

Developing a Machine Learning prediction model for bedside decision support by predicting readmission or death following discharge from the Intensive Care unit

Patrick J. Thoral (✉ p.thoral@amsterdamumc.nl)

Amsterdam UMC, location VUmc <https://orcid.org/0000-0001-6140-7195>

Mattia Fornasa

Pacmed BV

Daan P. de Bruin

Pacmed BV

Hidde Hovenkamp

Pacmed BV

Ronald H. Driessen

Amsterdam UMC, location VUmc, Vrije Universiteit Amsterdam

Armand R.J. Girbes

Amsterdam UMC, location VUmc, Vrije Universiteit Amsterdam

Mark Hoogendoorn

Computational Intelligence Group, Department of Computer Science, Vrije Universiteit Amsterdam

Paul W.G. Elbers

Amsterdam UMC, location VUmc, Vrije Universiteit Amsterdam

Research

Keywords: machine learning, prediction model, decision support, discharge, readmission, mortality

Posted Date: January 27th, 2020

DOI: <https://doi.org/10.21203/rs.2.21940/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Unexpected ICU readmission is associated with longer length of stay and an increase in mortality. Real time support systems could prevent untimely discharge from the ICU. We aim to develop a machine learning model for implementation at the bedside by predicting the risk of ICU readmission or death at time of potential discharge, showing feature importance and visualizing day-to-day changes in risk.

Methods

Data from adult patients, admitted to our mixed surgical-medical ICU between 2004 and 2016, were used in the analysis. Patient characteristics, clinical observations, (automated) physiological measurements, laboratory studies and treatment data were considered as model features. Different supervised learning algorithms were trained to predict ICU readmission and/or death, both within 7 days from ICU discharge, using 10-fold cross-validation. Feature importance was determined using SHapley Additive exPlanations. We constructed readmission probability-time curves to identify subgroups.

Results

Our dataset included 14,105 admissions. The combined readmission/mortality rate within seven days of ICU discharge was 5.3%. Using Gradient Boosting, the model achieved a Receiver Operating Characteristic AUC of 0.802 (95% CI 0.789-0.816) and a Precision-Recall AUC of 0.198 (95% CI 0.185-0.211). The most predictive features were well-known parameters, including physiological parameters, as well as less apparent features like nutritional support. Impact analysis using probability-time curves identified specific patients groups, that might lead to a change in discharge management with a relative risk reduction of 17%.

Conclusions

We developed a model that can accurately predict readmission and mortality after ICU discharge. Impact analysis showed that a relative risk reduction of 17% could be achievable. Given the large and increasing number of ICU admissions worldwide, this modest reduction may have significant impact for patients and society.

Background

The intensive care unit (ICU) is a well-equipped and well-staffed environment where treatment is delivered to patients that is not readily available elsewhere in the hospital. Thus, there is a so-called treatment gap between the ICU and the general hospital ward. Therefore, the decision to discharge patients from the ICU to the hospital ward following clinical recovery is often both challenging for the ICU team with respect to optimal timing and stressful to patients due to lack of continuous monitoring and lower availability of nurses and physicians on the ward [1]. The transition of care to different healthcare providers in a less

monitored environment may lead to preventable errors and adverse events, including ICU readmission and death [2, 3]. In addition, unexpected ICU readmission is associated with longer length of stay and an increase in mortality [4, 5].

ICU readmissions may be prevented by delaying discharge from the ICU, though optimal timing is subject of debate [6, 7]. To complicate matters, ICU stays that are unnecessarily long, prevent timely admission of other patients requiring ICU treatment which may affect their outcome. This problem may be especially prominent in tertiary teaching hospitals where ICU capacity is often strained [8]. Thus, preventing untimely ICU discharge is of pivotal importance.

The decision to discharge a patient from the ICU is usually based on clinical judgment using experience, clinical intuition, physiological parameters and scores to assess severity of illness and specific criteria, such as the need for mechanical ventilation or vasoactive medication [9–12]. However, the decision is also influenced by ICU capacity and pressure for beds as suggested by the fact that after-hours discharge is strongly associated with higher rates of readmission and mortality [7, 13]. To reduce the incidence of untimely ICU discharge, current guidelines suggest the use of checklists and explicit criteria for discharge, but marked heterogeneity in implementation and compliance exists [6, 14].

Real time decision support systems using prediction models that interact with electronic patient records at the bedside could be a promising solution to prevent untimely discharge from the ICU. Given the complexity of the decision and taking into account the vast amount of data gathered routinely in the ICU, machine learning is particularly suited for this task. Indeed, recent years have witnessed several attempts to develop these models [10, 15–27]. However, none of these models seem to have been specifically developed with the intention to directly implement them within the context of existing electronic health records to deliver real time decision support at the bedside. In fact, to the best of our knowledge, none of these models are currently in clinical use.

Therefore, our aim was to develop a machine learning model to predict readmission or death following discharge from the Intensive Care unit. We intended to greatly emphasize the need for the model to provide trends for its predictions and to provide the clinician insight in those predictions. In addition, it was imperative that the model used clinically relevant features that are readily available through interfaces with existing electronic health records allowing for implementation at the bedside. These considerations implied that it was of invaluable importance that the model was developed in close cooperation with experienced intensivists.

Methods

Anonymized data was extracted from the Patient Data Management System (MetaVision, iMDsoft, Tel Aviv, Israel) and administrative databases of Amsterdam UMC, location VUmc, Amsterdam, The Netherlands. VUmc is a 733-bed tertiary care hospital with a 24-bed mixed surgical-medical ICU. Data from patients older than 18 years and admitted to the ICU between 2004 and March 2016 were included

in the analysis. ICU admissions longer than 30 days were excluded, because these patients typically followed a discharge workflow that is more closely coordinated with the receiving ward.

Endpoints

The primary outcome was ICU readmission and/or death, both within 7 days of ICU discharge. This composite outcome was chosen because both are likely to influence the decision to discharge patients from the ICU and because they are competing risks. ICU readmission was defined as a transfer from the ICU to the general ward and back to the ICU or our Medium Care (high dependency) unit (MCU) during the same hospital stay. Palliative care patients and patients with do-not-resuscitate or do-not-intubate orders were excluded from the analysis. Patients transferred to other hospitals were also excluded.

Feature engineering

Patient demographics (e.g. age, sex), clinical observations (e.g. nursing scores, Glasgow Coma Scale score), automated physiological measurements from devices (e.g. patient monitor, ventilator, continuous renal replacement therapy), laboratory studies, medication (e.g. sedatives, vasopressors) and other support (e.g. enteral feeding, intermittent haemodialysis) were considered as input for the model. Extensive discussions in expert sessions with intensivists and visual analysis using heat maps on all available variables in the dataset were used to determine potential features. Demographics and characteristics of the admission (e.g. length of stay or time spent at the hospital before the ICU admission) were directly used as features. For variables that were measured or documented multiple times during the admission (e.g. all other input data), extensive pre-processing was performed in order to extract informative features on which to train the model. Since there is a large difference in sample frequency of the data, from data that is captured automatically every minute (devices) to scores that are performed only once a week if ever, we specifically chose aggregation functions that could capture trends, extremes and availability for all relevant features. The choice of feature aggregates was inspired by previous work [10, 18, 25] and we expanded the choice with additional aggregates that we suspected to be predictive (Additional file 1: Feature Engineering section with Table S1 and S2). Missing values were imputed using mean imputation. Additionally, we created features indicating whether or not a specific variable was measured and features indicating how often this variable was measured to allow informative missingness to be modelled.

Model development

After feature engineering, the total number of features was initially 5,466 per patient. Model development and analysis was performed using scikit-learn on Python [28]. Using scikit-learn Pipelines, feature selection, scaling and model training were performed sequentially with stratified 10-fold cross-validation. Feature selection was performed using Logistic Regression with an L1 penalty [29]. The regularization parameter was optimized using grid search. Only the features receiving a non-null coefficient were included in the final model. Features were scaled to a distribution with zero mean and standard deviations equal to 1 before training the models. The incidence of our combined endpoint is relatively low

compared to an uneventful (good) outcome, also known as an imbalanced dataset. We apply both traditional logistic regression and various machine learning algorithms to study which method leads to the most accurate results for this type of prediction problem. Algorithms were trained to predict the outcome using grid search to optimize hyperparameters (Additional file 1: Hyperparameter Tuning section and Table S3).

Performance

Model performance was measured using 10-fold stratified cross-validation. Since the primary outcome is an imbalanced class, we constructed both the Receiver Operating Characteristics (ROC) curve as well as the Precision-Recall (PR) curve, and calculated the areas under these curves (AUC) [30, 31]. To compare the performance of our model, we also trained an additional model for which the classification was based entirely on the Stability and Workload Index for Transfer (SWIFT) score, which has been designed to predict readmission after ICU discharge [9]. In addition, we evaluated the effect of feature engineering on model performance.

Feature importance

To determine feature importance, we used SHapley Additive exPlanations (SHAP) [32, 33]. SHAP determines for each patient individually the contribution of all features to that patient's prediction and thus can be used to interpret the model's prediction for an individual patient for complex machine learning algorithms.

Sensitivity analysis

A number of intensive care societies advocate using ICU readmission within 2 days of discharge as a quality indicator [34]. To evaluate robustness of our model design, we retrained the model on the composite outcome after 2 days and compared performance and feature importance with our model developed for 7 days.

Clinical relevance and impact

We used decision curve analysis to quantify the usefulness of our model based on the net benefit, defined as the difference between the true positives (actual readmissions/deaths) and the false positives (incorrectly identified patients, that could have been discharged), corrected by a factor determined by a threshold the clinician chooses to accept: the readmission probability (Additional file 1: Fig. S1) [35, 36].

To explore potential clinical impact, we first divided patients into (1) short stay patients and (2) long stay patients. Short stay patients (66% of the patients) were defined as patients with an ICU length of stay of less than two days, whereas long stay patients (34% of the patients) were patients with an ICU length of stay of two or more days. Using decision curve analysis, a clinically reasonable risk threshold was chosen to further divide the short stay patients into two subgroups: (a) high risk patients: patients with a combined risk of more than 6% of the primary outcome at the moment of discharge and (b) low risk patients: patients with a combined risk of less than 6% of the primary outcome at the moment of

discharge. For long stay patients, we assessed potential clinical impact based on analysis of readmission probability-time curves. These were obtained by using our prediction model to calculate the probability of the primary outcome for each day of ICU stay, thus describing its variation throughout the admission. It should be noted, however, since the model was trained on features of patients that were actually discharged, a low predicted probability doesn't necessarily imply that a patient is ready for discharge, since other conditions, such as the need for mechanical ventilation or vasoactive drugs may prevent this.

Finally, we explored the impact of our model on readmission rates and length of stay using two scenarios. In the first scenario we assume that the readmission rates of the groups that will be kept in the ICU longer to drop with 15% and in the second scenario to drop by 30%.

Results

After excluding patients that did not survive their first ICU admission and ICU admissions longer than 30 days, our data set included 14,105 admissions (Table 1). Most patients were ventilated and received vasoactive drugs during their admission: 86.3% and 68.2% respectively. The combined readmission/mortality rate within seven days of ICU discharge was 5.3%, while the readmission and mortality rates were 4.3% and 1.2%, respectively. As expected, patients that were readmitted or died within 7 days of ICU discharge were more often emergency patients (65 vs. 43%) and had a longer initial length of stay (6.4 vs 5.1 days). In addition, patients that died within 7 days of ICU discharge were older (63.2 vs 71.5 years).

Table 1

Study population. Patients are grouped by outcome events after ICU discharge: readmission and/or death within 7 days of discharge. SD Standard deviation. SOFA Sequential Organ Failure Assessment; score ranges from 0 to 24; higher ranges indicate greater severity of illness. MEWS Modified Early Warning Score; score ranges from 0 to 14; higher ranges indicate more abnormal physiological variables [42]; SWIFT Stability and Workload Index for Transfer; score ranges from 0 to 64; higher ranges indicate higher readmission risk [9]. Note: classification in Surgical/Medical admissions was only available for patients admitted in 2010 and later.

Characteristic	Total	No events	Readmission and/or death	Readmission only	Death only
ICU admissions, n (%)	14105 (100)	13354 (94.7)	751 (5.3)	610 (4.3)	173 (1.2)
Demographics					
Age, years, mean (SD)	63.3 (15.3)	63.2 (15.3)	65.0 (15.5)	63.5 (15.4)	71.5 (13.5)
Female, n (%)	4439 (32.1)	4180 (32.0)	259 (34.9)	206 (34.2)	71 (41.5)
Body Mass Index (BMI), kg/m ² , mean (SD)	26.3 (4.9)	26.4 (4.9)	25.1 (4.8)	25.1 (4.8)	24.7 (4.9)
Admission type					
Surgical, n (%)	3691 (66.7)	3529 (67.9)	162 (48.5)	150 (53.8)	22 (29.7)
Emergency admission, n (%)	6247 (44.3)	5761 (43.1)	486 (64.7)	368 (60.3)	139 (80.3)
Length of stay, days, mean (SD)	3.5 (5.2)	3.3 (5.1)	5.5 (6.4)	5.0 (6.1)	6.8 (7.1)
Supportive Care					
Received mechanical ventilation, n (%)	12172 (86.3)	11592 (86.8)	580 (77.2)	474 (77.7)	130 (75.1)
Received vasopressors/inotropes, n (%)	9623 (68.2)	9082 (68.0)	541 (72.0)	440 (72.1)	122 (70.5)
Risk Scores					
SOFA max, during admission, mean (SD)	4.4 (4.2)	4.3 (4.1)	5.7 (4.7)	5.6 (4.6)	6.2 (5.0)
MEWS at admission, mean (SD)	3.0 (1.9)	3.0 (1.9)	3.6 (2.1)	3.5 (2.1)	4.0 (2.2)

Characteristic	Total	No events	Readmission and/or death	Readmission only	Death only
SWIFT score at discharge, mean (SD)	6.4 (5.0)	6.4 (4.9)	5.9 (5.9)	5.9 (5.9)	5.6 (6.1)

Figure 1 shows model performance using different machine learning algorithms. The Gradient Boosting and the Support Vector Machine (SVM) were the highest performing algorithms. Logistic Regression showed good performance as well, suggesting that pre-processing and feature-engineering enabled the algorithm to be sensitive to non-linear effects and, thus, to perform similar to more complex algorithms. Using Gradient Boosting, our model achieved an area under the ROC curve of 0.802 (95% CI 0.789–0.816) and an area under the Precision-Recall (PR) curve of 0.198 (95% CI 0.185–0.211, baseline PR AUC 0.053). Performance did not improve significantly by combining the prediction of separate models for readmission and mortality (Additional file 1: Table S4). When validated on our cross-validation set, all algorithms performed better than the logistic regression model developed using the SWIFT score. The additional value of some feature engineering, such as squaring the features, was non-significant especially for the Gradient Boosting algorithm (logistic regression, AUC ROC without squared features: 0.782 vs with squared features: 0.79; Gradient Boosting without squared features: 0.800 vs with squared features: 0.802).

The distribution of the predicted outcomes for the Gradient Boosting model is shown in Fig. 2. The group with a readmission/mortality event has a much wider predicted probability distribution extending up to almost 0.7 (70%), while in the no event group predicted probabilities are mostly below 0.2 (20%), suggesting that the model is clearly capable of separating the groups. The overlap in the low probability range shows that still a number of patients falsely receive a low probability, even though they have been readmitted and/or died.

An overview of features for the Gradient Boosting model to predict ICU readmission/mortality within seven days of discharge is shown in Table 2. Figure 3 shows the distribution of SHAP values of the most predictive features in the Gradient Boosting model. Most features are well known parameters, such as type of patient (elective surgery), physiological variables (respiratory rate, mean arterial pressure, urine output, oxygen requirements) but also less apparent features such as the need for tube feeding have a significant effect on the predicted risk.

Table 2

Features used as input for the final Gradient Boosting model after manual selection, Logistic Regression with an L1 penalty and training using 10-fold cross-validation. The number of features includes aggregations of primary features.

Feature category	Examples	Number of features
General information		
Patient characteristics	age, length, weight on admission	4
Admission information	origin department, destination department	5
Laboratory results		
Blood gas analysis	pH, actual bicarbonate, partial CO ₂ pressure (P _a CO ₂), arterial oxygen saturation	10
Haematology	haemoglobin, white blood cell count, platelet count, activated partial thromboplastin time (APTT), prothrombin time (PT)	28
Routine chemistry	blood glucose, blood lactate, c-reactive protein (CRP), albumin, potassium, creatinine	36
Cardiac enzymes	creatinine kinase, creatinine kinase-muscle/brain (CK-MB), troponin-T, NT-proBNP	9
Liver tests	bilirubin, alanine aminotransferase (ALAT), aspartate aminotransferase (ASAT), alkaline phosphatase	14
Urinalysis	sodium, potassium	4
Vital signs and Device data		
Circulation	heart rate, mean arterial blood pressure, mean non-invasive blood pressure, pulse pressure, temperature, central venous pressure	43
Respiration	fraction of inspired oxygen (F _i O ₂), oxygen saturation, positive end-expiratory pressure (PEEP), tidal volume	40
Clinical observations and scores		
Neurology	Glasgow Coma Scale-score (GCS), Richmond Agitation-Sedation Scale (RASS), pupil response, Confusion Assessment Method for the ICU (CAM-ICU), Delirium Observation Screening (DOS)	24

Feature category	Examples	Number of features
Respiration	supplemental oxygen, bronchial suctioning, coughing reflex, P_aO_2/F_iO_2	12
Nephrology	urine output, fluid balance	3
Scores	Nursing activity score (NAS)	1
Diagnosics and Therapeutics		
Lines, drains and tubes	arterial line, renal replacement catheter, nasogastric feeding tube, urine catheter, endotracheal tube	7
Medication	anti-infectives, anti-bacterials, cardiovascular, feeding	8
Total		233

After retraining the model to predict ICU readmission and/or mortality within the first 2 days, analysis does show a significant drop in performance (ROC AUC 0.73), which is expected since our 7-day model post-discharge included relatively more patients that died (2 days mortality is 0.3% vs. 1.8% readmission, 7 days mortality 1.2% vs. 4.3% readmission), which has shown to be more predictable [25]. Interestingly, there is a substantial shift in predictive features for our 7-days model, compared to predicting the outcome after 2 days. For our 7-days model CK-MB (number of measurements per day), serum albumin (maximum value last 24 hours) and anti-infectives (time since treatment) are important predictors, but not for the 2-days model, where supplemental oxygen (last value during the last 24 hours) and white blood cell count (average last 24 hours) are most predictive.

Using decision curve analysis (Fig. 4), it's possible to identify the additional value of using the model compared to other strategies by showing the net benefit for a range of thresholds. The threshold is determined by how an intensivist weighs untimely discharging patients (false negatives) compared to unnecessarily keeping patients on the ICU (false positives). For example, using a threshold of 5%, we consider an untimely discharge 19 times more important than a unnecessarily prolonged stay. The discharge all line shows the strategy of discharging patients based on current practice, which has a net benefit of 0. The discharge none line shows the theoretical strategy of not discharging any patient from the ICU, which will only lead to a significant net benefit when using very low thresholds thereby leading to many unnecessarily and undesired prolonged stays. Using our prediction model (blue line), we can demonstrate that for clinically relevant thresholds (~ 3 to ~ 30%) net benefit is higher (green area) than the default strategies of discharge all or none.

For long stay (two or more days) patients, on average, predicted readmission probability decreased as patients get closer to ICU discharge. However, there was large variation between patients. Based on

decision curve analysis, we chose a threshold of 6% to identify high-risk patients, and used a 2% percentage point change in risk, the population average daily risk change, to identify different subgroups. Using these criteria, we identified five subgroups: 2a to 2e. Figure 5 shows readmission probabilities for those subgroups at different time points of ICU admission. Table S5 and the Impact Analysis section in Additional file 1 give detailed definitions for these groups, but in short, group 2a represents patients that have a relatively high-risk (> 6%) but were improving by at least 2% points; group 2b have a high risk but were improving by < 2%, group 2c are low-risk (< 6%) patients, group 2d patients recently improved towards a low-risk and group 2e patients are worsening (high-risk and increase in risk by > 2%). In our impact analysis, we suggest discharge strategies based on these groups, either postponing discharge, discharging as planned or even a day earlier. Using two possible scenarios, we show that integrating the readmission probability-time curve in a discharge workflow and changing management for high risk short stay patients (group 1b) and long stay groups 2a, 2c and 2e could lead to a decrease of up to 17% in readmission rate with an increase of only about 1.8% in average length of stay (Additional file 1: Impact Analysis section).

Discussion

In this paper we describe the development of a machine learning model for supporting ICU discharge with good performance using routinely collected ICU data. By using feature importance techniques and displaying risk prediction for readmission and mortality throughout the admission, application of our model as a bedside decision support tool seems feasible.

Several attempts to develop prediction models to prevent untimely discharge from the ICU for general adult intensive care patients have been made previously [10, 16–19, 21, 23–26, 37]. Earlier models use logistic regression for prediction focusing on very few parameters, whereas the newer models use more advanced machine learning algorithms (Additional file 1: Table S8). Our Gradient Boosting machine learning model represents an improvement over the models reported in the literature in terms of ROC AUC and outperforms the purposely built SWIFT score when validated on our own data set. The improvement in performance is modest and by choosing a time window of 7 days post-discharge more patients have been included that died after discharge, which is a more predictable task [25]. However, given the large and increasing number of ICU admissions worldwide, this modest reduction may have significant impact for patients and society.

Our paper has several strengths. Firstly, compared to the current literature, we performed more extensive feature engineering. Unsurprisingly, this allowed the logistic regression model to perform very similar to the Gradient Boosting model. Extending the feature engineering with for example log or inverse transformations and interactions between individual features would potentially further improve the performance of the logistic regression model towards the performance of the Gradient Boosting model.

Secondly, and perhaps more importantly, for most predictive models formal evaluation or bed side implementation is currently lacking [38]. Based on our previous experience, developing a bed side

decision support tool requires designing a model, pipeline and software with clinical implementation in mind [39, 40]. For predictive models, this involves close collaboration between intensivists and data scientists for extensive feature engineering with a focus on features that are available in real-time, innovative approaches with respect to interpretability, actionable insights and feature importance, as well as extensive performance evaluations and impact analyses.

Our paper also has some limitations. Firstly, performance was measured using internal cross-validation only and not on a separate test group. We specifically chose this approach because of the low incidence of adverse outcomes in our dataset and using a separate test set would further reduce the power during model development. Since many published models show moderate to good performance on data they were trained on, our next step to implement our model is to validate it on our current electronic health records and on data from other hospitals.

Secondly, in our Dutch setting, where ICU capacity is strained, we specifically chose to target readmissions and mortality until 7 days after ICU discharge to include patients that might suffer from complications that typically occur later, for example respiratory failure or sepsis and not the quality indicator of ICU readmission after 2 days. Since the performance of the retrained model outcome after 2 days dropped significantly, this does show that predicting early readmissions is a more difficult task.

Thirdly, predicting and possibly preventing readmissions may not influence outcome in some health care systems. In fact, even unexpected ICU readmissions may not unequivocally lead to an increase in hospital mortality as some authors have shown in a prospective study [41]. In addition, by only using routinely collected ICU data, some data that might be predictive of our endpoint, such as pre-hospital status or detailed reasons for (re)admission, could not be used to further improve performance.

For imbalanced datasets, the area under the ROC curve, which is often the only metric reported, may not be a good indicator of model performance. The reason for this is that the ROC curve is influenced by the large number of true negatives, which are clinically less relevant since those are patients that do well. False positive rate ($1 - \text{specificity}$), which forms the x-axis of the ROC curve, is calculated as $\text{false positives} / (\text{false positives} + \text{true negatives})$. In our case with a high number of true negatives, this will push the curve upwards and to the left, thereby increasing the area under the curve, which will give an overly optimistic view of model performance [30, 31]. The PR curve does not suffer from this limitation since precision, also known as the positive predictive value, is equivalent to $\text{true positives} / (\text{true positives} + \text{false positives})$ and therefore not influenced by the large number of negatives. In addition, for a perfect model, the area under the PR curve equals 1, similar to the ROC curve, but whereas the baseline AUC is fixed at 0.5 for the ROC curve, for the PR curve this baseline AUC equals the proportion of the 'positives' or in our case 0.053 (5.3% combined outcome of readmission and mortality). The PR curve (Fig. 1b) shows that the area under the curve (0.198) is much better than the baseline AUC of our combined outcome of 5.3% (0.053), but also that there is still room for improvement, even for the complex models we describe. Unfortunately, reporting of PR curves is rare in medicine [30] and they are also unavailable for previously developed models targeted at prevention of untimely ICU discharge. This makes rigorous comparisons

with our model cumbersome. Given the rapid progression of the field, it is likely that many more classification algorithms will be published for intensive care medicine. As imbalanced datasets are the norm in this setting, it is essential that both ROC and PR curves will be reported for future machine learning studies in the context of intensive care medicine.

Good model performance alone, is insufficient for a useful bedside tool on the ICU. Whether a model will be adopted in clinical practice will also depend heavily on ease of use as well as on the trade-off between the cost associated with a readmission (mortality, length of stay) and the cost of an unnecessary prolonged stay (length of stay, cancelled elective surgery or denied admissions). For the distinct groups of patients that we identified earlier using readmission probability-time curves, we showcase in our explorative impact analysis that using these curves may improve discharge management by preventing readmissions and deaths from premature ICU discharge with only a small increase in total length of stay. Furthermore, when a fraction of the patients that do not seem to improve during the last days of the admission (group 2b in Fig. 5) might even be discharged earlier as a consequence of using our model, the impact could potentially include both a reduction in total length of stay as well as a reduction in readmission rate. The promising result of our explorative analysis of clinical impact and the vast potential benefit for our future critically ill patients have prompted us to proceed with validation and implementation of our model at the bedside in Amsterdam UMC.

Conclusions

Our findings showed that the vast amount of data stored in ICU electronic health records can be used to develop a model that can accurately predict readmission and mortality after ICU discharge. By using readmission probability-time curves, our analysis showed that the model may decrease the readmission rate with a modest relative risk reduction of up to 17% while only minimally increasing the average length of stay.

Abbreviations

AUC
Area under the curve
CI
Confidence interval
ICU
Intensive Care unit
MEWS
Modified Early Warning Score
MCU
Medium Care (high dependency) unit
PR
Precision-recall

ROC
Receiver Operating Characteristic
SD
Standard deviation
SOFA
Sequential Organ Failure Assessment
SHAP
SHapley Additive exPlanations
SVM
Support Vector Machine
SWIFT
Stability and Workload Index for Transfer

Declarations

Ethics approval and consent to participate

The Medical Ethics Review Committee of VU University Medical Center reviewed the study protocol and considered it to be outside the scope of the Dutch Law on Scientific Research on Humans. In addition, they approved a waiver for the need for written informed consent for the use of data based on the very large number of included patients in accordance with the General Data Protection Regulation.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

Amsterdam UMC is entitled to royalties from Pacmed. HH is cofounder of Pacmed, a Dutch tech company for health care. PWGE is deputy chair of the Data Science section of the European Society of Intensive Care Medicine.

Funding

No funding was received for performing this study.

Authors' contributions

PJT and PWGE wrote the study protocol. RHD designed the data extraction process. PJT, MF, DPdB and HH analysed and interpreted the data. PJT, MF and DPdB drafted the manuscript. All authors read, commented, and approved the final manuscript.

Acknowledgements

The authors are grateful to the ICU and MCU nurses, residents, fellows and staff for their efforts in documenting and continuously improving the collected data in the patient data management system. The authors are also grateful to Jan Peppink for database administration and support of the patient data management system.

References

1. Oh H, Lee S, Kim J, Lee E, Min H, Cho O, et al. Clinical validity of a relocation stress scale for the families of patients transferred from intensive care units. *J Clin Nurs*. 2015;24:1805–14. doi:10.1111/jocn.12778.
2. Niven DJ, Bastos JF, Stelfox HT. Critical care transition programs and the risk of readmission or death after discharge from an ICU: A systematic review and meta-analysis. *Crit Care Med*. 2014;42:179–87.
3. Stelfox HT, Bastos J, Niven DJ, Bagshaw SM, Turin TC, Gao S. Critical care transition programs and the risk of readmission or death after discharge from ICU. *Intensive Care Med*. 2016;42:401–10. doi:10.1007/s00134-015-4173-7.
4. Chen LM, Martin CM, Keenan SP, Sibbald WJ. Patients readmitted to the intensive care unit during the same hospitalization: Clinical features and outcomes. *Crit Care Med*. 1998;26:1834–41. doi:10.1097/00003246-199811000-00025.
5. Alban RF, Nisim AA, Ho J, Nishi GK, Shabot MM. Readmission to surgical intensive care increases severity-adjusted patient mortality. *J Trauma - Inj Infect Crit Care*. 2006;60:1027–31. doi:10.1097/01.ta.0000218217.42861.b7.
6. Nates JL, Nunnally M, Kleinpell R, Blosser S, Goldner J, Birriel B, et al. ICU Admission, Discharge, and Triage Guidelines: A Framework to Enhance Clinical Operations, Development of Institutional Policies, and Further Research. *Crit Care Med*. 2016;44:1553–602. doi:10.1097/CCM.0000000000001856.
7. Vollam S, Dutton S, Lamb S, Petrinic T, Young JD, Watkinson P. Out-of-hours discharge from intensive care, in-hospital mortality and intensive care readmission rates: a systematic review and meta-analysis. *Intensive Care Med*. 2018;44:1115–29. doi:10.1007/s00134-018-5245-2.

8. Bagshaw SM, Wang X, Zygun DA, Zuege D, Dodek P, Garland A, et al. Association between strained capacity and mortality among patients admitted to intensive care: A path-analysis modeling strategy. *J Crit Care*. 2018;43:81–7. doi:10.1016/J.JCRC.2017.08.032.
9. Kareliusson F, De Geer L, Tibblin AO. Risk prediction of ICU readmission in a mixed surgical and medical population. *J Intensive Care*. 2015;3:30. doi:10.1186/s40560-015-0096-1.
10. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc*. 2018;15:846–53. doi:10.1513/AnnalsATS.201710-787OC.
11. Reini K, Oscarsson A. The prognostic value of the Modified Early Warning Score in critically ill patients: a prospective, observational study. *Eur J Anaesthesiol*. 2012;29:152–7. doi:10.1097/EJA.0b013e32835032d8.
12. Kramer AA, Higgins TL, Zimmerman JE. Can this patient be safely discharged from the ICU? *Intensive Care Med*. 2016;42:580–2. doi:10.1007/s00134-015-4148-8.
13. Bailey M, Duke G, Santamaria JD, Hart G, Renton J, Pilcher D V, et al. Factors associated with increased risk of readmission to intensive care in Australia. *Intensive Care Med*. 2011;37:1800–8.
14. Adviescommissie Kwaliteit van het Zorginstituut. Kwaliteitsstandaard Organisatie van Intensive Care. 2016;;1–18. <https://www.zorginzicht.nl/kwaliteitsinstrumenten/intensive-care-organisatie-kwaliteitsstandaard>. Accessed 1 Oct 2019.
15. Daly K, Beale R, Chang RW. Reduction in mortality after inappropriate early discharge from intensive care unit: logistic regression triage model. *BMJ*. 2001;322:1274–6. <http://www.ncbi.nlm.nih.gov/pubmed/11375229>.
16. Desautels T, Das R, Calvert J, Trivedi M, Summers C, Wales DJ, et al. Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach. *BMJ Open*. 2017;7:e017199. doi:10.1136/bmjopen-2017-017199.
17. Venugopalan J, Chanani N, Maher K, Wang MD. Combination of static and temporal data analysis to predict mortality and readmission in the intensive care. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. IEEE; 2017. p. 2570–3. doi:10.1109/EMBC.2017.8037382.
18. Fabes J, Seligman W, Barrett C, McKechnie S, Griffiths J. Does the implementation of a novel intensive care discharge risk score and nurse-led inpatient review tool improve outcome? A prospective cohort study in two intensive care units in the UK. *BMJ Open*. 2017;7:e018322. doi:10.1136/bmjopen-2017-018322.
19. Xue Y, Klabjan D, Luo Y. Predicting ICU readmission using grouped physiological and medication trends. *Artif Intell Med*. 2019;95:27–37. doi:10.1016/J.ARTMED.2018.08.004.
20. Fernandez R, Baigorri F, Navarro G, Artigas A. A modified McCabe score for stratification of patients after intensive care unit discharge: The Sabadell score. *Crit Care*. 2006;10. doi:10.1186/cc5136.
21. Gajic O, Malinchoc M, Comfere TB, Harris MR, Achouiti A, Yilmaz M, et al. The Stability and Workload Index for Transfer score predicts unplanned intensive care unit patient readmission: Initial

- development and validation. *Crit Care Med.* 2008;36:676–82. doi:10.1097/CCM.0B013E318164E3B0.
22. Fernandez R, Serrano JM, Umaran I, Abizanda R, Carrillo A, Lopez-Pueyo MJ, et al. Ward mortality after ICU discharge: A multicenter validation of the Sabadell score. *Intensive Care Med.* 2010;36:1196–201. doi:10.1007/s00134-010-1825-5.
 23. Frost SA, Tam V, Alexandrou E, Hunt L, Salamonson Y, Davidson PM, et al. Readmission to intensive care: development of a nomogram for individualising risk. *Crit Care Resusc.* 2010;12:83–9. <http://www.ncbi.nlm.nih.gov/pubmed/20513215>.
 24. Ouanes I, Schwebel C, Français Msc A, Bruel C, Soufir L, Adrie C, et al. A model to predict short-term death or readmission after intensive care unit discharge. *J Crit Care.* 2012;27:422.e1-422.e9.
 25. Badawi O, Breslow MJ. Readmissions and Death after ICU Discharge: Development and Validation of Two Predictive Models. *PLoS One.* 2012;7. doi:10.1371/journal.pone.0048758.
 26. Jo YS, Lee YJ, Park JS, Yoon H Il, Lee JH, Lee C-T, et al. Readmission to medical intensive care units: risk factors and prediction. *Yonsei Med J.* 2015;56:543–9. doi:10.3349/ymj.2015.56.2.543.
 27. Luo Y-F, Rumshisky A. Interpretable Topic Features for Post-ICU Mortality Prediction. *AMIA . Annu Symp proceedings AMIA Symp.* 2016;2016:827–36.
 28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2012;12:2825–30. <http://scikit-learn.sourceforge.net>. Accessed 11 Jul 2019.
 29. Talenti L, Luck M, Argy N, Houzé S, Yartseva A. L 1 Penalized logistic Regression Feature Preselection for Classification Tree: Application to the Diagnostic of Severe Imported Malaria Based on Heterogeneous Data. *JSM Math Stat.* 2015;2:1012.
 30. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One.* 2015;10:e0118432. doi:10.1371/journal.pone.0118432.
 31. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd international conference on Machine learning - ICML '06.* New York, New York, USA: ACM Press; 2006. p. 233–40. doi:10.1145/1143844.1143874.
 32. Lundberg SM, Erion GG, Lee S-I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv.* 2018;;preprint arXiv:1802.03888. <http://arxiv.org/abs/1802.03888>. Accessed 7 May 2019.
 33. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2018;2:749–60. doi:10.1038/s41551-018-0304-0.
 34. Maharaj R, Terblanche M, Vlachos S. The utility of ICU readmission as a quality indicator and the effect of selection. *Crit Care Med.* 2018;46:749–56.
 35. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26:565–74. doi:10.1177/0272989X06295361.

36. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol*. 2018;74:796–804. doi:10.1016/J.EURURO.2018.08.038.
37. Ng YH, Pilcher D V, Bailey M, Bain CA, MacManus C, Bucknall TK. Predicting medical emergency team calls, cardiac arrest calls and re-admission after intensive care discharge: creation of a tool to identify at-risk patients. *Anaesth Intensive Care*. 2018;46:88–96.
38. Hosein FS, Bobrovitz N, Berthelot S, Zygun D, Ghali WA, Stelfox HT. A systematic review of tools for predicting severe adverse events following patient discharge from intensive care units. *Crit Care*. 2013;17:R102. doi:10.1186/cc12747.
39. Weijs PJM, Stapel SN, De Groot SDW, Driessen RH, De Jong E, Girbes ARJ, et al. Optimal protein and energy nutrition decreases mortality in mechanically ventilated, critically ill patients: A prospective observational cohort study. *J Parenter Enter Nutr*. 2012;36:60–8. doi:10.1177/0148607111415109.
40. Elbers P, Girbes A, Malbrain MLNG, Bosman R. Right dose, right now: Using big data to optimize antibiotic dosing in the critically ill. *Anaesthesiol Intensive Ther*. 2015;47:457–63.
41. Santamaria JD, Duke GJ, Pilcher D V., Cooper DJ, Moran J, Bellomo R. Readmissions to Intensive Care. *Crit Care Med*. 2017;45:290–7. doi:10.1097/CCM.0000000000002066.
42. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *QJM*. 2001;94:521–6. doi:10.1093/qjmed/94.10.521.

Additional File

Additional file 1: Microsoft Word (.docx). Supplementary information: Feature engineering, Hyperparameter Tuning, Model performance, Decision curve analysis, Impact analysis, Published ICU Readmission models.

Figures

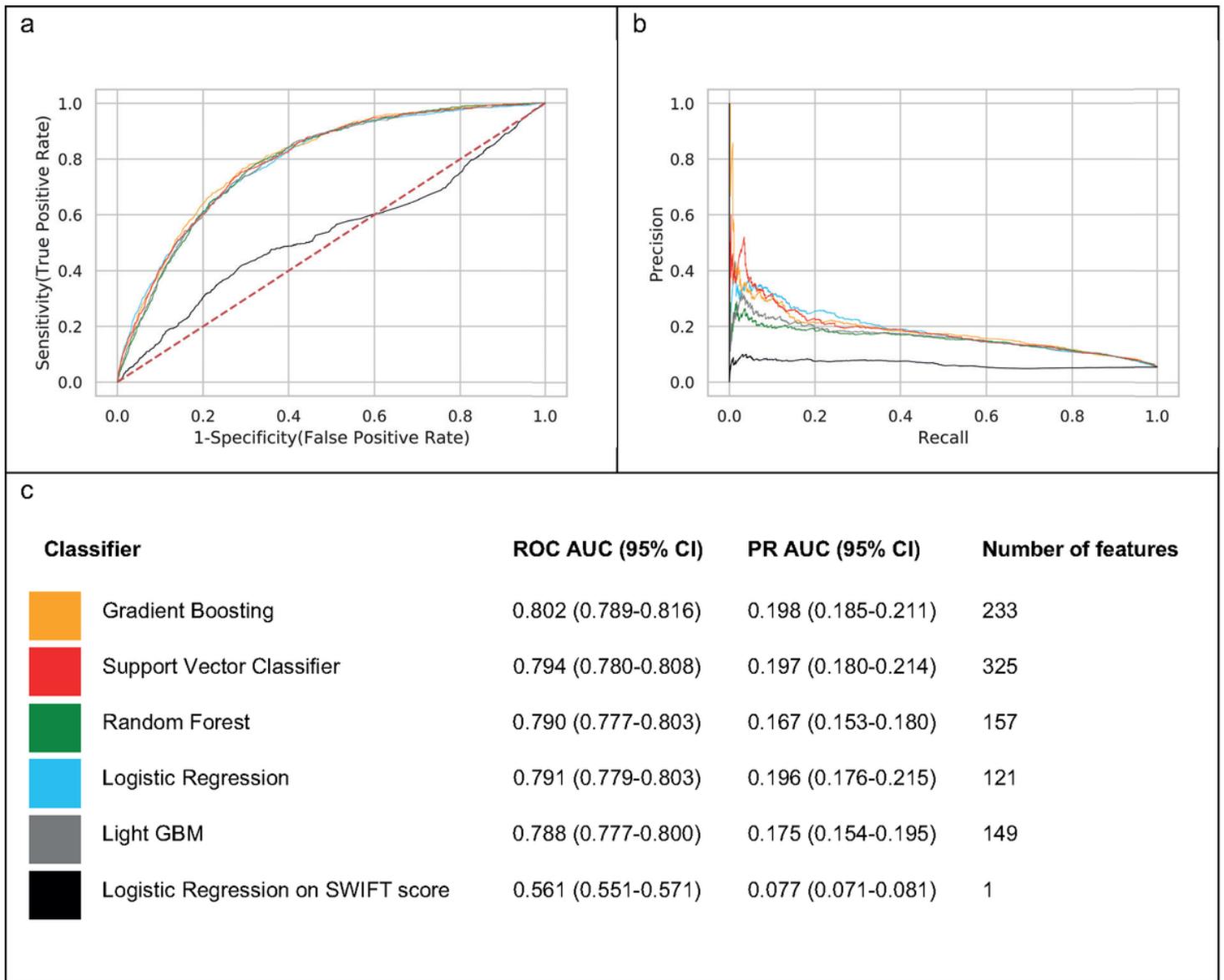


Figure 1

Model performance for different algorithms. The performance of the ICU readmission and mortality model is shown by using the Receiver Operating Characteristic curve (Fig. 1a) and Precision-Recall curve (Fig. 1b). The AUC of both curves and the number of features used in the model are shown in Fig. 1c. The SWIFT score is added for comparison. AUC area under the curve. PR Precision-Recall curve. CI Confidence interval. ROC Receiver Operating Characteristic

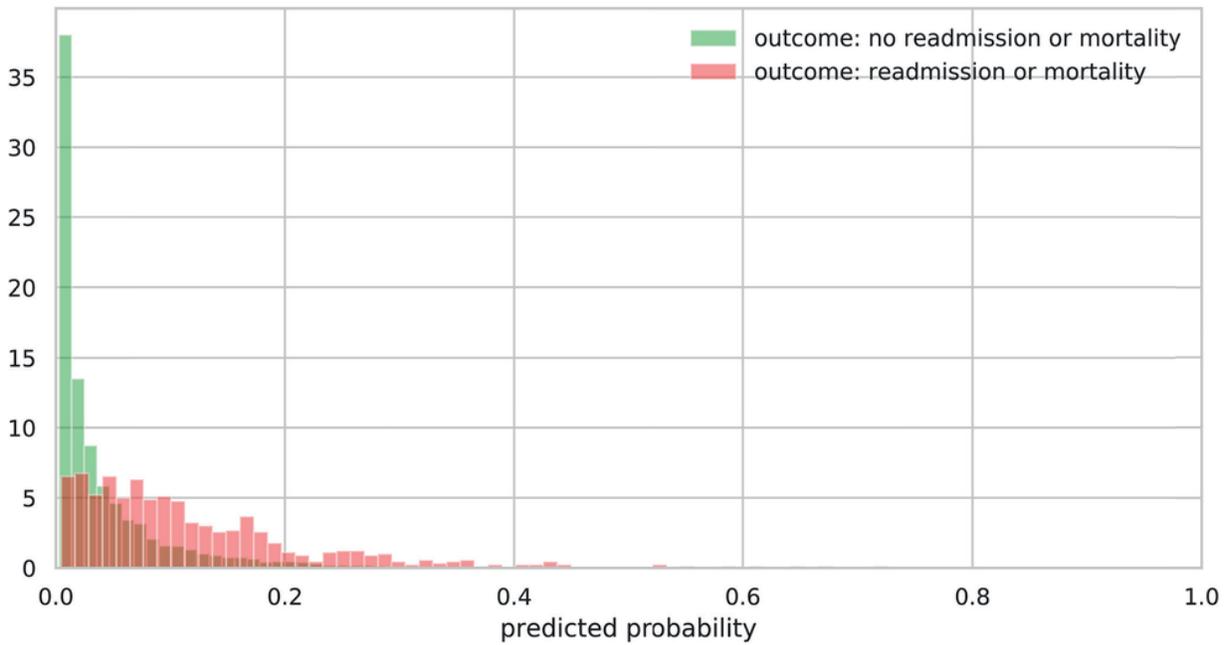


Figure 2

Distribution of the predicted combined readmission and mortality probability for patients in the cross-validation set. Patients that have been readmitted or died within seven days are plotted in red while the no event group in green. The two groups are normalized so that the area under each curve is equal to 1, since the number of patients in the readmission/mortality group is much smaller than the no event group. Patients in the readmission/mortality group receive higher probabilities (up to 0.7) than the no event group, but there is an overlap between the two groups.

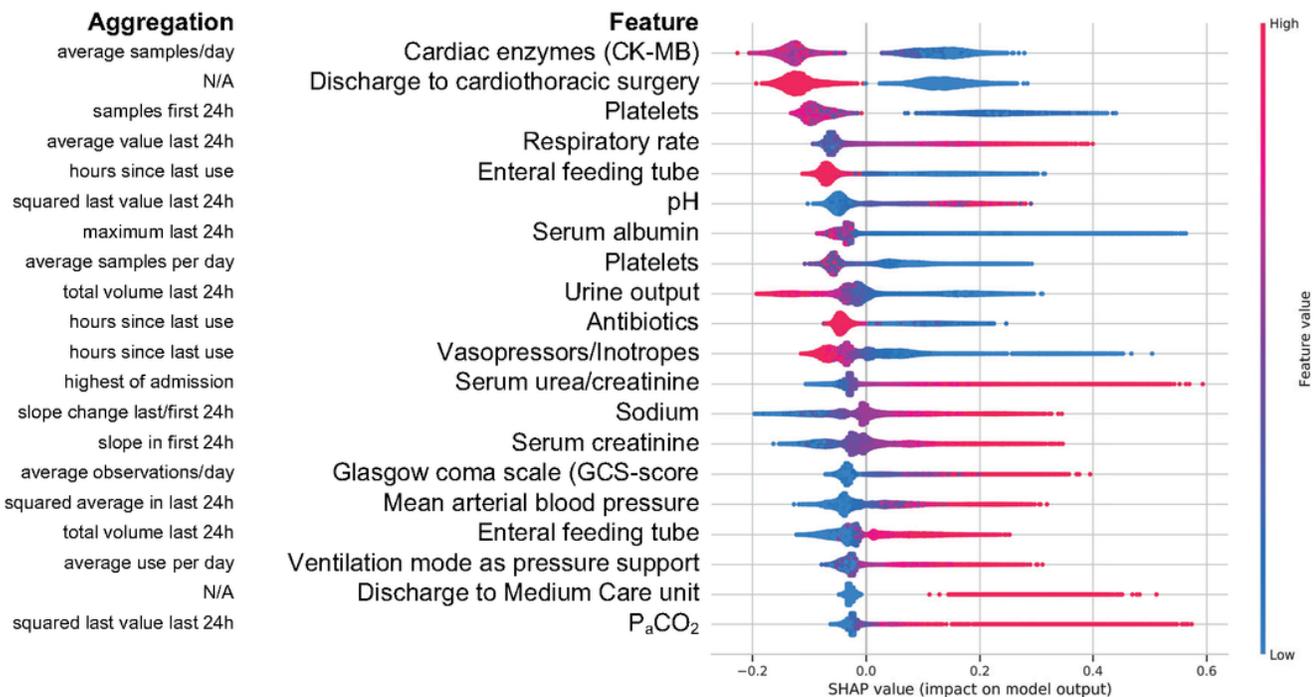


Figure 3

Feature importance. For the Gradient Boosting model feature performance is plotted for the first twenty features using SHAP. For each feature, each red or blue dot represents the impact of that feature on the prediction for that patient. Red dots represent patients with a high value for the specific feature, while blue dots represent patients with a low value for the specific feature. Red dots on to the right side of the distribution indicate that high values of the feature are associated with a high risk of readmission or mortality. Conversely, blue dots on the right side of the distribution indicate that low values of the feature are associated with a high risk.

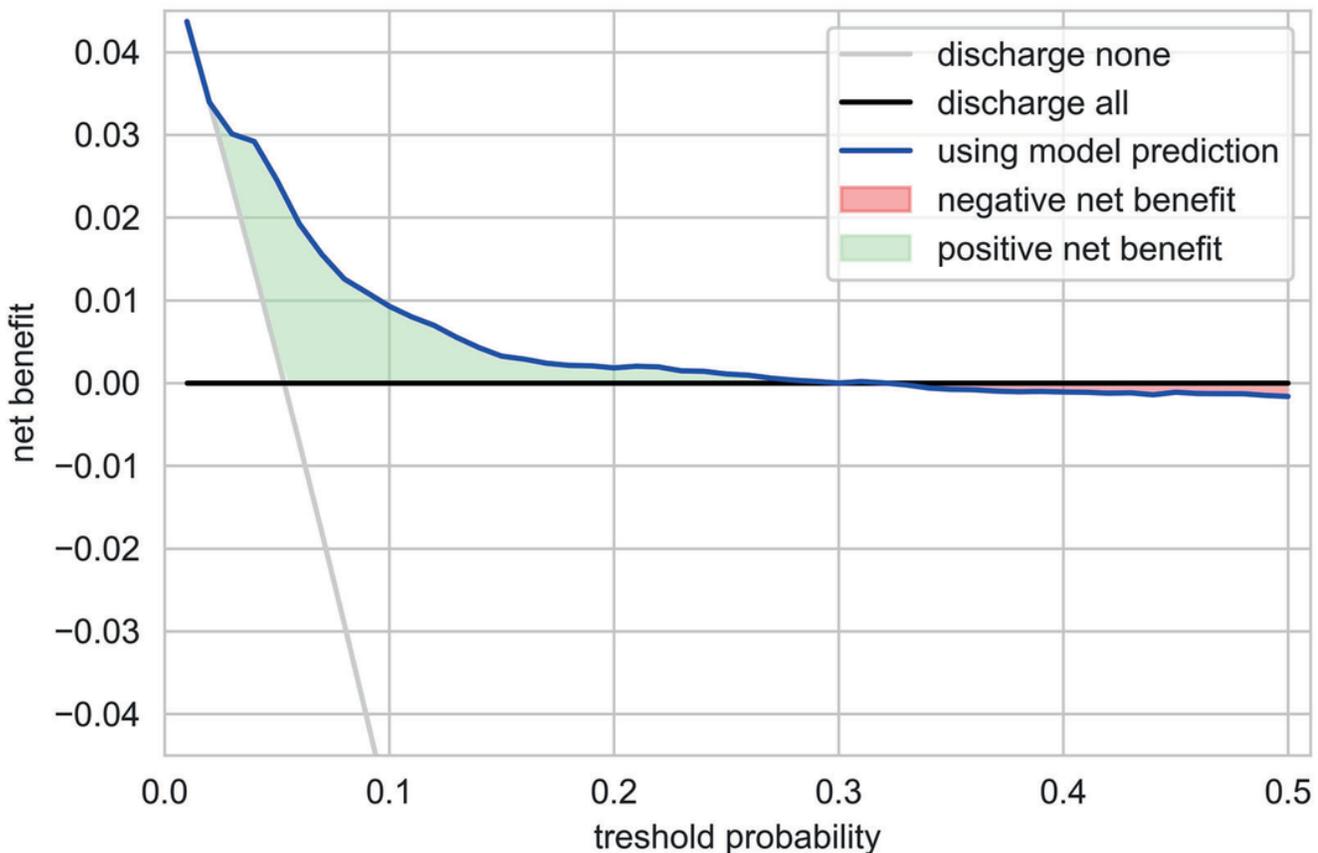


Figure 4

Decision curve analysis for our model showing net benefit for different scenarios. The discharge all line shows the strategy of discharging patients based on current practice, which has a net benefit of 0. The discharge none line shows the strategy of not discharging any patient from the ICU, which will only lead to a significant net benefit when using very low thresholds, leading to many unnecessarily prolonged stays. Using our prediction model (blue line), we can demonstrate that for clinically relevant thresholds (~3 to ~30%) net benefit is higher (green area) than the discharge all or none strategies.

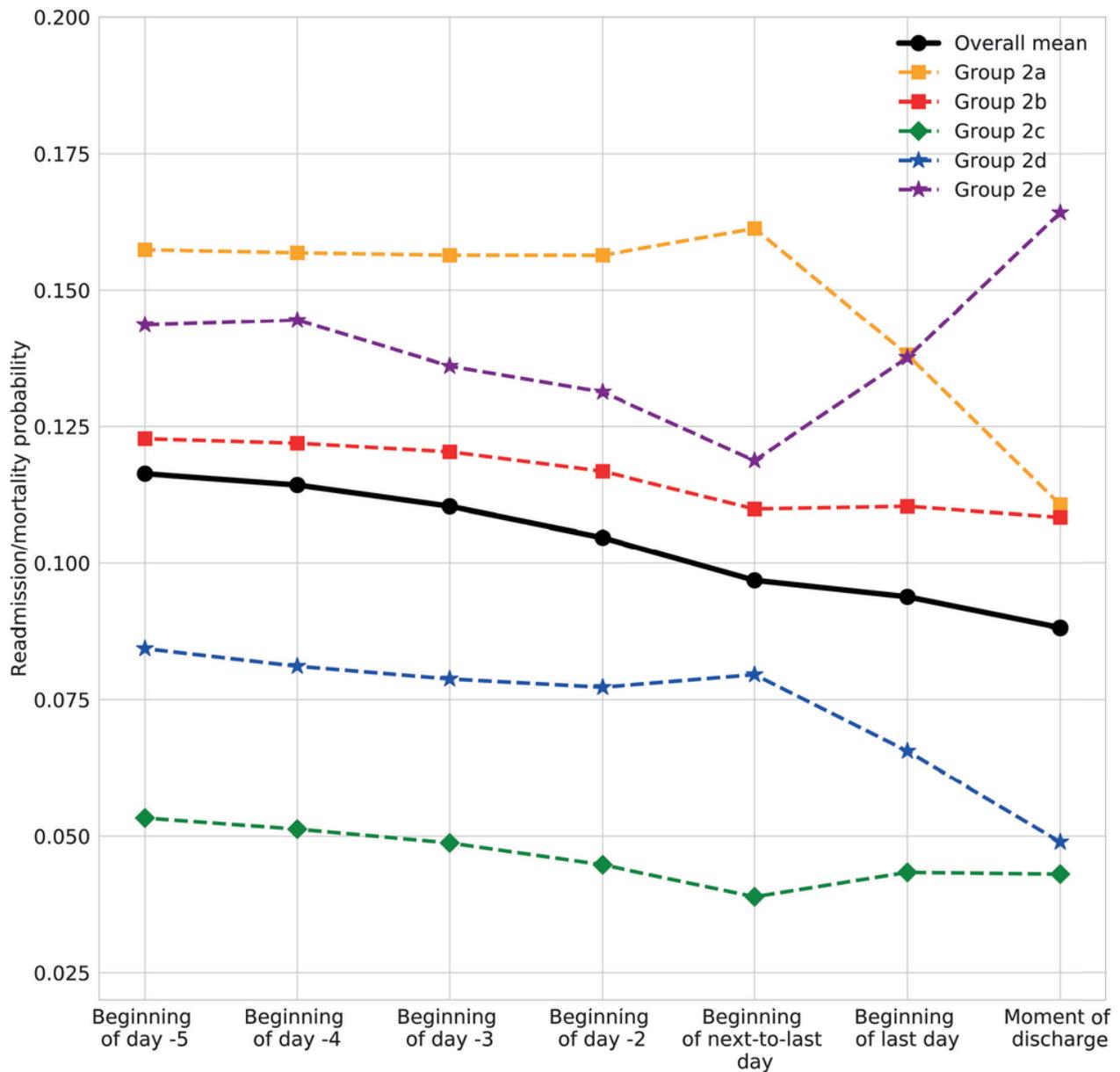


Figure 5

Mean readmission probability at different moments of ICU admission. The solid black line is the average among all patients with an ICU length of stay of at least 5 days. The coloured lines are the average for 5 different groups of patients. See the Impact Analysis section in Additional file 1 for the definition of the groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MLICUreadmissionAdditionalfile1.docx](#)