



Article

Outcome Prediction in Critically-Ill Patients with Venous Thromboembolism and/or Cancer Using Machine Learning Algorithms: External Validation and Comparison with Scoring Systems

Vasiliki Danilidou ^{1,2,*} , Stylianos Nikolakakis ³, Despoina Antonakaki ⁴ , Christos Tzagkarakis ⁴,
Dimitrios Mavroidis ⁴, Theodoros Kostoulas ⁵ and Sotirios Ioannidis ^{3,4}

¹ Sphynx Technology Solutions, 6300 Zug, Switzerland

² School of Medicine, European University of Cyprus, 2404 Nicosia, Cyprus

³ School of Electrical and Computer Engineering, Technical University of Crete, 73100 Chania, Greece; snikolakakis@tuc.gr (S.N.); sotiris@ece.tuc.gr (S.I.)

⁴ Institute of Computer Science (ICS)-Foundation for Research and Technology-Hellas (FORTH), 70013 Heraklion, Greece; despoina@ics.forth.gr (D.A.); ctzagkarakis@gmail.com (C.T.); dmavroid@gmail.com (D.M.)

⁵ Department of Information and Communication Systems Engineering, School of Engineering, University of the Aegean, 83200 Samos, Greece; theodoros.kostoulas@aegean.gr

* Correspondence: v.danilidou@euc.ac.cy or v.danilidou@sphynx.ch



Citation: Danilidou, V.; Nikolakakis, S.; Antonakaki, D.; Tzagkarakis, C.; Mavroidis, D.; Kostoulas, T.; Ioannidis, S. Outcome Prediction in Critically-Ill Patients with Venous Thromboembolism and/or Cancer Using Machine Learning Algorithms: External Validation and Comparison with Scoring Systems. *Int. J. Mol. Sci.* **2022**, *23*, 7132. <https://doi.org/10.3390/ijms23137132>

Academic Editor: Ekaterini Chatzaki

Received: 27 May 2022

Accepted: 19 June 2022

Published: 27 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Intensive care unit (ICU) patients with venous thromboembolism (VTE) and/or cancer suffer from high mortality rates. Mortality prediction in the ICU has been a major medical challenge for which several scoring systems exist but lack in specificity. This study focuses on two target groups, namely patients with thrombosis or cancer. The main goal is to develop and validate interpretable machine learning (ML) models to predict early and late mortality, while exploiting all available data stored in the medical record. To this end, retrospective data from two freely accessible databases, MIMIC-III and eICU, were used. Well-established ML algorithms were implemented utilizing automated and purposely built ML frameworks for addressing class imbalance. Prediction of early mortality showed excellent performance in both disease categories, in terms of the area under the receiver operating characteristic curve (*AUC-ROC*): VTE-MIMIC-III 0.93, eICU 0.87, cancer-MIMIC-III 0.94. On the other hand, late mortality prediction showed lower performance, i.e., *AUC-ROC*: VTE 0.82, cancer 0.74–0.88. The predictive model of early mortality developed from 1651 VTE patients (MIMIC-III) ended up with a signature of 35 features and was externally validated in 2659 patients from the eICU dataset. Our model outperformed traditional scoring systems in predicting early as well as late mortality. Novel biomarkers, such as red cell distribution width, were identified.

Keywords: venous thromboembolism; cancer; mortality; ICU; interpretable machine learning

1. Introduction

Venous thromboembolism (VTE) and cancer are major causes of death worldwide [1] and their prevalence is continuously rising due to the doubling of life expectancy [2], the tripling of the world population over the last 70 years, changes in lifestyle, the increased prevalence of chronic diseases [3], and the COVID-19 pandemic [4]. VTE can present with clots in the veins, most frequently as deep vein thrombosis (DVT) and pulmonary embolism (PE). Despite the recent advances in treatment, fatality rates have increased in the last decade, with notable racial and geographical disparities [5]. Patients with VTE or cancer occasionally need advanced support and suffer from significant morbidity,

prolonged intensive care unit (ICU) stay, and high mortality rates, not only early during hospitalization but even after several months.

Mortality prediction in ICU patients has been a major challenge in the area of medical informatics [6] and has long been used as a quality indicator. Mortality is a major end point in epidemiological and interventional studies in the ICUs [7]. Published studies so far mainly focus on the prediction of in-hospital or early mortality, irrespective of the primary diagnosis and related comorbidities [8–12], whereas studies focusing on late mortality prediction are quite rare [13,14]. It is generally accepted that the initial diagnosis of a patient and the evaluation of the reasons for ICU admission could significantly affect the overall survival; thus, it would be interesting to study individually the different disease outcomes and identify the specific disease-related clinical features with prognostic significance. Even more importantly, it is necessary to predict post-discharge mortality, which is a very challenging task, since patients admitted to the ICU usually suffer from a high comorbidity burden [15]. Prompt identification of factors associated with late mortality could help physicians to re-orientate medical clinical practices (e.g., extended anticoagulation), identify modifiable factors, and reasonably allocate health resources in the ICUs, which are extremely restricted, especially during the COVID-19 pandemic period.

On the other hand, there is a growing availability of automated ICU patient surveillance systems and healthcare big data that remains unexploited, whereas they could provide the opportunity to examine data-driven research solutions using modern machine learning (ML) methods. The wealth of available ICU data combined with the use of ML-based approaches have impacted medical predictive analytics and clinical decision support systems, since ML algorithms can learn from complex data patterns and identify associations that could help in the improvement of patient care and survival, as well as lowering hospitalization costs [16].

This article is organized as follows. Section 2 reviews previous work on ICU mortality prediction. Specifically, traditional clinical scores considering the standard clinical practice based on a limited number of features, collected during admission to the ICU, are compared with recent approaches using state-of-the-art ML algorithms and routinely collected data from medical records. Section 3 presents the motivation and the aim of our work, which is early and late mortality ML-based prediction focusing on patients with thrombosis or cancer. Section 4 describes the datasets used in our study, the selection of cohorts, the time-dependent features used to build the prognostic model, the data manipulation, and the two adopted ML frameworks. Sections 5 and 6 provide an extensive discussion of the results on early and late prediction in ICU patients with thrombosis or cancer, the main predictive features, along with their interpretation and the external validation of the model. Finally, Section 7 summarizes the main outcomes and highlights the fundamental contributions of the current study.

2. Related Work

Since the early 1980s, several clinical scores have been introduced into clinical practice to predict in-hospital ICU mortality. The Simplified Acute Physiology Score (SAPS) [17], Acute Physiology and Chronic Health Evaluation (APACHE) [18], and Sequential Organ Failure Assessment (SOFA) [19] score are considered as validated tools in predicting ICU mortality [20]. Nevertheless, these scores have several limitations. They have been developed based on different target populations with heterogeneous inclusion and exclusion criteria; thus, during the validation process, they provide modest performance. They are based on multivariate statistical methods, such as logistic regression models, disregarding the non-linear relationships that exist between variables in real-life medical data. Since the scores are computed on health data collected during the first 24 h of ICU admission or instant-based measurements (e.g., the worst or average value), they do not consider time-based measurements, which could contain important information about clinical deterioration [21]. It is worth noting that inter-rater agreement is low between the various clinicians, mostly affected by personal experience, and thus a potential bias in the scores'

interpretation exists [22]. Finally, these clinical prediction scores have poor generalization and inadequate model calibration, especially in the high-risk patients, which is the type of patients that we study [23].

Other thrombosis-specific clinical scores, such as the Pulmonary Embolism Severity Index (PESI) [24], have been mostly used to identify low-risk patients that could benefit from outpatient treatment or early discharge, although they have also been validated to assess the probability of 30- and 90-day mortality post-PE [25–28]. Overall, it is unclear if clinicians routinely use these risk stratification tools, since only a small proportion of low-risk patients based on the PESI score are discharged earlier [24]. Thus, the clinical usefulness of these risk stratification scores needs to be proven. A recent meta-analysis combining all available risk stratification tools for PE showed that although most of them have high sensitivity, the low specificity in the range of 50% discourages clinicians from universal acceptance and employment in everyday clinical practice [29].

There is increasing interest in the literature in using modern ML algorithms, such as random forests (RF) and gradient boosting machines, since it has been shown that they can predict more accurately clinical outcomes (e.g., mortality) in comparison with classical logistic regression [30]. Some of the recent works on mortality prediction in ICU patients show the superiority of the ML-based predictive models against the traditional clinical scores [10,31,32]. These studies are not focused on a specific diagnostic group of patients and only use a limited feature set [8,11,12,33].

Prediction of mortality in critically ill patients with thrombosis or cancer using ML algorithms has not been extensively studied so far. To the best of our knowledge, the only study in the literature that investigates ICU patients with thrombosis is [32]. The study is based on a relatively small number of patients, and the authors used only admission data, developing two models for 30-day mortality prediction based on logistic regression, and least absolute shrinkage and selection operator (Lasso) for feature selection. Regarding patients with cancer, there is a paucity of studies predicting mortality in the ICU. In [34], the authors provide a systematic review on ML-based early mortality prediction in cancer patients (but not critically ill patients). The average area under the receiver operating characteristic curve (*AUC-ROC*) is 0.72–0.92 and differs between the various cancer types. The absence of studies in this field can be partly explained by the fact that the admission of cancer patients to the ICU is considered frequently unavailing, and physicians sometimes discourage their patients and families from proceeding with aggressive treatment. Nevertheless, it is highly important to recognize the patients who would benefit from intensive care treatment. Traditional clinical scores, such as APACHE, fail to predict accurately individual outcomes in these patients [35,36], and thus there is a need for more sophisticated models.

3. Motivation of the Current Work

Motivated by the lack of studies in the two particular diagnostic groups of patients mentioned above, as well as by the limitations of the traditional clinical scores, we aim to develop and validate prognostic models for early and late mortality using state-of-the-art ML algorithms, while exploiting almost all available data in the electronic health record. Since healthcare datasets are unstructured and heterogeneous, manual features' engineering and extraction is time-consuming; thus, combined bioinformatician and clinician expertise is needed. This can be attenuated by adopting ML models trained on multidimensional data [33]. Our initial hypothesis was that the utilization of all the information stored in the electronic health record, i.e., demographics, laboratory tests, medications, diagnosis, and relevant procedures, along with their detailed timeline information, could help a clinician to dramatically improve the decision-making process. We were also interested in identifying the most informative features in a medical record, and thus we performed feature discriminative analysis during the analysis of the various groups of features.

Towards this direction, two different open-access high-dimensional multicenter retrospective datasets were used (MIMIC-III and eICU databases). The derived models were

compared with traditional clinical scores, externally validated to prove generalizability, and finally interpreted by identifying clinically meaningful predictive features. Additionally, we introduced a custom-based ML approach combined with an oversampling method to address the dataset imbalance. We recognized some interesting biomarkers, which, although readily available, are currently disregarded during clinical practice. We strongly believe that our work will be of added value in refining conventional clinical scores and rediscovering easily measurable and low-cost biomarkers.

4. Materials and Methods

4.1. Ethics Statement

This is a retrospective study based on two freely accessible databases, MIMIC-III [37] and eICU [38], both created in accordance with the Health Insurance Portability and Accountability Act (HIPAA) standards, where all investigators with data access were approved by PhysioNet [39]. Patient data were de-identified. MIMIC-III patient data are date-shifted and age over 89 years old was set to 300 years old, whereas in eICU, older patients are referred to as >89 years old. All pre-processing and data analysis were performed under Physionet regulations.

4.2. Dataset Description and Cohort Selection

4.2.1. MIMIC-III Database

The Medical Information Mart for Intensive Care Database (MIMIC-III, version 1.4) [37] is a single institutional ICU database from Beth Israel Deaconess Medical Center, comprising health-related data from 38,597 adult patients and 49,785 admissions in the ICU between 2001 and 2012. Two groups of patients were studied, namely patients with (i) thrombosis and (ii) cancer. Initially, we selected all patients aged >15 years old hospitalized in the ICU with a primary diagnosis of thrombosis based on 35 distinct ICD-9 codes. We excluded age <15 years ($n = 3$), pregnancy and puerperium complications ($n = 40$), and patients with a “do not resuscitate” (DNR) code ($n = 169$). Overall, 2468 patients (6.4% of total patients in MIMIC-III) were selected. As a second group, we utilized all patients aged >15 years old hospitalized in the ICU with a primary diagnosis of solid cancer or hematological malignancy based on 101 different ICD-9 and ICD-10 codes. Overall, 5691 patients (14.74% of total patients in MIMIC-III) were selected. Patients with age <15 years ($n = 0$), pregnancy and puerperium complications ($n = 15$), and patients with DNR code ($n = 358$) were excluded. After the exclusion, a total number of 5318 patients remained in the study.

4.2.2. eICU Database

The eICU Collaborative Research Database (eICU, v2.0) [38] is a multi-center ICU database with over 200,000 admissions for almost 140,000 patients, admitted to more than 200 hospitals between 2014 and 2015 across the USA. We selected patients aged >15 years old with a primary diagnosis of VTE based on 18 different ICD-9 and ICD-10 codes. The same inclusion and exclusion criteria were used, as described above. Overall, 4385 VTE patients (3.15% of total patients in the eICU database) were identified. Detailed demographic and clinical characteristics of patients from both databases are shown in Table 1.

Table 1. Demographic and clinical characteristics of patients with VTE or cancer from eICU and MIMIC-III databases. SD denotes the standard deviation, and LOS length of stay. *p*-Values between “surviving” and “non-surviving” patients are reported.

Characteristic	VTE		Cancer			
	eICU	<i>p</i> -Value	MIMIC-III	<i>p</i> -Value	MIMIC-III	<i>p</i> -Value
Thrombosis	3724		2468		5128	
- PE	2739 (62.4%)	0.71	960 (38.9%)	0.92	-	
- DVT	2220 (50.6%)	0.95	1543 (62.5%)	-	-	
Sex						
- Female	1788 (48 %)	0.24	1024 (41.5%)	0.53	2192 (41.2%)	0.07
- Male	1934 (52 %)	0.24	1444 (58.5%)	0.53	3126 (58.8%)	0.07
- Unknown	1		-		-	
Ethnicity						
- White	2845 (76.4%)	0.36	1801 (73%)	0.27	4049 (76%)	0.57
- Black	522 (14%)	0.16	246 (10%)	0.47	378 (7.1 %)	0.83
- Other	357 (9.6%)		421 (17%)		891 (16.8%)	
Age, years						
Average (SD)	60.28 (16.26)	<0.001	62.64 (16.7)	<0.001	66.2 (14.2)	<0.001
Min	15		17.4		18.9	
Max	90		98.7		98.9	
Cancer diagnosis	409 (9.3 %)	<0.001	605 (24.5%)	<0.001		
LOS, days						
Mean (SD)	11.22 (12.24)	0.002	7.06 (10.06),	<0.001	4.4 (6.5)	<0.001
Median:	7.01		153.9 days		2.2	
Mortality (%):						
Early:	267 (7.16 %)		348 (14.1%)		902 (17%)	
Late:	N/A		817 (33.1%)		2659 (49.9%)	
“Survivor”:	3457 (92.84%)		1303 (52.8%)		1757 (33.1%)	
Death time(days)						
Average (SD):	N/A		390 (647)		328.5 (536.55)	
Median:			83		365	

4.3. Feature Pre-Analysis Selection

To investigate potential novel discriminatory attributes, features that were extracted from the database and used to build the prognostic model were chosen based on the clinical experts’ opinion and, following a liberal approach, we tried to simulate a real-life scenario where the medical practitioner exploits all relevant clinico-laboratory information available on the electronic health record.

4.3.1. MIMIC-III

A total amount of 1471 features including demographics, clinico-laboratory information, medications, and procedures were extracted. Apart from the features extracted directly from the database, we computed through SQL queries various meta-features (known as concepts), such as clinical severity scores and first-day labs available as scripts on GitHub. Unstructured notes written by clinicians in free text format were extracted as text entities using the Sequence Annotator for Biomedical Entities and Relations (SABER) [40], which is a deep learning tool for information extraction in the medical domain. A detailed description of these features’ and meta-features’ processing, as well as a natural language processing overview, can be found in [41].

4.3.2. eICU

Out of 31 tables, we selected variables from the following tables: *patient* (demographics, admission, and discharge information), *diagnosis*, *admission_dx* (primary and other diagnoses), *physicalExam*, *vitalPeriodic* (vital signs), *apachePredVar*, *apacheApsVar* (clinical scores), *lab* (laboratory measurements), *admissiondrug*, *infusion*, *treatment* (drugs administered prior to and during ICU stay). The description of the selected features for each group is given in Appendix A Table A1.

4.3.3. Traditional Clinical Scores

The eICU database contains information from APACHE version IV and IVa and Acute Physiology Score (APS) version III scores, whereas the MIMIC-III database contains SOFA, SAPS, Outcome and Assessment Information Set (OASIS), Logistic Organ Dysfunction Score (LODS), Multiple Organ Dysfunction (MODS), APS III, as well as comorbidity scores such as Elixhauser. We compared the performance (*AUC-ROC*) of the derived model using the extended feature set with the available medical scores recorded in the databases.

4.4. Data Manipulation and Transformation

Age adjustment has been originally applied in both datasets to comply with privacy regulations. Older patients were all assigned as 90 years old, given that the risk of thrombosis is homogeneously high in those more than 85 years old.

JADBio automatically performs pre-processing of the data—that is, mean and mode imputation of missing data, constant feature removal (features that contain only one value for all the outputs in the dataset and therefore are meaningless are removed), and standardization of the feature range.

A significant challenge that had to be addressed manually during attribute selection was the redundancy of the features in both datasets. For example, several drugs are prescribed either with the trade name or the active compounds (e.g., enoxaparin or lovenox), whereas misspellings of the names are frequent (e.g., agatroban or argatoban instead of argatroban). The most important medication groups extracted were anticoagulants, antiplatelets, cardiovascular, antidiabetic, antilipidemics, thrombolytics, vasopressors, antibiotics, chemotherapeutics, and corticosteroids. Dosage and duration of treatment was not taken into account in the current experiments. Laboratory tests, such as complete blood count, kidney and liver function tests, acid base balance, clotting, and biochemical and enzyme tests, were extracted. For each of these features, the first and the average value from the whole ICU stay were selected. Time stamps were defined as the average values per 6 h, during the first 48 h of the ICU stay. Finally, vital signs were recorded as average values per 1 h, during the first 48 h in the ICU, first, and average values during the ICU stay.

4.5. Automated Machine Learning Framework: JADBio

The automated ML (AutoML) platform JADBio automatically tries and evaluates numerous ML pipelines, optimizing the pre-processing, feature selection, and modeling the steps and their hyper-parameters [42]. It employs the Bootstrap Bias Corrected Cross-Validation (BBC-CV) to provide an unbiased estimation of performance that adjusts (controls) for trying multiple ML pipelines [43,44]. Pre-processing, normalization, mode and mean imputation, constant feature removal, and feature selection are not applied to the complete dataset before cross-validation, thus avoiding overestimation and over-fitting. The classification algorithms used are linear, ridge, and Lasso regression, decision trees, random forests (RF), and support vector machines (SVMs) with Gaussian and polynomial kernels. For feature selection, JADBio uses the Lasso and Statistically Equivalent Signature (SES) [45] algorithms. JADBio applies best practices of ML to eliminate any over-fitting of the model and overestimation of its out-of-sample predictive performance, even for small sample sizes; a detailed evaluation of JADBio can be found in [43].

4.6. Custom Machine Learning Framework: XGBoost

The custom ML framework is depicted in Figure 1. A stratified train–test split of 80–20% is applied to the initial dataset, where the missing values are imputed by using mean imputation (for numerical features) and mode imputation (for categorical features). Then, the data preparation phase corresponds to the one-hot encoding of categorical features and correlation-based feature selection, where features with high Pearson correlation values are more linearly dependent, having almost the same effect on the dependent variable, and thus can be removed. Specifically, the pairwise Pearson correlation matrix is computed and the features with a correlation ratio higher than 0.9 are dropped. Bayesian optimization [46] is adopted during the hyper-parameter tuning and a stratified five-fold cross-validation strategy is followed. It is also important to notice that JADBio addresses imbalanced classes through stratified cross-validation and diversified class weights during SVM learning. Since the imbalanced ratio in the eICU dataset is 3457 vs. 267 class samples for classes 0 and 1, respectively, we aim at examining the class balancing effect in light of oversampling combined with a state-of-the-art ML classifier, where, in the current work, we investigate the XGBoost (eXtreme Gradient Boosting) classifier’s performance. Towards achieving a balanced ratio between the two classes, the Synthetic Minority Oversampling Technique (SMOTE) [47] is adopted. In each cross-validation fold, resampling is applied to the data, and as a final step, we use the hyper-parameters for which the cross-validation achieved the best performance as well as SMOTE on the total training set.

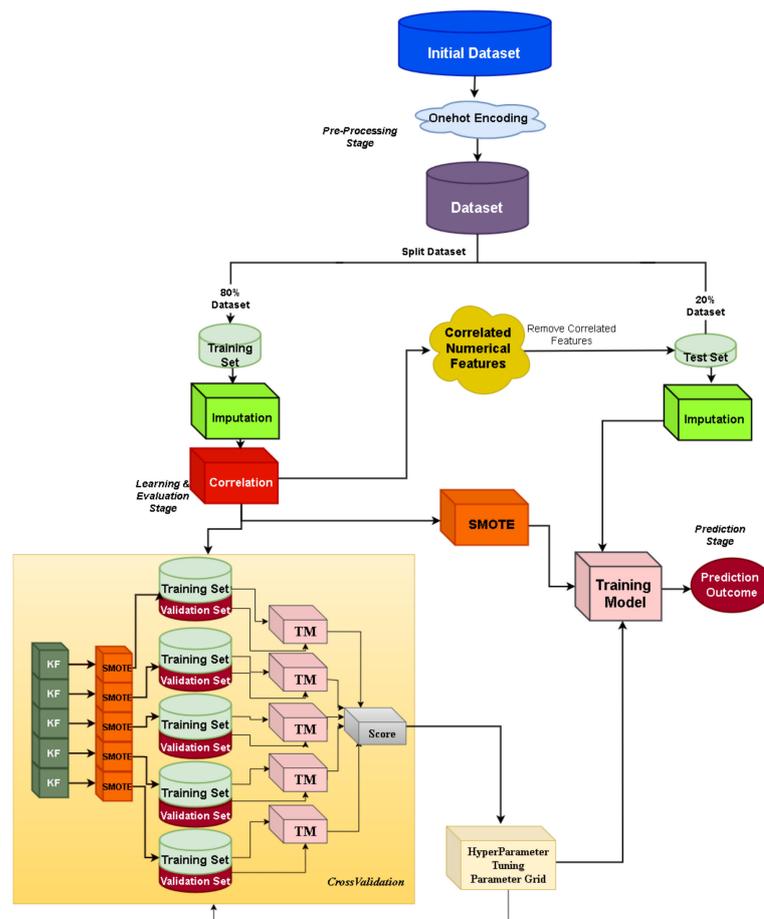


Figure 1. Custom machine learning pipeline: XGBoost.

4.7. Machine Learning Models’ Performance Evaluation

The performance of ML classification tasks is typically assessed by various well-known metrics, such as sensitivity (or recall), accuracy, specificity, F1 score, and area under

the receiver operating characteristic curve (*AUC-ROC*). The significance of the *AUCs* was measured using a significance level of $\alpha = 0.05$ [48]. The classification threshold was optimized for F1 score.

5. Results

In this section, we describe the predictive models of early and late mortality for patients with thrombosis or cancer, using an automated ML tool. Performance metrics, feature discriminative analysis, comparison with conventional scores, and validation of the model are reported in detail. In the last part of this section, the results of the custom ML framework that addresses the class imbalance problem are presented.

5.1. Automated Machine Learning

5.1.1. Prediction of Early Mortality in ICU Patients with Thrombosis

As early or in-hospital mortality define the outcomes of patients at discharge from the hospital, two different datasets were used and two groups of patients with thrombosis were extracted: from the MIMIC-III database, 348 non-survivors vs. 1303 survivors, and from the eICU database, 267 non-survivors vs. 3457 survivors.

MIMIC-III Dataset

Prediction of early mortality in patients with thrombosis, using all clinico-laboratory features, achieved excellent performance (*AUC-ROC* = 0.93, CI 0.91–0.95), where CI denotes the confidence interval. The winning algorithm was random forest (RF), training 500 trees with the deviance splitting criterion and minimum leaf size = 3. Detailed information regarding metrics of performance and feature discriminative analysis can be found in [41]. As shown in Figure 2 and Appendix A Table A3, our model significantly outperformed nine well-known traditional medical scores, such as SAPS [17] and SOFA [19]. Specifically, among the various scores, the best performance was achieved by SAPSII (*AUC-ROC* = 0.85, CI 0.81–0.89), and when a combination of all available scores was used, the achieved performance was *AUC-ROC* = 0.86, CI 0.81–0.88.

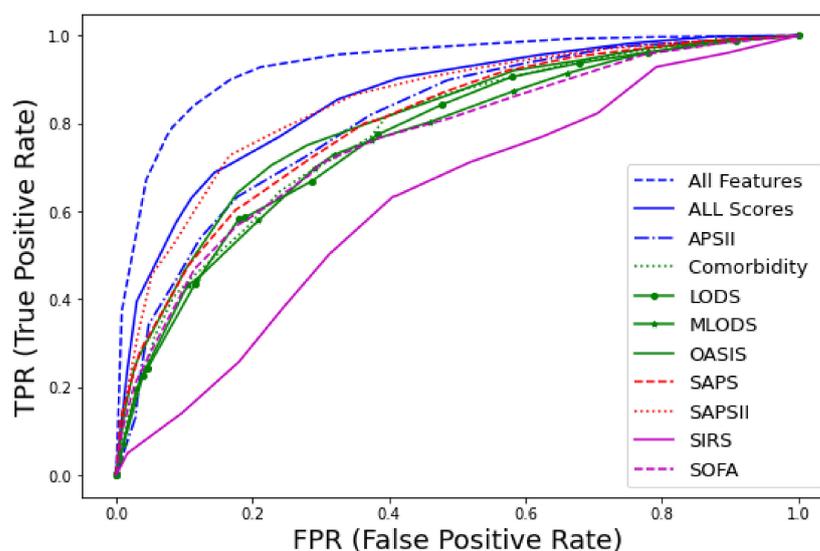


Figure 2. Cont.

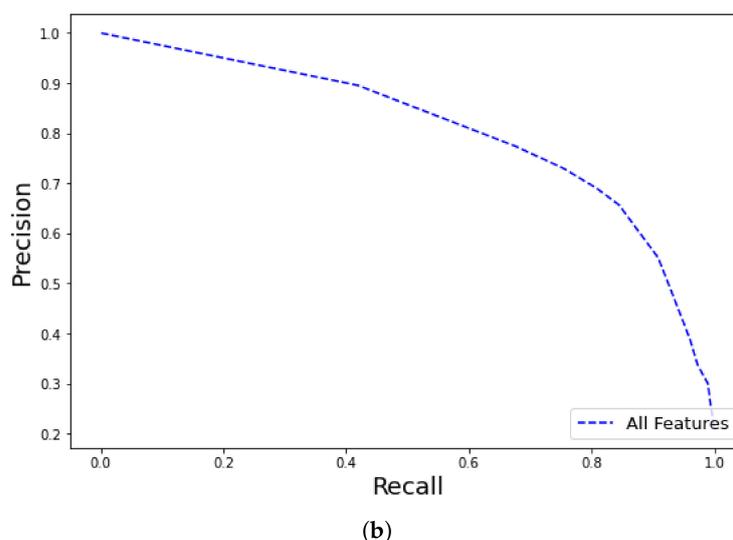


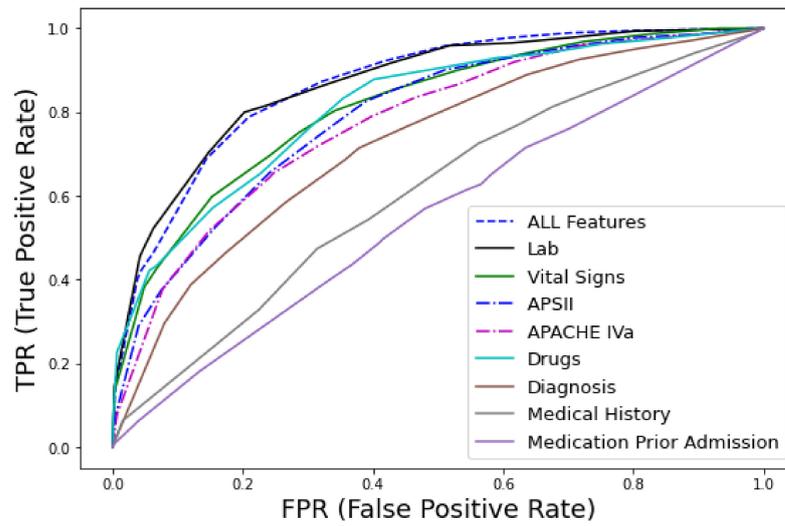
Figure 2. Receiver operating characteristic (ROC) curves for early mortality in ICU patients with thrombosis from MIMIC-III database using all features: (a) AUC–ROC curves and comparison with traditional scores (APS, Acute Physiology Score; LODS, Logistic Organ Dysfunction Score; MLODS, Multiple Logistic Organ Dysfunction Score; OASIS, Outcome and Assessment Information Set; SAPS, Simplified Acute Physiology Score; SIRS, Systemic Inflammatory Response Syndrome; SOFA, Sequential Organ Failure), (b) Precision–recall (PR) curve.

eICU Dataset

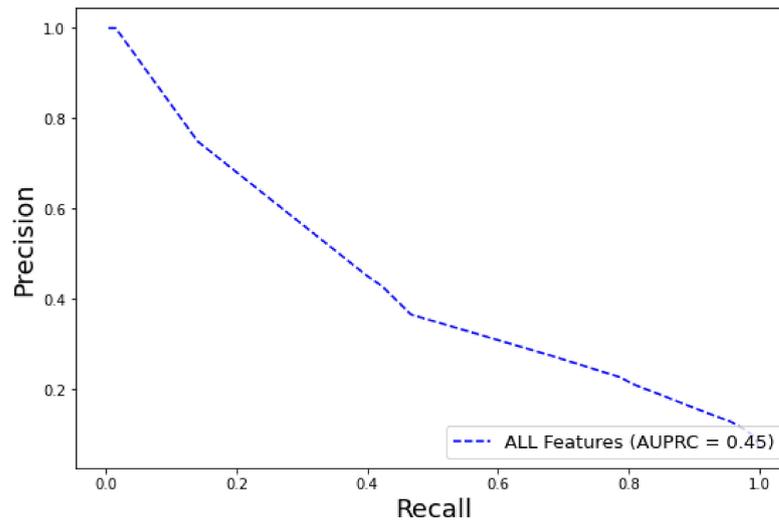
The RF classifier (training 1000 trees with deviance splitting criterion, minimum leaf size = 5) was chosen as the winning algorithm in the AutoML approach, using the extended feature approach. Among the various feature groups, the model with the best performance was trained with the dataset containing *all features* (AUC–ROC = 0.87, CI 0.84–0.9) and *labs* (AUC–ROC = 0.87, CI 0.83–0.9), followed by *drugs* (AUC–ROC = 0.82, CI 0.77–0.86) *vital signs* (AUC–ROC = 0.81, CI 0.76–0.85), whereas *medical history* (AUC–ROC = 0.6, CI 0.56–0.64) and *medications prior to admission* (AUC–ROC = 0.55, CI 0.5–0.59) had the worst performance, as shown in Figure 3. The AUC of the precision–recall curve was 0.45 regarding early mortality in ICU patients with VTE from the eICU dataset. The use of the extended feature set significantly outperformed traditional clinical scores APSIII and APACHE IVa. A detailed comparison of the various metrics of performance between the various feature groups and clinical scores is shown in Table 2. From an initial number of 2300 attributes in *all features* and 891 in the *labs* subsets, the Test-Budgeted Statistically Equivalent Signature (SES) algorithm (hyper-parameters: maxK = 2, alpha = 0.05, and budget = 3 · nvars) used a signature of only 25 features, predictive of early mortality, as shown in Figure 4a,b, respectively.

Table 2. Detailed metrics of performance for the predictive models of early mortality in patients with thrombosis from the eICU dataset (JADBio).

	AUC–ROC	F1 Score	Accuracy	Specificity	Sensitivity
All features	0.87 (0.84, 0.9)	0.42	0.92	0.95	0.42
Labs	0.87 (0.83, 0.9)	0.43	0.92	0.95	0.50
Vital Signs	0.83 (0.78, 0.87)	0.37	0.92	0.95	0.38
APSIII	0.79 (0.75, 0.84)	0.32	0.92	0.96	0.29
APACHE IVa	0.78 (0.73, 0.82)	0.30	0.89	0.92	0.38
Drugs	0.82 (0.77, 0.86)	0.53	0.94	0.99	0.23
Diagnosis	0.73 (0.68, 0.77)	0.26	0.82	0.92	0.30
Medical History	0.60 (0.56, 0.64)	0.27	0.93	0.98	0.06
Medication prior to admission	0.55 (0.5, 0.59)	0.41	0.92	0.99	0.01

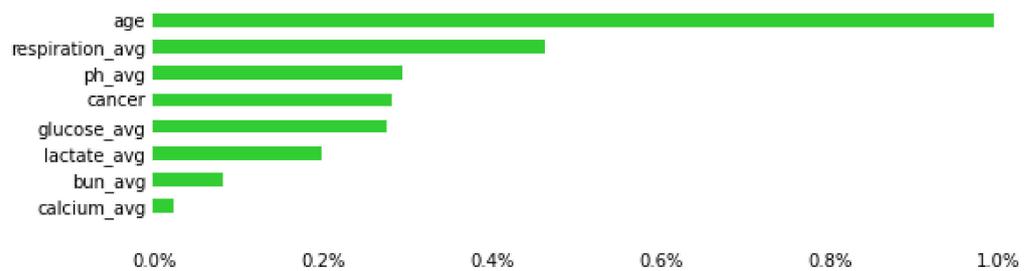


(a)



(b)

Figure 3. ROC curves for early mortality in ICU patients with VTE from eICU dataset: (a) AUC–ROC curves, feature discriminative analysis and comparison with clinical scores (APS, Acute Physiology Score; APACHE, Acute Physiology Age Chronic Health Evaluation), (b) Precision–recall curve.



(a)

Figure 4. Cont.

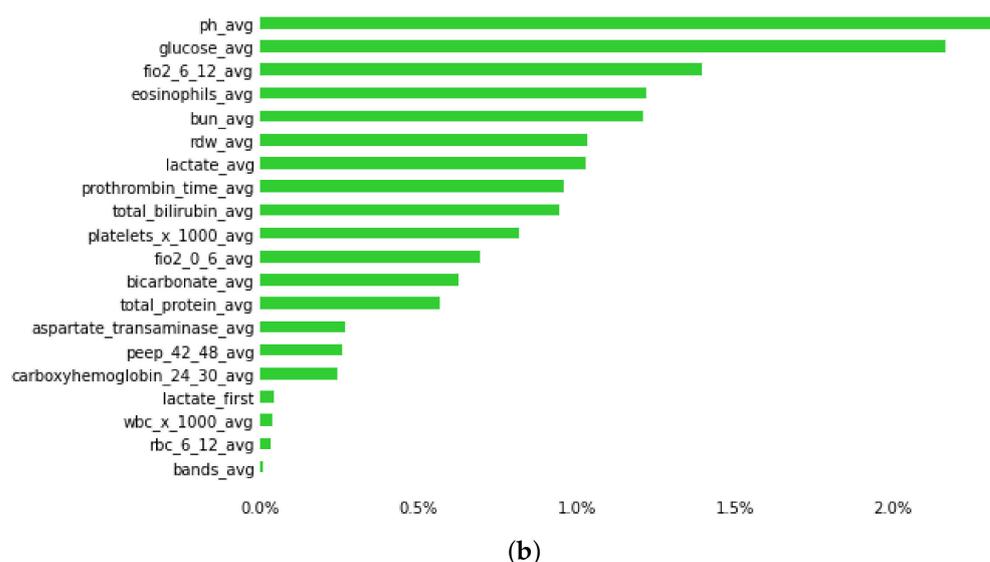


Figure 4. Plot showing predictive features of early mortality in ICU patients with thrombosis, using (a) *allfeatures* and (b) *labs* datasets. Green bars represent the average percentage drop in predictive performance when the feature is removed from the model. (Abbreviations: avg, average; bun, blood urea nitrogen; fio2: fraction of inspired oxygen; rdw, red cell distribution width; peep, positive end expiratory pressure; wbc, white blood cells; 0–6, 6–12, 24–30, 42–48 is the time in hours after admission).

5.1.2. Prediction of Early Mortality in ICU Patients with Cancer

We extracted 902 non-survivors vs. 1757 surviving patients with various solid cancer and hematological malignancies from the MIMIC-III database. SVM of type C-SVC with linear kernel (cost = 1.0) was the best-performing model for early mortality prediction ($AUC-ROC = 0.94$, CI 0.92–0.96). The Lasso feature selection [45] (penalty = 1.0, $\lambda = 4.183 \times 10^{-2}$) algorithm revealed the following features with high predictive performance: endotracheal intubation, 1st day respiratory rate, coexistence of metastatic cancer, albumin, systolic blood pressure, and Red Cell Distribution Width (RDW) (Appendix A Table A2). Using the all features set outperformed traditional medical scores OASIS ($AUC-ROC = 0.83$, CI 0.79–0.87), SAPS ($AUC-ROC = 0.86$, CI 0.82–0.9), and SOFA ($AUC-ROC = 0.78$, CI 0.73–0.83), as shown in Figure 5 and Table 3.

Table 3. Detailed metrics of performance for prediction of early and late mortality in ICU patients with cancer using all features.

Mortality Prediction	AUC-ROC	Accuracy	F1 Score	Specificity	Sensitivity
m0	0.94 (0.92, 0.96)	0.88	0.82	0.91	0.82
m1	0.88 (0.85, 0.91)	0.85	0.89	0.84	0.78
m3	0.84 (0.8, 0.87)	0.79	0.85	0.74	0.79
m6	0.78 (0.74, 0.82)	0.72	0.49	0.72	0.71
m12	0.76 (0.71, 0.8)	0.72	0.5	0.72	0.7
m > 12	0.74 (0.69, 0.74)	0.68	0.76	0.64	0.74

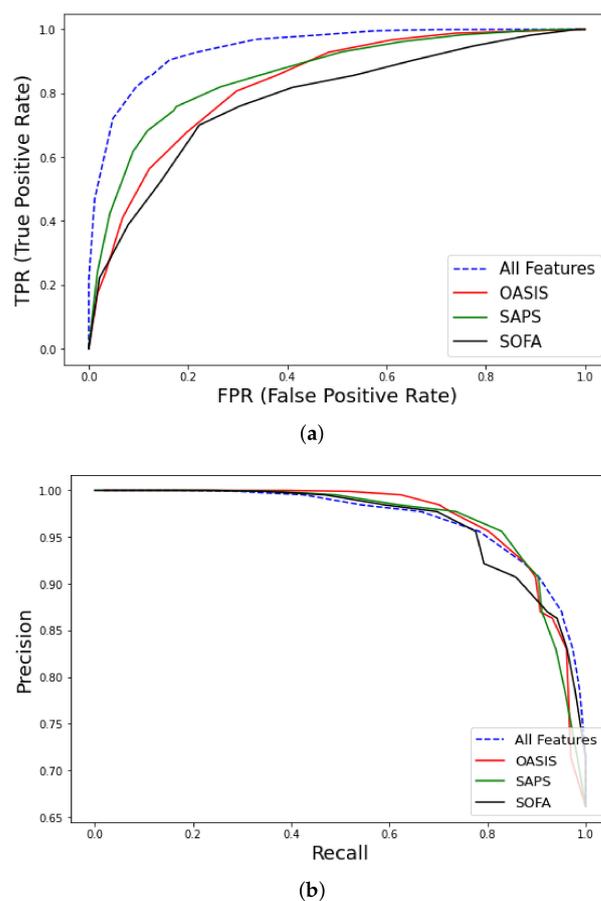


Figure 5. ROC curves for early mortality in ICU patients with cancer from MIMIC-III database using all features. (a) *AUC-ROC* curve and comparison with clinical scores (OASIS, Outcome and Assessment Information Set; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure), (b) Precision-Recall curve.

5.1.3. Prediction of Late Mortality in ICU Patients with Thrombosis

Late mortality is defined as mortality after ICU or hospital discharge, as recorded in a later admission or outside the hospital. The MIMIC-III database is suitable for late mortality studies, since it offers longitudinal follow-up information regarding survival, for months after their admission to the ICU, in contrast to the eICU database, where such information is missing and it is not possible to chronologically order hospital admissions for the same patient within the same year. Moreover, no censor outcomes are recorded, since each admission has specific time stamps. For this binary classification task, 817 non-survivors vs. 1303 survivor patients with VTE from MIMIC-III were included. On average, patients with VTE died 549 days after admission, with a median of 225 days.

The best ML model was RF training 500 trees with the deviance splitting criterion and minimum leaf size = 3. As expected, the task of predicting late mortality was less efficient than that of the early mortality even using the whole feature space (*AUC-ROC* = 0.82, CI 0.79–0.84 vs. 0.93, CI 0.91–0.95). Although traditional clinical scores were originally designed to predict in-hospital mortality, their performance in predicting late mortality is moderate and acceptable. The SAPSII (*AUC-ROC* = 0.76, CI 0.72–0.81) and Elixhauser comorbidity scores (*AUC-ROC* = 0.74, CI 0.69–0.79) had the highest predictive performance from the rest of the studied scores and were close to the use of a combination of the available scores as shown in Figure 6.

Detailed performance metrics for both tasks of predicting early and late mortality using all features in patients with thrombosis, and from both datasets, are summarized in Table 4.

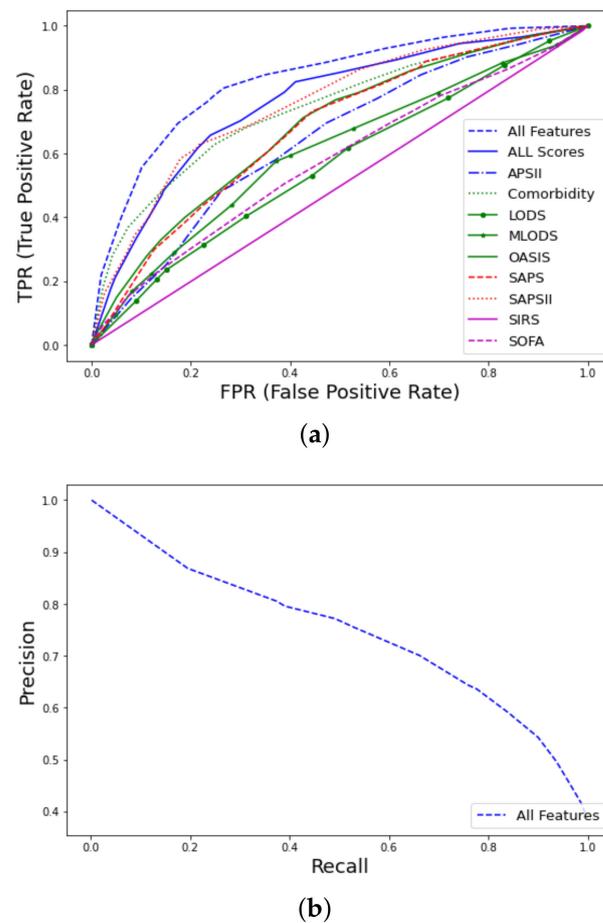


Figure 6. ROC curves for late mortality in ICU patients with thrombosis from MIMIC-III database using all features: (a) *AUC–ROC* curves and comparison with clinical scores. (b) PR curve.

Table 4. Detailed metrics of performance for prediction of early and late mortality of ICU patients with thrombosis using all features.

	Early Mortality		Late Mortality
	MIMIC-III	eICU	MIMIC-III
<i>AUC–ROC</i>	0.93	0.87	0.82
Accuracy	0.89	0.92	0.76
F1 score	0.72	0.97	0.60
Sensitivity	0.67	0.99	0.49
Specificity	0.95	0.1	0.90

5.1.4. Prediction of Late Mortality in ICU Patients with Cancer

Prediction of late mortality for cancer patients was further stratified based on the time of death after admission (in months). When constructing the predictive models based on time (months) after admission, *AUC–ROC* for months m_1 , m_3 , m_6 , m_{12} and $m_{>12}$ were 0.88, 0.84, 0.78, 0.76, 0.74, respectively.

Detailed performance metrics are shown in Table 3 and Figure 7. The best performance in the prediction of late mortality was achieved using *all features*, which outperformed classic clinical scores, especially in terms of F1 score, specificity, and sensitivity. Among the clinical scores, SAPS showed modest performance to predict mortality at 1 month after admission (*AUC–ROC* = 0.82, CI 0.78–0.85) and at 3 months after admission (*AUC–ROC* = 0.77, CI 0.73–0.81), although significantly lower than using the *all features* set. Features that were selected to have high predictive performance can be found in

Appendix A Table A2. The presence of metastatic cancer was the strongest predictor of mortality, whereas, interestingly, transfusions with red blood cells were the strongest predictor of mortality for more than a year after ICU admission.

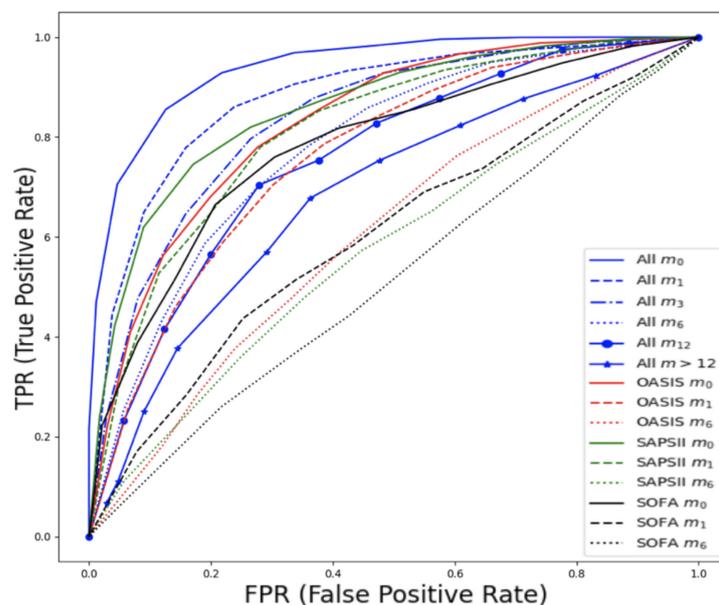


Figure 7. ROC curves for late mortality in ICU patients with cancer from MIMIC-III database, stratified based on the time of death after admission expressed in months (m_0 , m_1 , m_3 , m_6 , m_{12} , $> m_{12}$). Comparison with traditional clinical scores.

5.1.5. Model Validation

The predictive model for early mortality in patients with thrombosis derived from the MIMIC-III database ended with a signature of 35 predictive features that included features from full blood cell count (white blood cells (WBC), eosinophils, lymphocytes, and red cell distribution width (RDW)), biochemistry markers (glucose, blood urea nitrogen (BUN), calcium, total protein, albumin, potassium, lactate, lactate dehydrogenase (LDH), creatine phosphokinase (CPK), anion gap, arterial oxygen saturation (SaO₂), pH), vital signs (heart rate, respiration, lowest systolic blood pressure (BP), and current diastolic BP), hemostasis (prothrombin time (PT), international normalised ratio (INR)), clinical information (cancer, sepsis, Glasgow Coma Scale (GCS)), and medications (vasopressors such as epinephrine, dopamine, and vasopressin and warfarin use). This model has been later validated in the eICU dataset, as shown in Table 5 and Figure 8.

Table 5. External validation of a predictive model for early mortality in patients with VTE based on a signature of 35 features.

	AUC-ROC	F1 Score	Accuracy	Specificity	Sensitivity
MIMIC-III	0.93 [0.91, 0.96]	0.74	0.77	0.93	0.74
eICU	0.89 [0.87, 0.91]	0.56	0.86	0.89	0.7

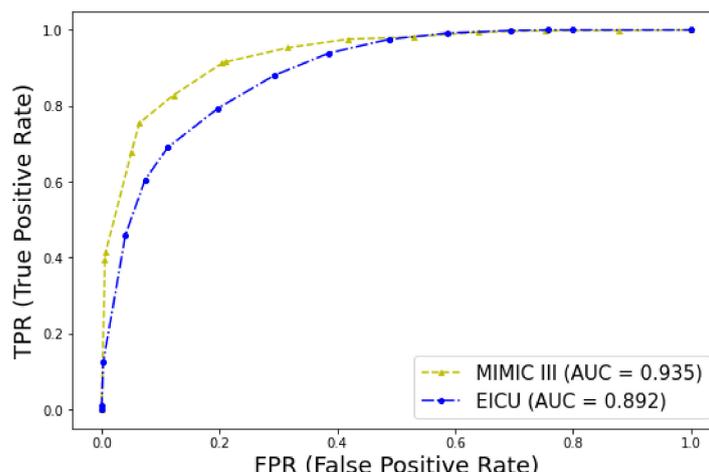


Figure 8. ROC curves for early mortality in ICU patients with thrombosis. Validation of the model in the two datasets.

5.2. Custom Machine Learning Framework: XGBoost

The best XGBoost classifier configuration (after hyper-parameter tuning) is with learning rate: 0.15, maximum tree depth: 12 gamma: 0.1, minimum sum of instance weight: 1, colsample_bytree: 0.7. Similarly to the AutoML framework, among the various feature groups, the model that achieved the best performance was the one trained with the dataset containing *all features* (AUC–ROC = 0.84, CI 0.83–0.85), as shown in Figure 9. The precision–recall curve AUC was 0.3 regarding early mortality in ICU patients with VTE from the eICU dataset, and this low score can be attributed to the extremely high imbalance class ratio. The detailed metrics of performance are shown in Table 6.

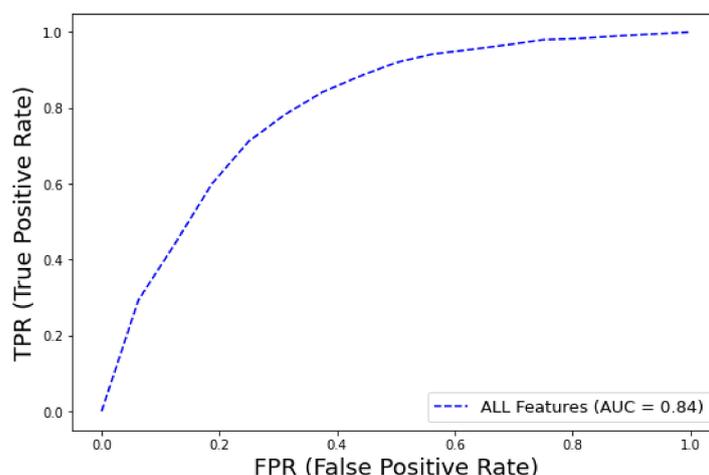


Figure 9. ROC curve for early mortality in ICU patients with VTE from eICU dataset (XGBoost).

Table 6. Detailed metrics of performance for prediction of early mortality of ICU patients with thrombosis using all features.

XGBoost	All Features
AUC–ROC	0.84 (0.83–0.85)
Accuracy	0.87
F1 score	0.63
Sensitivity	0.29
Specificity	0.95

6. Discussion

The main contribution of the current work is the investigation of an extended number of clinico-laboratory features normally stored in the electronic health record, grouped into clinically meaningful sets (e.g., vital signs, labs, medications, and procedures) and in time stamps to examine their impact on the prediction of early and late mortality in ICU patients. Two homogeneous population cohorts based on their diagnosis (specifically venous thromboembolism and/or cancer) were derived using two different open-access large healthcare datasets, MIMIC-III and eICU. Patients with a DNR code were excluded as it is known that DNR is an independent factor of dismal survival outcomes in the ICU [49]. Specifically, we compared different state-of-the-art ML algorithms, addressed the class imbalance problem in the medical datasets, developed interpretable models and identified clinically meaningful predictive signatures, compared the performance of the ML approach with existing scoring systems, and finally provided an external validation of the model, proving the generalizability of it. Similarly to Choi et. al. [10], our model outperformed traditional clinical scores not only in the prediction of early mortality but also in the prediction of late mortality. A multi-dimensional time series data-driven research approach was used, as well as stratification over the different groups of features, to identify feature groups with the highest predictive performance.

Towards constructing a robust model, training of two different ML strategies has been employed, an AutoML and a custom approach based on the XGBoost algorithm [50]. The primary aim of using the two approaches was not to compare them, since this it would be unfair, but to address the class imbalance through an oversampling method such as SMOTE [47]. Typically, class-imbalanced datasets constitute a common problem in medical informatics, which might lead to degraded performance depending on the type/number of data, features, etc., and thus an additional analysis should be performed in order to tackle this issue. JADBio addresses imbalanced classes through stratified cross-validation and diversified class weights during SVM learning. We thought that adopting the SMOTE method, which is considered a typical class balancing algorithm within the oversampling techniques framework [51], would be of added value. In the custom approach, we observed that the predictive accuracy was improved when features with more than XX missing values were removed. However, the accuracy appeared to remain constant despite the adoption of the SMOTE oversampling technique. Since SMOTE is a deterministic resampling method that selects examples being close in the local feature space, a probabilistic approach, such as Generative Adversarial Networks [52], which originally have been used in the area of image processing to produce synthetic but realistic images, could be used to produce new samples within a medical data analysis framework by learning from the overall class distribution. Moreover, the experimentation and extraction of the best-performing ML model in the custom approach is time-consuming since it requires substantial human and computational effort, artificial intelligence expertise, and extensive tuning of hyper-parameters; for this reason, automated ML tools are becoming popular among non-specialists in this area.

The AutoML approach based on the JADBio platform has been widely tested in biomedical data and follows all good practices for analysis and efficiency reporting [42]. During the experiments, the RF model was consistently found to be the winning algorithm, with a few exceptions. A frequent problem in various studies [53] that use ML predictive models is whether proper ML guidelines for over-fitting prevention and accurate performance metrics are reported. It is obvious from the experimental evaluation that JADBio can handle more efficiently than the custom ML framework high-dimensional datasets having a high level of missing values. Prediction of early mortality in patients with thrombosis was less efficient in the highly imbalanced dataset from the eICU compared with the one from the MIMIC-III dataset. The difference in performance, as illustrated in the precision–recall curve when compared to the ROC curve, is attributed to the highly imbalanced dataset. In this sense, for the minority class, high recall can only come at the cost of low precision. Feature discriminative analysis revealed that follow-up laboratory tests and vital signs in the medical record have the higher predictive performance when building prognostic

models for early as well as late mortality, outperforming traditional clinical scores. Among traditional scores, the best performance for the prediction of early mortality in patients with thrombosis or cancer was demonstrated for SAPS [17]. Regarding late mortality in patients with thrombosis, SAPS and comorbidity showed modest performance, whereas in cancer patients, SAPS could modestly predict mortality up to 3 months after ICU admission.

The interpretability of the ML predictive models is considered a prerequisite for physicians in order to accept them as clinical decision support systems. To this end, both of our approaches produced interpretable and comparable models. The automated ML approach can produce “biosignatures” that can be intuitively explored and explained by physicians [43], as confirmed by this study. For feature selection, Statistically Equivalent Signature (SES) algorithms, inspired by the principles of constrained-based learning of Bayesian networks, were consistently found to be superior to the feature selection method Lasso. XGBoost has a built-in function for feature importance but different metrics can be used, which could lead to misleading results [54]. Although both approaches derived similar features, the ranking of importance was different.

Among the thousands of features that we extracted from the electronic health records of the patients with thrombosis, a few features were selected that were clinically meaningful, such as older age, cancer, respiratory, cardiovascular, and renal disease, vasopressor support, and mechanical ventilation, which are well established clinical predictors of ICU mortality [13]. Similarly to [13], sex was not found to be a predictor of ICU mortality. Moreover, individual feature analysis confirmed that warfarin [55], RDW [56], red blood cell transfusions [57], and blood urea nitrogen [58] are significant predictors of early and possibly long-term mortality. RDW has been shown to play a significant negative predictive role in ICU early mortality through the deregulation of erythropoiesis from inflammatory cytokines and oxidative stress [59]. It has also been reported to be an independent risk factor for cardiovascular diseases, dyslipidemia, diabetes, and renal and liver diseases. Surprisingly, high RDW has been shown to correlate with cancer stage irrespective of comorbidities, and with early mortality in VTE patients. For all these reasons, it is not paradoxical that RDW could be an easily applicable, new biomarker, useful not only for the prediction of early but possibly of late mortality in patients with thrombosis and/or cancer. Another interesting feature revealed by our prognostic model is eosinophil count, which is known to have prognostic significance in ICU patients [60]. Markers such as RDW and eosinophils are attractive since they are easily available and of low cost.

In patients with cancer, the strongest predictor of early and late ICU mortality is the presence of metastatic cancer. Red blood cell transfusions are a negative predictor of early mortality, as already known [61]. Interestingly, transfusions were found to be the strongest predictor of late mortality, more than one year after admission to the ICU, for patients with cancer. Red cell transfusions in patients with cancer not only increase the risk of death but also the risk of relapse [61]. Unfortunately, information regarding transfusion (red blood cells, plasma, and platelets) is missing in a significant number of patients, whereas in MIMIC-III, this information is scattered across two different information systems that collect data (Metavision and Carevue), and again, a significant number of data points are missing.

Head to head comparisons of the various studies in ICU mortality prediction are difficult, since the various studies have different inclusion and exclusion criteria, different types of studied features, and various definitions of mortality. Our study targeted two specific groups of patients, patients with venous thromboembolism and patients with cancer. Both diagnostic groups are high-risk patients, with a substantial risk of ICU admission and mortality. Mortality prediction models for ICU patients with thrombosis or cancer that are based on ML algorithms and use a large amount of clinical and laboratory data, structured and unstructured, are almost completely absent in the literature. Moreover, traditional scoring systems are not specific for these two diseases. To the best of our knowledge, only one publication on a relatively small number of patients with venous thromboembolism has been published [32]. Similarly to Runnan et al., we compared state-of-the-art ML algorithms

with traditional scores, and we achieved comparable performance and identified similar predictive features.

In medical machine learning/predictive models, external validation in different datasets is imperative before clinical application. Classifiers usually perform well in the original dataset from which they were trained but then perform poorly in independent datasets. Another important consideration is that ML models in clinical practice are to be deployed in different institutions or countries. A recent systematic review identified only 5 out of 70 publications that used independent data to externally validate their model to predict mortality [53]. One of the strengths of our study is that it used two independent datasets to externally validate the ML prognostic model of early mortality derived from the MIMIC-III dataset to the eICU dataset, which are derived from different institutions and have different time periods. The results of the external validation were promising, with the exception of the F1 score, which was inferior.

Some limitations of this study should be considered. First of all, the study was based on retrospective data. Since the data were collected in the past, it is possible that many medical practices have changed over time, such as the case of warfarin use in the ICU. Second, the selection of the studied diagnostic groups was based solely on ICD-9 codes [62] and DRG codes [63], and not on imaging studies. Third, time series data were processed in specific time stamps, which increased significantly the dimensionality of the data. Moreover, we observed that labs and vital signs in both datasets were infrequently reported in the first 48 h, thus leading to a dramatically increased number of missing data on the various time stamps. Fourth, a direct comparison of our model with the only PE-specific score, PESI, was not possible, since this is not included in the datasets.

One of the primary goals of our future work is to directly compare our model with the PESI score and in a prospective cohort study. Inclusion of more features, such as genetic information and imaging studies, would be ideal and would probably improve the predictive performance. We could also focus on features extracted on the day of discharge to predict other outcomes, such as ICU readmission. Our future vision is to develop an intelligent ML-based system that is continuously updated with new clinical events and detailed information of the current clinical status of the patient, which could be a useful assistant for the physician and their clinical decision-making. To this end, the use of deep learning models, such as long short-term memory (LSTM) [14] for importing time series data in high-frequency datasets, and neural networks [64] could probably achieve better generalization performance with a significantly lower error rate. Shapley additive explanation (SHAP) analysis could be used to explain the output of our predictive model [65]. Handling of the high imbalance ratio of the datasets could be performed with other advanced resampling methods, such as Generative Adversarial Networks [52].

7. Conclusions

The presented research could be used as a proof of concept study that could be further validated in prospective or more recent datasets. Prediction of in-hospital mortality in patients with thrombosis or cancer is highly feasible, whereas prediction of late mortality is a more difficult and complex task. The results of this study are promising and, most importantly, interpretable, since the predictive features included in the model were clinically meaningful. The discovery of novel biomarkers, such as RDW and eosinophils, and their incorporation into the traditional clinical scores could possibly refine their performance.

Author Contributions: Conceptualization: V.D.; Methodology: V.D., S.N., D.A., C.T. and D.M.; Software: S.N.; Validation: V.D., S.N., C.T., D.A. and D.M.; Formal analysis: V.D., S.N. and C.T.; Investigation: V.D.; Resources: V.D.; Data curation: V.D. and D.M.; Writing—original draft preparation: V.D. and S.N.; Writing—review and editing: V.D., S.N., C.T., D.A., T.K. and S.I.; Visualization: V.D., S.N. and C.T.; Supervision: V.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the European Union’s Horizon 2020 Research and Innovation Program, ASCAPE, under grant agreement No. 875351.

Institutional Review Board Statement: No Institutional Review Board Statement was required for the completion of this study.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data from MIMIC-III and eICU datasets are available upon request from Physionet (www.physionet.org, accessed on 19 April 2022). Researchers must first complete a course regarding HIPAA requirements and then sign a data use agreement for appropriate data usage. ICD-9 and ICD-10 codes used in the research to find patients with thrombosis or cancer, source codes for replicating the custom-ML pipeline, and links to JADBio experiments are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AC	Acute
ALP	Alkaline phosphatase
APACHE	Acute Physiology Age Chronic Health Evaluation
APS	Acute Physiology Score
AST	Aspartate transaminase
AUC	Area under the curve
AVG	Average
BUN	Blood urea nitrogen
COPD	Chronic obstructive pulmonary disease
CPK	Creatine phosphokinase
CV	CareVue
DRG	Disease-related group
DOACs	Direct oral anticoagulants
DVT	Deep vein thrombosis
FFP	Fresh frozen plasma
fiO ₂	Fraction of inspired oxygen
FPR	False Positive Rate
HIPAAA	Health Insurance Portability and Accountability Act
GCS	Glasgow Coma Scale
ICD	International Classification Code
ICU	Intensive care
LMWH	Low-molecular-weight heparin
LOS	Length of stay
LODS	Logistic Organ Dysfunction Score
m	Month
mean1stRR	Mean value of respiratory rate on the 1st ICU day
MD	Missing data
ML	Machine learning
MODS	Multiple Organ Dysfunction Score
MV	Metavision
NOF	Number of features
NOS	Not otherwise specified
OASIS	Outcome and Assessment Information Set
PaO ₂	Partial pressure of arterial oxygen
PaCO ₂	Partial pressure of arterial carbon dioxide
PE	Pulmonary embolism
PESI	Pulmonary Embolism Severity Index
sPESI	Simplified Pulmonary Embolism Severity Index
PLT	Platelet

PT	Prothrombin time
RBC	Red blood cell
RDW	Red cell distribution width
RF	Random Forest
ROC	Receiver operating characteristics
SBP	Systolic blood pressure
SD	Standard deviation
SaO2	Oxygen arterial saturation
SABER	Sequence Annotator for Biomedical Entities and Relations
SAPS	Simplified Acute Physiology Score
SES	Statistically Equivalent Signature
SIRS	Systemic Inflammatory Response Syndrome
SMOTE	Synthetic Minority Oversampling Technique
SOFA	Sequential Organ Failure
SVM	Support Vector Machine
TPR	True Positive Rate
VTE	Venous thromboembolism
WBC	White blood cells

Appendix A

Table A1. Description of attributes selected for patients with thrombosis and/or cancer from the various tables of eICU dataset. NOF denotes the number of features, and MD the missing data.

Group Name (as in eICU)	Table in eICU	Description	Number of Features	Missing Values (Merged with Patient)	Most Common Features
Patient ("patient")	Patient	Basic demographic information, LOS, discharge status	11	0	Gender, age, ethnicity, length of stay, admission height, admission weight, discharge status
Diagnosis "admission_dx"	Diagnosis	Diagnoses documented during ICU stay	61	395	PE, DVT, hypertension acute respiratory failure, altered mental status, diabetes, pneumonia, acute renal failure, congestive heart, anemia failure, chronic kidney, cancer, disease, sepsis
APSIII score	Apache-ApsVar	Acute Physiology Score (APSIII)	24	529	Variables used in APS III score, e.g., Glasgow respiratory rate, heart rate, temperature, coma scale, paO ₂ , etc.
APACHE IV score	Apache-PredVar	APACHE IV and IVa versions	34	529	Variables used in APACHE predictions, e.g., presence of cirrhosis, or metastatic carcinoma
Labs	Lab	Laboratory tests	80	442	Hematocrit hemoglobin, white blood cells, platelets, potassium, sodium, creatinine, nitrogen, chloride, glucose, calcium, bicarbonate, PT, INR, aPTT, blood urea MCH, MCHC, RDW
Vital signs	Physical Exam	Vital signs	8	597	Blood pressure diastolic and systolic current, highest, lowest, heart rate, GCS, fiO ₂ , SatO ₂ in 1 h intervals
	Vital Periodic	Vital signs	10	433	Heart rate, respiration, SatO ₂ , blood pressure, temperature

Table A1. *Cont.*

Group Name (as in eICU ²)	Table in eICU	Description	Number of Features	Missing Values (Merged with Patient)	Most Common Features
Medications prior to admission	Admission Drug	Medications taken prior to ICU admission	36	3263	Aspirin, furosemide bronchodilator metoprolol, warfarin, lisinopril, atorvastatin, inhaled corticosteroids insulin, amlodipin, levothyroxin, metformin, carbedilol, LMWH, clopidogrel, DOACs
	Infusion	Medications Transfusions Parenteral	30	2503	Heparin, epinephrine, norepinephrine, insulin, amiodarone, phenylephrine vasopressin, t-PA, diltiazem, dopamine
	Treatment	Medications	52	1040	Medications in groups, e.g., anticoagulants antibiotics, antiarrhythmics vasopressor, ulcer prophylaxis and several procedures, e.g., mechanical ventilation, IVC filter, thrombolysis, dialysis, embolectomy
Medical history	Pasthistory	Past history of chronic diseases	111	422	Hypertension, insulin-dependent diabetes, disease, cancer chronic heart failure, asthma, COPD, atrial fibrillation, stroke, DVT or PE within 6 months, dementia, peripheral vascular disease

Table A2. Most important selected features with predictive performance for early and late mortality in ICU patients with cancer. (Abbreviations: LOS, length of stay; SBP, systolic blood pressure; RDW, red cell distribution width; RR, respiratory rate; AST, aspartate transaminase; GCS, Glasgow Coma Scale; PaO₂, partial pressure of arterial oxygen; FFP, fresh frozen plasma).

1st Admission	Endotracheal Intubation	Min 1st RR	Metastatic Cancer	Albumin	Mean 1st RR	SBP	RDW
m1	Metastatic cancer	OASIS Elective surgery	LOS	Renal insufficiency	1st day chloride max	SAPSII score	PaO ₂
m3	Metastatic cancer	Open heart surgery	ALP	OASIS Pre-ICU LOS	GCS	1st day chloride max	Age
m6	Metastatic cancer	AST	Albumin	RDW	Aorto-coronary bypass	OASIS Elective surgery	
m12	Metastatic cancer	Age	SAPSII score	Etoposide	PaO ₂	FFP	Lung biopsy
>m12	Transfusions						

Table A3. Analytical metrics of performance of predictive models for early mortality in patients with thrombosis from MIMIC-III database (Abbreviations: APS, Acute Physiology Score; LODS, Logistic Organ Dysfunction Score; OASIS, Outcome and Assessment Information Set; SAPS, Simplified Acute Physiology Score; SIRS, Systemic Inflammatory Response Syndrome; SOFA, Sequential Organ Failure).

	AUC-ROC	F1 Score	Accuracy	Specificity	Sensitivity
	Early/Late	Early/Late	Early/Late	Early/Late	Early/Late
All features	0.94 (0.91–0.96)	0.87	0.83	0.93	0.79
All scores	0.85 (0.81, 0.88)	0.6	0.81	0.85	0.7
APS II	0.81 (0.76, 0.85)	0.55	0.8	0.85	0.6
Comorbidity	0.78 (0.74, 0.82)	0.5	0.736	0.76	0.65
LODS/MLODS	0.77 (0.72, 0.81)	0.5	0.77	0.8	0.6
OASIS	0.8 (0.76, 0.84)	0.55	0.76	0.77	0.7
SAPS	0.8 (0.76, 0.84)	0.53	0.78	0.82	0.6
SAPSII	0.85 (0.81, 0.89)	0.61	0.81	0.83	0.73
SIRS	0.64 (0.59, 0.68)	0.39	0.53	0.5	0.71
SOFA	0.76 (0.72, 0.81)	0.5	0.7	0.7	0.71

Table A4. Analytical metrics of performance of predictive models for late mortality in patients with thrombosis from the MIMIC-III database (Abbreviations: APS, Acute Physiology Score; LODS, Logistic Organ Dysfunction Score; OASIS, Outcome and Assessment Information Set; SAPS, Simplified Acute Physiology Score; SIRS, Systemic Inflammatory Response Syndrome; SOFA, Sequential Organ Failure).

	AUC-ROC	F1 Score	Accuracy	Specificity	Sensitivity
	Early/Late	Early/Late	Early/Late	Early/Late	Early/Late
All features	0.83 (0.79, 0.87)	0.72	0.74	0.75	0.78
All scores	0.76 (0.71, 0.81)	0.64	0.69	0.7	0.7
APS II	0.65 (0.59, 0.7)	0.56	0.62	0.53	0.69
Comorbidity	0.74 (0.69, 0.79)	0.62	0.69	0.69	0.68
LODS/MLODS	0.57 (0.51, 0.62)	0.46	0.62	0.56	0.53
OASIS	0.68 (0.63, 0.69)	0.59	0.63	0.51	0.77
SAPS	0.68 (0.63, 0.72)	0.59	0.62	0.55	0.73
SAPSII	0.76 (0.72, 0.81)	0.63	0.7	0.78	0.63
SIRS	0.49 (0.48, 0.5)	0.54	0.62	0.34	0.65
SOFA	0.58 (0.52, 0.63)	0.52	0.62	0.35	0.73

References

- Lozano, R.; Naghavi, M.; Foreman, K.; Lim, S.; Shibuya, K.; Aboyans, V.; Abraham, J.; Adair, T.; Aggarwal, R.; Ahn, S.Y.; et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2095–2128. [\[CrossRef\]](#)
- Crimmins, E.M. Lifespan and healthspan: Past, present, and promise. *Gerontologist* **2015**, *55*, 901–911. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hajat, C.; Stein, E. The global burden of multiple chronic conditions: A narrative review. *Prev. Med. Rep.* **2018**, *12*, 284–293. [\[CrossRef\]](#) [\[PubMed\]](#)
- Boonyawat, K.; Chantrathammachart, P.; Numthavaj, P.; Nanthatanti, N.; Phusanti, S.; Phuphuakrat, A.; Niparuck, P.; Angchaisuksiri, P. Incidence of thromboembolism in patients with COVID-19: A systematic review and meta-analysis. *Thromb. J.* **2020**, *18*, 34. [\[CrossRef\]](#)
- Martin, K.A.; Molsberry, R.; Cuttica, M.J.; Desai, K.R.; Schimmel, D.R.; Khan, S.S. Time trends in pulmonary embolism mortality rates in the United States, 1999 to 2018. *J. Am. Heart Assoc.* **2020**, *9*, e016784. [\[CrossRef\]](#)
- Zimmerman, J.E.; Kramer, A.A. A history of outcome prediction in the ICU. *Curr. Opin. Crit. Care* **2014**, *20*, 550–556. [\[CrossRef\]](#)
- Teixeira, C.; Kern, M.; Rosa, R.G. What outcomes should be evaluated in critically ill patients? *Rev. Bras. Ter. Intensiv.* **2021**, *33*, 312–319. [\[CrossRef\]](#)
- Sadeghi, R.; Banerjee, T.; Romine, W. Early hospital mortality prediction using vital signals. *Smart Health* **2018**, *9*, 265–274. [\[CrossRef\]](#)
- El-Rashidy, N.; El-Sappagh, S.; Abuhmed, T.; Abdelrazek, S.; El-Bakry, H.M. Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model. *IEEE Access* **2020**, *8*, 133541–133564. [\[CrossRef\]](#)

10. Choi, M.H.; Kim, D.; Choi, E.J.; Jung, Y.J.; Choi, Y.J.; Cho, J.H.; Jeong, S.H. Mortality prediction of patients in intensive care units using machine learning algorithms based on electronic health records. *Sci. Rep.* **2022**, *12*, 7180. [[CrossRef](#)]
11. Patel, S.; Singh, G.; Zarbiv, S.; Ghiassi, K.; Rachoin, J.S. Mortality Prediction Using SaO₂/FiO₂ Ratio Based on eICU Database Analysis. *Crit. Care Res. Pract.* **2021**, *2021*, 6672603. [[CrossRef](#)]
12. Holmgren, G.; Andersson, P.; Jakobsson, A.; Frigyesi, A. Artificial neural networks improve and simplify intensive care mortality prognostication: A national cohort study of 217,289 first-time intensive care unit admissions. *J. Intens. Care* **2019**, *7*, 44. [[CrossRef](#)] [[PubMed](#)]
13. Ho, K.M.; Knuiman, M.; Finn, J.; Webb, S.A. Estimating long-term survival of critically ill patients: The PREDICT model. *PLoS ONE* **2008**, *3*, e3226. [[CrossRef](#)] [[PubMed](#)]
14. Thorsen-Meyer, H.C.; Nielsen, A.B.; Nielsen, A.P.; Kaas-Hansen, B.S.; Toft, P.; Schierbeck, J.; Strøm, T.; Chmura, P.J.; Heimann, M.; Dybdahl, L.; et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: A retrospective study of high-frequency data in electronic patient records. *Lancet Digit. Health* **2020**, *2*, e179–e191. [[CrossRef](#)]
15. Simpson, A.; Puxty, K.; McLoone, P.; Quasim, T.; Sloan, B.; Morrison, D.S. Comorbidity and survival after admission to the intensive care unit: A population-based study of 41,230 patients. *J. Intens. Care Soc.* **2021**, *22*, 143–151. [[CrossRef](#)]
16. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)]
17. Le Gall, J.R.; Lemeshow, S.; Saulnier, F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* **1993**, *270*, 2957–2963. [[CrossRef](#)]
18. Knaus, W.A.; Wagner, D.P.; Draper, E.A.; Zimmerman, J.E.; Bergner, M.; Bastos, P.G.; Sirio, C.A.; Murphy, D.J.; Lotring, T.; Damiano, A.; et al. The APACHE III prognostic system: Risk prediction of hospital mortality for critically III hospitalized adults. *Chest* **1991**, *100*, 1619–1636. [[CrossRef](#)]
19. Lopes Ferreira, F.; Peres Bota, D.; Bross, A.; Mélot, C.; Vincent, J.L. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* **2001**, *286*, 1754–1758. [[CrossRef](#)]
20. Vincent, J.L.; Moreno, R. Clinical review: Scoring systems in the critically ill. *Crit. Care* **2010**, *14*, 207. [[CrossRef](#)]
21. Rapsang, A.G.; Shyam, D.C. Scoring systems in the intensive care unit: A compendium. *Indian J. Crit. Care Med.-Peer-Rev. Off. Publ. Indian Soc. Crit. Care Med.* **2014**, *18*, 220. [[CrossRef](#)] [[PubMed](#)]
22. Chen, L.M.; Martin, C.M.; Morrison, T.L.; Sibbald, W.J. Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Crit. Care Med.* **1999**, *27*, 1999–2004. [[CrossRef](#)] [[PubMed](#)]
23. Cosgriff, C.V.; Celi, L.A.; Ko, S.; Sundaresan, T.; Armengol de la Hoz, M.Á.; Kaufman, A.R.; Stone, D.J.; Badawi, O.; Deliberato, R.O. Developing well-calibrated illness severity scores for decision support in the critically ill. *NPJ Digit. Med.* **2019**, *2*, 76. [[CrossRef](#)] [[PubMed](#)]
24. Shafiq, A.; Lodhi, H.; Ahmed, Z.; Bajwa, A. Is the pulmonary embolism severity index being routinely used in clinical practice? *Thrombosis* **2015**, *2015*, 175357. [[CrossRef](#)]
25. Aujesky, D.; Roy, P.M.; Le Manach, C.P.; Verschuren, F.; Meyer, G.; Obrosky, D.S.; Stone, R.A.; Cornuz, J.; Fine, M.J. Validation of a model to predict adverse outcomes in patients with pulmonary embolism. *Eur. Heart J.* **2006**, *27*, 476–481. [[CrossRef](#)] [[PubMed](#)]
26. Aujesky, D.; Perrier, A.; Roy, P.M.; Stone, R.; Cornuz, J.; Meyer, G.; Obrosky, D.; Fine, M. Validation of a clinical prognostic model to identify low-risk patients with pulmonary embolism. *J. Intern. Med.* **2007**, *261*, 597–604. [[CrossRef](#)]
27. Jiménez, D.; Yusen, R.D.; Otero, R.; Uresandi, F.; Nauffal, D.; Laserna, E.; Conget, F.; Oribe, M.; Cabezudo, M.A.; Diaz, G. Prognostic models for selecting patients with acute pulmonary embolism for initial outpatient therapy. *Chest* **2007**, *132*, 24–30. [[CrossRef](#)]
28. Donzé, J.; Le Gal, G.; Fine, M.J.; Roy, P.M.; Sanchez, O.; Verschuren, F.; Cornuz, J.; Meyer, G.; Perrier, A.; Righini, M.; et al. Prospective validation of the pulmonary embolism severity index. *Thromb. Haemost.* **2008**, *100*, 943–948.
29. Kohn, C.G.; Mearns, E.S.; Parker, M.W.; Hernandez, A.V.; Coleman, C.I. Prognostic accuracy of clinical prediction rules for early post-pulmonary embolism all-cause mortality: A bivariate meta-analysis. *Chest* **2015**, *147*, 1043–1062. [[CrossRef](#)]
30. Churpek, M.; Yuen, T.; Winslow, C.; Meltzer, D.; Kattan, M.; Edelson, D. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit. Care Med.* **2016**, *44*, 368–374. [[CrossRef](#)]
31. Pirracchio, R.; Petersen, M.; Carone, M.; Rigon, M.; Chevret, S.; Laan, M. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *Lancet Respir. Med.* **2014**, *3*, 42–52. [[CrossRef](#)]
32. Runnan, S.; Gao, M.; Tao, Y.; Chen, Q.; Wu, G.; Guo, X.; Xia, Z.; You, G.; Hong, Z.; Huang, K. Prognostic nomogram for 30-day mortality of deep vein thrombosis patients in intensive care unit. *BMC Cardiovasc. Disord.* **2021**, *21*, 11. [[CrossRef](#)]
33. Purushotham, S.; Meng, C.; Che, Z.; Liu, Y. Benchmarking deep learning models on large healthcare datasets. *J. Biomed. Inform.* **2018**, *83*, 112–134. [[CrossRef](#)] [[PubMed](#)]
34. Lu, S.C.; Xu, C.; Nguyen, C.; Geng, Y.; Pfob, A.; Sidey-Gibbons, C. Machine Learning-based Short-term Mortality Prediction Models for Cancer Patients Using Electronic Health Record Data: A Systematic Review and Critical Appraisal (Preprint). *JMIR Med. Inform.* **2021**, *10*, e33182. [[CrossRef](#)]
35. Staudinger, T.; Stoiser, B.; Müllner, M.; Locker, G.; Laczika, K.; Knapp, S.; Burgmann, H.; Wilfing, A.; Kofler, J.; Thalhammer, F.; et al. Outcome and prognostic factors in critically ill cancer patients admitted to the intensive care unit. *Crit. Care Med.* **2000**, *28*, 1322–1328. [[CrossRef](#)]

36. Boer, S.; de Keizer, N.; Jonge, E. Performance of prognostic models in critically ill cancer patients—A review. *Crit. Care* **2005**, *9*, R458–R463. [CrossRef]
37. Johnson, A.E.; Pollard, T.J.; Lu, S.; Lehman, L.-w.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L.E.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035. [CrossRef]
38. Pollard, T.; Johnson, A.; Raffa, J.; Celi, L.A. Mark, R.G.; Badawi, O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci. Data* **2018**, *5*, 180178. [CrossRef]
39. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **2000**, *101*, e215–e220. [CrossRef]
40. Bader Lab. Saber (Sequence Annotator for Biomedical Entities and Relations). 2022. Available online: <https://baderlab.github.io/saber/> (accessed on 19 April 2022).
41. Danilatou, V.; Antonakaki, D.; Tzagkarakis, C.; Kanterakis, A.; Katos, V.; Kostoulas, T. Automated Mortality Prediction in Critically-ill Patients with Thrombosis using Machine Learning. In Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 26–28 October 2020; pp. 247–254.
42. Tsamardinos, I.; Charonyktakis, P.; Borboudakis, G.; Lakiotaki, K.; Zenklusen, J.C. Just Add Data: Automated Predictive Modeling for Knowledge Discovery and Feature Selection. *Nat. Precis. Oncol.* **2022**, *6*, 38. [CrossRef]
43. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. Just add data: Automated predictive modeling and biosignature discovery. *BioRxiv* **2020**. [CrossRef]
44. Tsamardinos, I.; Greasidou, E.; Borboudakis, G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach. Learn.* **2018**, *107*, 1895–1922. [CrossRef] [PubMed]
45. Lagani, V.; Athineou, G.; Farcomeni, A.; Tsagris, M.; Tsamardinos, I. Feature selection with the R package MXM: Discovering statistically-equivalent feature subsets. *arXiv* **2016**, arXiv:1611.03227. [CrossRef]
46. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. In Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012, NIPS'12, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 2951–2959.
47. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
48. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [CrossRef]
49. Fuchs, L.; Anstey, M.; Feng, M.; Toledano, R.; Kogan, S.; Howell, M.; Clardy, P.; Celli, L.; Talmor, D.; Novack, V. Quantifying the Mortality Impact of Do-Not-Resuscitate Orders in the ICU. *Crit. Care Med.* **2017**, *45*, 1019–1027. [CrossRef]
50. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'16, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
51. Santos, M.S.; Soares, J.P.; Abreu, P.H.; Araujo, H.; Santos, J. Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]. *IEEE Comput. Intell. Mag.* **2018**, *13*, 59–76. [CrossRef]
52. Ghosheh, G.; Li, J.; Zhu, T. A review of Generative Adversarial Networks for Electronic Health Records: Applications, evaluation measures and data sources. *arXiv* **2022**, arXiv:2203.07018.
53. Shillan, D.; Sterne, J.A.; Champneys, A.; Gibbison, B. Use of machine learning to analyse routinely collected intensive care unit data: A systematic review. *Crit. Care* **2019**, *23*, 284. [CrossRef]
54. Science, T. The Multiple faces of 'Feature Importance' in XGBoost. 2022. Available online: shorturl.at/oGU12 (accessed on 21 April 2022).
55. Fernando, S.M.; Mok, G.; Castellucci, L.A.; Dowlatshahi, D.; Rochweg, B.; McIsaac, D.I.; Carrier, M.; Wells, P.S.; Bagshaw, S.M.; Fergusson, D.A.; et al. Impact of anticoagulation on mortality and resource utilization among critically ill patients with major bleeding. *Crit. Care Med.* **2020**, *48*, 515–524. [CrossRef]
56. Fernandez, R.; Cano, S.; Catalan, I.; Rubio, O.; Subira, C.; Masclans, J.; Rognoni, G.; Ventura, L.; Macharete, C.; Winfield, L.; et al. High red blood cell distribution width as a marker of hospital mortality after ICU discharge: A cohort study. *J. Intensive Care* **2018**, *6*, 74. [CrossRef] [PubMed]
57. Wong, C.C.; Chow, W.W.; Lau, J.K.; Chow, V.; Ng, A.C.; Kritharides, L. Red blood cell transfusion and outcomes in acute pulmonary embolism. *Respirology* **2018**, *23*, 935–941. [CrossRef] [PubMed]
58. Arihan, O.; Wernly, B.; Lichtenauer, M.; Franz, M.; Kabisch, B.; Muessig, J.; Masyuk, M.; Lauten, A.; Schulze, P.C.; Hoppe, U.C.; et al. Blood Urea Nitrogen (BUN) is independently associated with mortality in critically ill patients admitted to ICU. *PloS ONE* **2018**, *13*, e0191697. [CrossRef] [PubMed]
59. Salvagno, G.L.; Sanchis-Gomar, F.; Picanza, A.; Lippi, G. Red blood cell distribution width: A simple parameter with multiple clinical applications. *Crit. Rev. Clin. Lab. Sci.* **2015**, *52*, 86–105. [CrossRef]
60. Yang, J.; Yang, J. Association between blood eosinophils and mortality in critically ill patients with acute exacerbation of chronic obstructive pulmonary disease: A retrospective cohort study. *Int. J. Chronic Obstr. Pulm. Dis.* **2021**, *16*, 281. [CrossRef] [PubMed]

61. Petrelli, F.; Ghidini, M.; Ghidini, A.; Sgroi, G.; Vavassori, I.; Petrò, D.; Cabiddu, M.; Aiolfi, A.; Bonitta, G.; Zaniboni, A.; et al. Red blood cell transfusions and the survival in patients with cancer undergoing curative surgery: A systematic review and meta-analysis. *Surg. Today* **2021**, *51*, 1535–1557. [[CrossRef](#)]
62. AHRQ. Clinical Classifications Software (CCS) for ICD-9-CM. 2022. Available online: <https://cutt.ly/7H0o4f8> (accessed on 19 April 2022).
63. Busse, R.; Geissler, A.; Aaviksoo, A.; Cots, F.; Häkkinen, U.; Kobel, C.; Mateus, C.; Or, Z.; O'Reilly, J.; Serdén, L.; et al. Diagnosis related groups in Europe: Moving towards transparency, efficiency, and quality in hospitals? *BMJ* **2013**, *346*, f3197. [[CrossRef](#)]
64. Ali, L.; Bukhari, S.A.C. An Approach Based on Mutually Informed Neural Networks to Optimize the Generalization Capabilities of Decision Support Systems Developed for Heart Failure Prediction. *IRBM* **2020**, *42*, 345–352. [[CrossRef](#)]
65. Lee, K.; Kha, H.; Nguyen, V.; Chen, Y.C.; Cheng, S.J.; Chen, C.Y. Machine Learning-Based Radiomics Signatures for EGFR and KRAS Mutations Prediction in Non-Small-Cell Lung Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 9254. [[CrossRef](#)]