, B. Huang, K. Zeng, J. Lei, J. Mai, J. Pan, Y. Chen, J. Wang, Q. Guo, *Minimum heart rate and mortality in critically ill myocardial infarction patients: An analysis of the MIMIC-III database*, Annals of Translational Medicine 9 (6) (2021) 496 9 pages.

[13] Y. Chen, Y. Du, C. Sun, W. Tan, *Lactate is associated with increased 30-day mortality in critically ill patients with alcohol use disorder*, International Journal of General Medicine 14 (2021) 2741–2749.

[14] Y. Ji, L. Li, *Lower serum chloride concentrations are associated with increased risk of mortality in critically ill cirrhotic patients: An analysis of the MIMIC-III database*, BMC Gastroenterology 21 (1) (2021) 200–208.

[15] H. Yang, L. Kuang, F. Xia, *Multimodal temporal-clinical note network for mortality prediction*, Journal of Biomedical Semantics 12 (2021) 1–14.

[16] S. Upadhyay, A. L. Stephenson, D. G. Smith, *Readmission rates and their impact on hospital financial performance: A study of Washington hospitals*, INQUIRY: The Journal of Health Care Organization, Provision, and Financing 56 (2019) 1–10.

[17] Y.-W. Lin, Y. Zhou, F. Faghri, M. J. Shaw, R. H. Campbell, *Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory*, Plos One 14 (7) (2019) e0218942 22 pages.

[18] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, E. Liu, *Predicting all-cause risk of 30-day hospital readmission using artificial neural networks*, Plos One 12 (7) (2017) e0181173 14 pages.

[19] R. Assaf, R. Jayousi, *30-day hospital readmission prediction using MIMIC data*, in: 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT), IEEE, 2020, pp. 1–6.

[20] U. Bayram, L. Benhiba, *Determining a person's suicide risk by voting on the short-term history of tweets for the clpsych 2021 shared task*, in: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access, 2021, pp. 81–86.

[21] U. Bayram, J. Pestian, D. Santel, A. A. Minai, *What's in a word? Detecting partisan affiliation from word use in congressional speeches*, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.

[22] 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, H. E. Stanley, *PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals*, Circulation 101 (23) (2000) e215–e220.

[23] A. Johnson, T. Pollard, R. Mark, *MIMIC-III clinical database (version 1.4)*, PhysioNet 10 (2016) 1–2.

[24] A. E. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R. G. Mark, *MIMIC-III, a freely accessible critical care database*, Scientific Data 3 (1) (2016) 1–9.

[25] S. Lesser, S. Zakharkin, C. Louie, M. R. Escobedo, J. Whyte, T. Fulmer, *Clinician knowledge and behaviors related to the 4Ms framework of age-friendly health systems*, Journal of the American Geriatrics Society 70 (3) (2022) 789–800.

[26] F. Zakirov, A. Krasilnikov, *Age-related differences in decision-making process in the context of healthy aging*, in: BIO Web of Conferences, Vol. 22, EDP Sciences, 2020, pp. 1–6.

[27] Y. Zhu, J. Zhang, G. Wang, R. Yao, C. Ren, G. Chen, X. Jin, J. Guo, S. Liu, H. Zheng, Y. Chen, Q. Guo, L. Li, B. Du, X. Xi, W. Li, H. Huang, Y. Li, Q. Yu, *Machine learning prediction models for mechanically ventilated patients: Analyses of the MIMIC-III database*, Frontiers in medicine 8 (2021) 662340 9 pages.

[28] F. Li, H. Xin, J. Zhang, M. Fu, J. Zhou, Z. Lian, *Prediction model of in-hospital mortality in intensive care unit patients with heart failure: Machine learning-based, retrospective analysis of the MIMIC-III database*, BMJ Open 11 (7) (2021) e044779 17 pages.

[29] P. Domingos, *A few useful things to know about machine learning*, Communications of the ACM 55 (10) (2012) 78–87.

[30] U. Bayram, *Applying machine learning to online data?: Beware! computational social science requires care*, in: Opportunities and Challenges for Computational Social Science Methods, IGI Global, 2022, pp. 100–125.

[31] J. Pestian, D. Santel, M. Sorter, U. Bayram, B. Connolly, T. Glauser, M. DelBello, S. Tamang, K. Cohen, *A machine learning approach to identifying changes in suicidal language*, Suicide and Life-Threatening Behavior 50 (5) (2020) 939–947.

[32] T. Chen, C. Guestrin, *Xgboost: A scalable tree boosting system*, in: Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[33] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, *Lightgbm: A highly efficient gradient boosting decision tree*, Advances in Neural Information Processing Systems 30 (2017) 3149–3157.

[34] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, *Optuna: A next-generation hyperparameter optimization framework*, in: Proceedings of the 25th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, 2019, pp. 2623–2631.

[35] T. Yang, Y. Ying, *AUC maximization in the era of big data and AI: A survey*, ACM Computing Surveys 55 (8) (2022) 1–37.

[36] N. Ding, C. Guo, C. Li, Y. Zhou, X. Chai, *An artificial neural networks model for early predicting in-hospital mortality in acute pancreatitis in MIMIC-III*, BioMed Research International 2021 (1) (2021) Article ID 6638919 8 pages.

[37] A. Pakbin, P. Rafi, N. Hurley, W. Schulz, M. H. Krumholz, J. B. Mortazavi, *Prediction of ICU readmissions using data at patient discharge*, in: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2018, pp. 4932–4935.