

Journal Pre-proof

Machine Learning for Predicting Long-Term Outcomes in Neurocritical Care: A sub-analysis from the SYNAPSE-ICU Study

Anna S. Scholze, MD, Laura Borgstedt, MD, Stefan J. Schaller, MD, Prof., Giuseppe Citerio, MD, Prof., Francesca Graziano, MSc, Volker Hüge, MD, Bernhard Ulm, MSc, Kristina Fuest, MD

PII: S1878-8750(26)00607-8

DOI: <https://doi.org/10.1016/j.wneu.2026.125391>

Reference: WNEU 125391

To appear in: *World Neurosurgery*

Received Date: 13 May 2026

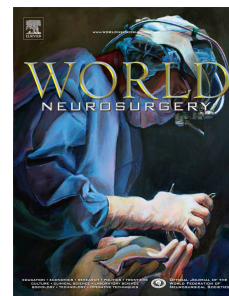
Revised Date: 24 September 2026

Accepted Date: 26 September 2026

Please cite this article as: Scholze AS, Borgstedt L, Schaller SJ, Citerio G, Graziano F, Hüge V, Ulm B, Fuest K, Machine Learning for Predicting Long-Term Outcomes in Neurocritical Care: A sub-analysis from the SYNAPSE-ICU Study, *World Neurosurgery* (2026), doi: <https://doi.org/10.1016/j.wneu.2026.125391>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 The Author(s). Published by Elsevier Inc.



Machine Learning for Predicting Long-Term Outcomes in Neurocritical Care:

A sub-analysis from the SYNAPSE-ICU Study

Anna S Scholze (MD)¹, Laura Borgstedt (MD)¹, Stefan J Schaller
(MD/Prof.)^{2,9}, Giuseppe Citerio (MD/Prof.)^{3,4}, Francesca Graziano
(MSc)^{5,6}, Volker Hüge (MD)⁸, Bernhard Ulm (MSc)^{1,7}, Kristina Fuest
(MD)¹

¹ TUM School of Medicine and Health, Department Clinical Medicine, Department of Anaesthesiology and Intensive Care Medicine, Munich, Germany

² Charité - Universitätsmedizin Berlin, Department of Anesthesiology and Operative Intensive Care Medicine (CVK, CCM), Berlin, Germany

³ School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

⁴ Neurointensive Care Unit, Neuroscience Department, Fondazione IRCCS San Gerardo Dei Tintori, Monza, Italy

⁵ Biostatistics and Clinical Epidemiology, Fondazione IRCCS San Gerardo Dei Tintori, Monza, Italy.

⁶ Bicocca Bioinformatics Biostatistics and Bioimaging Center B4, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.

⁷ University of Ulm, Faculty of Medicine, Department of Anaesthesiology and Intensive Care Medicine, Ulm, Germany

⁸ Department of Anesthesiology and Critical Care, Schoen Klinik, Bad Aibling, Germany

⁹ Medical University of Vienna, Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Division of General Anaesthesia and Intensive Care Medicine, Vienna, Austria

* Corresponding Author:

Kristina E. Fuest

Ismaninger Str. 22, 81675 Munich

Email: kristina.fuest@tum.de

Phone: +49 89 4140 8462

Keywords Neurocritical care - Machine learning - Outcome prediction

- Glasgow Outcome Scale Extended (1) - ICU management

Declaration of interests

KF, LB, VH and BU declare no competing interests. SJS received grants and non-financial support from Reactive Robotics GmbH (Munich, Germany), ASP GmbH (Attendorn, Germany), STIMIT AG (Biel, Switzerland), ESICM (Geneva, Switzerland), grants, personal fees, and non-financial support from Fresenius Kabi Deutschland GmbH (Bad Homburg, Germany), grants from the Innovationsfond of The Federal Joint Committee (G-BA), personal fees from Springer Verlag GmbH (Vienna, Austria) for educational purposes and Advanz Pharma GmbH (Bielefeld, Germany), non-financial support from national and international societies (and their congress organizers) in the field of anesthesiology and intensive care medicine, outside the submitted work. Dr. Schaller holds stocks in small amounts from Alphabet Inc., Bayer AG, and Siemens AG; these holdings have not affected any decisions regarding his research or this study. GC reports institutional research grants from Integra and NeurOptics and received personal fees as a speakers' bureau member and advisory board member from Integra, NeurOptics, Natus, all outside the submitted work. Giuseppe Citerio participated in the manuscript preparation during their personal involvement in the Italian Minister of University MUR Dipartimenti di Eccellenza 2023-2027 (l. 232/2016, art. 1, commi 314 - 337) and in PRIN 2022SYXEH.

Machine Learning as a Predictor of Long-Term Outcome

Prospective Cohort



SYNAPSE-ICU Study

Prospective, multinational cohort
3 neurocritical care entities (TBI / SAH / ICH)

n = 2,395 patients enrolled

n = 2,202 with 6-month outcome

Machine Learning Pipeline

Outcome: 6-month GOSE (favorable vs. unfavorable)

H2O AutoML (GLM / DRF / GBM / XGBoost / Stacked Ensembles)

**Training
(80%)**

**Test
(20%)**

Patient-level split, 4:1 ratio

Leakage-free split performed BEFORE multiple imputation (m = 5, pooled predictions across all imputations)

4 timepoints modelled:

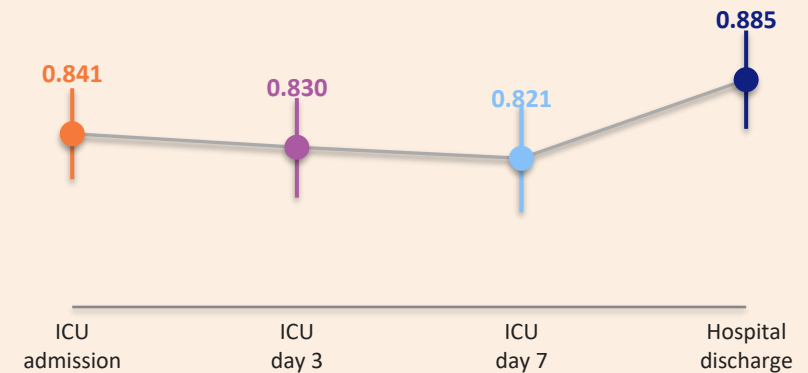
ICU admission → ICU day 3 → ICU day 7 → Hospital discharge

Predictive Performance

Top predictors (hospital discharge model)

1. GOSE at hospital discharge
2. Age
3. Discharge destination

Pooled AUROC (95% CI) by timepoint



Conclusion: Machine learning models predict 6-month functional outcome after neurocritical illness across the ICU admission-to-discharge trajectory (pooled AUROC 0.82-0.89, leakage-free, multiply-imputed pipeline). GOSE at discharge, age, and discharge destination are the strongest prognostic factors for the hospital-discharge model.

Machine Learning for Predicting Long-Term Outcomes in Neurocritical Care: A sub-analysis from the SYNAPSE-ICU Study

Anna S Scholze (MD)¹, Laura Borgstedt (MD)¹, Stefan J Schaller (MD/Prof.)^{2,9},
Giuseppe Citerio (MD/Prof.)^{3,4}, Francesca Graziano (MSc)^{5,6}, Volker Hude (MD)⁸,
Bernhard Ulm (MSc)^{1,7}, Kristina Fuest (MD)¹

¹ TUM School of Medicine and Health, Department Clinical Medicine, Department of Anaesthesiology and Intensive Care Medicine, Munich, Germany

² Charité - Universitaetsmedizin Berlin, Department of Anesthesiology and Operative Intensive Care Medicine (CVK, CCM), Berlin, Germany

³ School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

⁴ Neurointensive Care Unit, Neuroscience Department, Fondazione IRCCS San Gerardo Dei Tintori, Monza, Italy

⁵ Biostatistics and Clinical Epidemiology, Fondazione IRCCS San Gerardo Dei Tintori, Monza, Italy.

⁶ Bicocca Bioinformatics Biostatistics and Bioimaging Center B4, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.

⁷ University of Ulm, Faculty of Medicine, Department of Anaesthesiology and Intensive Care Medicine, Ulm, Germany

⁸ Department of Anesthesiology and Critical Care, Schoen Klinik, Bad Aibling, Germany

⁹ Medical University of Vienna, Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Division of General Anaesthesia and Intensive Care Medicine, Vienna, Austria

* Corresponding Author:

Kristina E. Fuest

Ismaninger Straße 22, 81675 München

Email: kristina.fuest@tum.de

Phone: +49 89 4140 8462

Abstract

Purpose Neurocritical conditions present significant challenges in the intensive care unit (ICU), especially during the initial hours. Accurate early risk assessment is essential for guiding acute care and influencing long-term outcomes. This study aims to evaluate the potential of machine learning (ML) models for risk prediction in neurocritical patients.

Methods: This retrospective sub-analysis of the SYNAPSE-ICU Study applied machine learning to predict the 6-month outcome for patients with three different neurocritical conditions: subarachnoid hemorrhage (SAH), traumatic brain injury (TBI), or intracerebral hemorrhage (ICH). Models were trained at ICU admission, ICU day 3, ICU day 7, and Hospital discharge using the H2O framework, including GLM, DRF, GBM, XGBoost, and Stacked Ensembles. Predictive performance was evaluated using AUPRC and AUROC metrics.

Results 2,202 patients were included with a complete data set to predict the 6-month outcome. At 6 months, 836 patients had a favorable outcome (Glasgow Outcome Scale Extended (GOSE) ≥ 5) and had a significantly higher GOSE at Hospital discharge ($p < 0.001$). Predictive performance (pooled across five multiply-imputed datasets) was highest at ICU admission (AUROC 0.841 [95% CI 0.804-0.878]) and Hospital discharge (AUROC 0.885 [0.845-0.925]), with a transient reduction during the ICU stay (Day 3: AUROC 0.830 [0.789-0.870]; Day 7: AUROC 0.821 [0.777-0.865]). Age was the most important parameter at the respective time points, except for Hospital discharge, where the GOSE at discharge had the highest impact on the 6-month outcome.

Conclusion This sub-analysis demonstrates that machine learning models can effectively predict long-term outcomes for neurocritical patients, with accuracy improving from ICU admission to Hospital discharge. GOSE at discharge and age are

critical parameters. The findings underscore the potential of ML models to improve prognostic assessments in neurocritical care.

Keywords Neurocritical care - Machine learning - Outcome prediction - Glasgow Outcome Scale Extended [1] - ICU management

Abbreviations

AUROC Area under the receiver operating characteristic curve

AUPRC Area under the precision-recall curve

CPP Cerebral Perfusion Pressure

CSF Cerebrospinal Fluid Drainage

DRF Distributed Random Forest

GBM Gradient Boosting Machine

GCS Glasgow Coma Scale

GLM Generalized Linear Model with regularization

GOSE Glasgow Outcome Scale Extended

ICH Intracerebral Hemorrhage

ICP Intracranial Pressure

ICPm Intracranial Pressure Monitoring

ICU Intensive Care Unit

IQR Interquartile Range

LOS Length of Stay

SAH Subarachnoid Hemorrhage

SHAP Shapley Additive Explanations Plots

TBI Traumatic Brain Injury

TIL Therapy Intensity Level

XGBoost eXtreme Gradient Boosting

1. Introduction

Neurocritical conditions such as SAH, TBI, or ICH present a complex challenge, especially in the critical first hours in the ICU. Accurate risk assessment of neurocritical patients is crucial for optimizing acute care and has a lasting impact on long-term outcomes [2]. Adequately assessing long-term outcomes is often difficult to determine, as prolonged rehabilitation periods are necessary for neurocritical illness [3, 4]. In the first six months after the critical illness, the most significant dynamics in terms of improvement can often be seen, but many patients remain in rehabilitation units beyond this period, and a transfer to a nursing home due to impaired functionality is common [5-7]. For this reason, identifying patients at an early stage is highly relevant to evaluating who may benefit from early neurorehabilitation [8]. It is equally pertinent to identify patients with an expected poor outcome to enable adequate counseling of relatives and to compare the outcome with the patient's presumed wishes [9]. The most suitable time point for predicting a reliable and effective outcome is currently unclear. This study aims to assess the ability to predict 6-month outcomes at various points along the hospital trajectory, beginning at ICU admission. Specifically, we investigate how the performance of risk prediction models evolves over time by identifying patient characteristics and clinical parameters at defined time points (ICU admission, ICU day 3, ICU day 7, and Hospital discharge) that are associated with a favorable long-term outcome, measured by the GOSE at 6 months, within the SYNAPSE study population [10].

In this context, the aim is to explore whether artificial intelligence—specifically machine learning algorithms—can develop risk prediction models based on routinely collected clinical data and patient trajectories during critical illness. Furthermore, we investigate whether such models can identify the most important individual predictors of long-term outcome at respective time points of the hospital trajectory. Improving risk

assessments and incorporating advanced predictive models using artificial intelligence can lead to better-informed decisions, ultimately enhancing patient outcome prediction [11].

This study hypothesizes that outcome prediction at an early time point is imprecise due to incomplete clinical and diagnostic information, and that prognostic accuracy improves continuously until hospital discharge.

2. Methods

This study is a retrospective sub-analysis of the Synapse-ICU Study. It was approved at the University of Milano-Bicocca by the ethics committee Brianza ASST-Monza on 21 November 2017, and was carried out according to the Declaration of Helsinki and the International Conference on Harmonization Good Clinical Practice guidelines. National or local approvals at study sites were obtained by national coordinators and local principal investigators, in accordance with local ethical regulations. This study was registered with ClinicalTrials.gov, NCT03257904.

2.1 The SYNAPSE-ICU Study

The SYNAPSE-ICU Study is a prospective, observational cohort study of 146 ICUs in 42 countries [10]. To be included in the study, participants at study sites had to be 18 years or older, with a diagnosis of TBI, ICH, or SAH. Additionally, patients had to present either with an altered level of consciousness (defined as a Glasgow Coma Scale [GCS] eye response score of 1 [no eye opening] and a GCS motor response score ≤ 5 [not obeying commands] on admission to ICU) or with neurological deterioration up to this level within 48 hours after ICU admission. The study aimed to evaluate the usage of intracranial pressure (ICP) monitoring and to compare outcomes between patients with and without it [10, 12]. Collected data included patient characteristics, clinical evaluation as per standard clinical practice, therapy intensity

level (TIL) with sedation management, neuroimaging, and 6-month outcomes. Primary endpoints were 6-month mortality and GOSE; secondary endpoints included hospital mortality and GOSE at ICU and hospital discharge [12].

2.2 Endpoints

The primary endpoint of our sub-analysis is the prediction of functional status after six months, as measured by the GOSE [13]. The 6-month GOSE was converted into a binary outcome. An unfavorable neurological outcome was defined as a GOSE score less than 5 (severe disability, vegetative status, and death); higher GOSE scores indicate a more favorable neurological outcome (good and moderate recovery) than lower scores [14, 15].

2.3 Statistical methods

All analyses were done using R (version 4.5.1) with R-Studio (Version 2025.05.0+496). Type I error was set at 0.05. Continuous variables were described as median with interquartile range (IQR), and categorical data were expressed as frequency (%). Comorbid conditions, neurological assessment on admission, type, and severity of the brain injury, and TIL during the first week were compared between the endpoint (favorable and unfavorable outcome) by the Mann-Whitney U test for continuous data and the χ^2 test for categorical data.

To adjust for the fact that different patient numbers were present at different observation times (due to discharge, transfer, or death), we decided to calculate four different models to avoid bias (named "*ICU admission*," "*ICU day 3*," "*ICU day 7*" and "*Hospital discharge*"). These models each include the patients present in the hospital at the respective time point.

Model development

For the prediction of the primary endpoint in the subsets of variables known at the time points "*ICU admission*," "*ICU day 3*," "*ICU day 7*," and "*Hospital discharge*" machine

learning algorithms were trained using the `h2o.automl()` function from the H2O framework [16]. This function automates the training and tuning process for various machine learning models within a set time limit, optimizing for a user-defined metric. It manages preprocessing, model selection, and hyperparameter tuning, ultimately generating a leaderboard of the top-performing models for easy comparison and deployment. The models used included a Generalized Linear Model with regularization (GLM), Distributed Random Forest (DRF), Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (XGBoost), a fully connected multilayer neural network (Deep Learning), and Stacked Ensembles [17]. The Stacked Ensemble algorithm in `h2o.automl` combines predictions from multiple base models to create a single, robust predictive model [18]. By leveraging the strengths of various individual models, it typically achieves better performance and accuracy than any single model alone, using a non-negative generalized linear model with Lasso regularization.

The data was split into a training and a test set using stratified (by endpoint) splitting with a 4:1 ratio [17]. Model training and tuning were conducted using stratified 5-fold cross-validation in the training set. Each training phase was set to run for 4 hours. Calculations were carried out on the Linux Cluster at the Leibniz Supercomputing Centre in Munich, utilizing 10 Intel Xeon CPU cores with 350 GB of RAM and a peak performance of 1400 TFlop/s.

For all prediction models, the area under the receiver operating characteristic curve (AUROC) with its 95% confidence interval (CI), and the area under the precision-recall curve (AUPRC) with its 95% CI for their identificatory performance were calculated as a quality measure for their discriminatory performance. An AUROC >0.9 is considered excellent, and an AUROC >0.8 is considered suitable for medical tests [19]. Model calibration was assessed using calibration plots, calibration intercept and slope, and the Brier score. The clinical utility of each time-point-specific prediction model was

evaluated using decision curve analysis across a range of threshold probabilities. Net benefit was calculated and compared with the default strategies of treating all or no patients as having an unfavorable outcome.

Models were ranked based on their mean AUPRC from the 5-fold cross-validation results. The AUPRC values with 95% CI for the test set are reported, along with the values of the AUROC with 95% CI for all univariate scores, the combination of scores, and the AutoML models.

Feature importance plots were used to determine which variables contributed most to the overall predictive performance of the model, while were used to explain how individual features influenced specific predictions and the model as a whole. For illustrative purposes, we plotted each individual prediction and connected the predictions over the 4 time points. Two patients were randomly selected, and their individual SHAP values were displayed as waterfall plots, highlighting how each predictor contributed to the final predicted risk in these specific cases over the study trajectory.

Sensitivity analysis

To separate the effects of changes in the study population and the availability of additional clinical information over time, we performed a sensitivity analysis at ICU day 3, ICU day 7, and hospital discharge. For each time point, the admission model, restricted to variables available at ICU admission, was applied to the corresponding survivor population to assess the population effect. The variable effect was then assessed by comparing this performance with that of a model incorporating all variables available at the respective time point. Additionally, five-fold cross-validation was performed at the country level, with no country represented in both the training and test set within the same fold.

Missing values

To deal with missing baseline data, we first coded missing values in categorical variables as ‘missing information’. Subsequently, multivariate imputation by chained equations [20] was applied to perform multiple imputation [21]. Specifically, the mice package was used with 5 imputed data sets ($m = 5$) and 10 iterations ($\text{maxit} = 10$). To prevent data leakage, patients were first split into training and test sets; imputation models were then fit using mice()'s `ignore=` argument, which restricts model fitting to the training rows while still producing imputed values for all rows (training and test). The imputation method was assigned automatically per variable type (mice::make.method()): predictive mean matching/normal regression for continuous variables, logistic regression for binary categorical variables, and multinomial logistic regression (polyreg) for unordered categorical variables with more than two levels. H2O AutoML was then trained separately on each of the five imputed training sets, and the resulting predicted probabilities on the matching imputed test sets were averaged (pooled) per patient across all five imputations; all reported performance metrics are based on these pooled predictions. For the model at “Hospital discharge” the following procedure was used to transform and prepare the dataset. For each patient, multiple measurements were available for specific variables at different time points, namely “ICU admission”, “ICU day 3”, and “ICU day 7”. To streamline the data structure and avoid redundancy, these separate measurements were combined into a single composite variable. The composite was created by applying a hierarchical rule. If a measurement was available for ICU day 7, it was selected first; if ICU day 7 was missing, the algorithm reverted to ICU day 3; and if both ICU day 7 and ICU day 3 were unavailable, it reverted to the value of ICU admission.

3. Results

The SYNAPSE-ICU study included 2,395 patients. A total of 193 patients were lost

during follow-up, leaving 2,202 patients in the data sets (Figure 1).

Patient characteristics

The cohort's median age was 56 years, with 35% of the participants being female. The admission diagnosis was TBI in 53%, SAH in 22% and ICH in 25%. Table 1 presents the main characteristics of the study population, stratified by GOSE after 6 months. Patients with an unfavorable outcome (GOSE <5) after 6 months were older than those with a favorable outcome (62 vs 46 years, $p < 0.001$) and had a lower GCS on admission (4 [IQR 3, 7] vs 6 [IQR 3, 8], $p < 0.001$). Those with an unfavorable outcome presented more often with a history of cardiovascular disease (53% vs. 27%, $p < 0.001$) and/or neurological disease (15% vs. 9%, $p < 0.001$) than those with a favorable outcome. Length of stay in the ICU as well as in hospital differed significantly between both groups, with longer duration in the group with favorable outcome (12 [IQR 6, 12] vs 10 [IQR 3, 20], $p < 0.001$ for the ICU stay and 25 [IQR 15, 43] vs 14 [IQR 5, 36], $p < 0.001$ for the hospital stay, respectively). 788 patients died during hospital stay (36%) (Figure 1 and Table 1). A complete list of the variables and standardized mean difference can be found in the Supplementary Material (Table S1 and S2).

Primary endpoint

The leading algorithms for the four respective time points were StackedEnsemble, GBM and XGBoost (Table S3 and S4, Figure S1). The complete leaderboards for the five imputed data sets for each time point can be provided upon request. Both AUROC and AUPRC values decreased from ICU admission through ICU day 3 and ICU day 7, followed by an increase at Hospital discharge. The AUROC at Hospital discharge exceeded the value observed at ICU admission (all Figure 2). The receiver operating characteristic (ROC) curves demonstrated sensitivity and specificity at the subsequent

time points, with the highest ROC curve at Hospital discharge. The precision-recall curves further illustrate the course of predictive precision and recall (Figure 2). Brier score varied between 0.132 and 0.174 for the four different time points (Table S5). Calibration curves and decision curve analysis are added in the Supplementary Material (Figure S2).

As shown in Figure 3, the most influential predictors at ICU admission were age, early neurological worsening, pupillary response, and initial neurological status (especially motor and verbal components of the GCS score). On ICU day 3 and ICU day 7, parameters known at ICU admission (age and admission diagnosis) continued to have the highest importance (Figure 3, A “ICU day 3” and “ICU day 7”). Additionally, variables reflecting neurological trajectory, such as GCS and therapy intensity level, gained more importance. At the time of Hospital discharge, the GOSE had the most importance for long-term outcome, followed by age, discharge destination, and GCS at last observation (Figure 3, A “Hospital discharge”). Model performance stratified by admission diagnosis revealed no statistical significant difference at all time points (Table S6 and Figure S3).

These findings were further supported by SHAP value analyses: increasing age consistently contributed negatively to the model output, indicating an association with poor outcome. In contrast, higher GCS values at any time point showed positive SHAP contributions, supporting their association with favorable outcomes (Figure 3, B). At the time of Hospital discharge, SHAP summary plots revealed a clear gradient in feature contributions across outcome categories: patients discharged in a vegetative state or with severe disability showed predominantly negative SHAP values, whereas those with moderate disability or good recovery exhibited consistently positive SHAP contributions to the model output (Figure 3, B “Hospital discharge”).

To further visualize the results in the context of the individual patients, Figure 4

illustrates the probability of an unfavorable outcome for each respective patient according to the ML models throughout the hospital stay. At the time of Hospital discharge, most of the purple-colored trajectories (patients with a favorable outcome after 6 months) have a low probability of an unfavorable outcome. In contrast, the red-coloured trajectories (patients with unfavorable outcome after 6 months) mostly have a high probability. Additionally, the trajectories for two sample patients from the cohort, who had a similar predicted risk upon ICU admission, are depicted (37.6% for the favorable and 37.7% for the unfavorable outcome, respectively). For one patient, the predicted probability steadily increased, with an unfavorable outcome (Figure 4, bold red line). In contrast, the second patient's risk decreased over time, ultimately leading to the prediction of a favorable outcome (Figure 4, bold purple line).

As shown in Figure 5, the most relevant variables for predicting the 6-month outcome differ between the highlighted trajectories of the two sample patients (Figure 4, bold red and purple lines). For the patient with a favorable long-term outcome, the age of 28 years is the most important factor (Figure 5, left column “ICU admission”, “ICU day3” and “Hospital discharge”), followed by variables reflecting neurological status (“no neuroworsening”, components of the GCS, and TIL). In contrast, for the patient with the unfavorable outcome, parameters of neurological status are the most important, leading to a vegetative state on Hospital discharge (Figure 5, right column). Here, the positive contribution of the age of 45 years is not capable of shifting the probability of an unfavorable outcome to a more favorable one (Figure 5, right column, “ICU Admission”, “ICU day3”, and “Hospital discharge”).

Sensitivity analysis

The sensitivity analysis demonstrated competing population and variable effects over time. Applying the admission model to the survivor population resulted in a decrease in AUROC from 0.841 to 0.809 at ICU day 3 (Δ AUROC -0.033) and to 0.796 at ICU

day 7 (ΔAUROC -0.045), while the addition of time-specific variables partially restored discrimination to 0.830 at ICU day 3 and 0.821 at ICU day 7, respectively. At hospital discharge, the population effect was more pronounced, reducing the AUROC to 0.731 (ΔAUROC -0.110); however, the incorporation of variables available at discharge substantially increased the AUROC to 0.885 (ΔAUROC +0.154) (Table S7 and Figure S4). Model performance was consistent across the country-level cross-validation folds. AUROC ranged from 0.793 to 0.847, with overlapping 95% CIs. The fold with the highest proportion of patients from low- and middle-income countries (28%) had an AUROC of 0.814 [95% CI 0.797-0.832] (Table S8 and S9).

Missing values

The Fraction of Missing Information (FMI) for all seven imputed variables across the four timepoints was between 0.001 and 0.004 (Table S1).

4. Discussion

This study used machine learning models to predict 6-month outcomes of neurocritical patients, based on findings from the large, multinational, prospective SYNAPSE-ICU Study. The results underscore the potential of ML in enhancing prognostic accuracy throughout the ICU stay. Predictive performance improved markedly from the “ICU admission” model to the “Hospital discharge” model, with a nearly complete discrimination between favorable and unfavorable outcome groups at discharge. Notably, the trajectory of clinical features during hospitalization superseded the predictive value of static admission parameters.

Risk prediction in neurocritical care has traditionally relied on established scores, such as the ICH score for intracerebral hemorrhage. Maljaars et al. demonstrated that the ICH score can overestimate mortality by up to 31% in young adults, depending on the severity of the disease (24). Early studies explored whether ML could surpass

conventional tools for risk prediction in neurocritical illness [22-25]. Subsequent work confirmed that ML models can indeed predict outcomes successfully, yet persistent challenges remain: outcomes are often restricted to short-term surrogates such as hospital mortality [26, 27] or functional outcomes at hospital discharge [24, 28-31]. Moreover, many models rely primarily on admission variables and static imaging [23, 25, 32, 33] rather than incorporating dynamic treatment parameters. Only one previous study, by Maldaner et al. in 2021, compared machine learning models at different time points during the hospital trajectory. They showed that the ‘late’ outcome model outperformed the ‘early’ one, although the analysis was limited to subarachnoid hemorrhage and did not include long-term follow-up [29]. These findings align with our results, as predictive precision improves when clinical parameters from the hospital trajectory are integrated [34]. In contrast, as we studied four time points, we observed a decline in precision after admission, which resolved by the time of hospital discharge. As a possible explanation, we investigated whether there was a difference between patients from the ICU admission model who died within the first 3 days and those who survived or were discharged. We were able to show that the two groups differed significantly in almost all parameters (Table S10). This suggests that the high precision of the ICU admission model may be attributed to the predictability of early deaths. This interpretation was further supported by a sensitivity analysis demonstrating that the decline in model performance at ICU day 3 and day 7 was primarily attributable to changes in the population at risk, whereas the additional clinical information available at later time points only partially compensated for this effect. Other large and multicenter studies focused solely on single disease entities and short-term outcomes, such as Moriya et al. and Ikawa et al. on SAH [28, 30], Ballesterio et al. on TBI [24], and Yap et al. on ICH [26, 27]. A secondary analysis of the multicenter CENTER-TBI dataset extended this approach by predicting 1-year employment status—a patient-

centered functional outcome—with 81% accuracy using in-hospital data and 88% accuracy when 3-month follow-ups were included [35]. These data confirm our hypothesis that incorporating information along the clinical pathway improves long-term outcome prediction.

Our analysis addresses several of these shortcomings by integrating all three major neurovascular conditions—TBI, SAH, and ICH—within a large multicenter, international cohort. The SYNAPSE-ICU dataset captures the entire ICU trajectory, including information on therapeutic interventions such as intracranial pressure management across multiple days. This approach reflects the evolving course of neurocritically ill patients.

In summary, machine-learning models can enhance the estimation of short- and long-term outcomes in patients with neurocritical illnesses. Yet, as McLernon et al. demonstrated in a qualitative analysis of neurocritical care for spontaneous intracerebral hemorrhage, decisions on treatment intensity are shaped not only by prognostic information but also by contextual factors such as family preferences and resource availability. Consequently, there is a need for more individual predictive tools to support clinicians in decision-making. Numerous studies used Shapley values to explore the contribution of individual parameters in neurocritical care [23, 25, 30, 33, 36], although only a few investigated predictions on an individual patient level. While Huang et al. also illustrated the interpretable predictions for an exemplary patient but limited their analysis to in-hospital mortality (up to day 28) using the extensive MIMIC-IV database for stroke patients [37]. Zeng et al. and Shu et al. presented individual long-term outcome predictions (90 days and 12 months, respectively) but limited their work to a single disease entity and a single-center setting [38, 39]. To our knowledge, our study uniquely combines several strengths, in particular multicentricity, multiple disease entities, and long-term functional outcome. Furthermore, the development of

prediction performance over the course of hospitalization with different models using additional clinical data from the respective time points has rarely been investigated to date. Despite the rather widespread approach to risk prediction for individual patients, our study is the only one to examine the risk prediction for representative patients at several time points over the hospital trajectory.

Strengths and limitations

This sub-analysis has several strengths and limitations: The study's strengths lie in its robust dataset from diverse ICU settings globally, allowing for generalizable insights into neurocritical care practices. We opted for a dichotomous outcome classification (favorable vs unfavorable), which reduces granularity and limits the ability to capture more nuanced functional outcomes. However, we considered this binary approach more meaningful for an initial assessment, particularly in the context of communicating prognosis to relatives (e.g., dependent vs independent functional status).

Limitations include potential variations in clinical practices and data collection across different sites, which may influence model performance and external validation. Besides that, this broad approach has the limitation that details of the clinical stay, such as secondary infections or organ failure outside the brain, are not reflected. In addition, the reasons for death (which might be due to withholding or withdrawal of therapy, especially in the first seven days) or poor outcome are not known in detail. Although the 6-month GOSE provides a validated long-term measure, details of ongoing rehabilitation remain unknown. Future models could benefit from additional clinical variables, such as more detailed information on injury characteristics and disease severity, as well as information on the patient's clinical course after hospital discharge, particularly during rehabilitation, which may improve prediction of 6-month outcome. Furthermore, 12-month outcomes might better reflect the rehabilitative trajectory. Finally, while the SYNAPSE-ICU dataset enables international, multicenter analysis,

its predefined structure limits the addition of further data points. Similarly, socioeconomic status, access to rehabilitation services, and cultural or regional factors affecting care decisions were not captured in the SYNAPSE-ICU dataset and could not be included as predictors or examined as potential confounders of the observed associations. As with any outcome-prediction model in neurocritical care, we cannot rule out the possibility that early prognostic pessimism may have influenced subsequent decisions to withhold or withdraw life-sustaining therapies. This could create a self-fulfilling prophecy. Consequently, such treatment decisions may contribute to the perceived accuracy of early severity-based predictions. This represents an inherent limitation of observational outcome-prediction studies in neurocritical care and is not specific to our modeling approach. However, it may still affect model performance and its external validation.

5. Conclusion

The study demonstrates that machine learning models at different timepoints can effectively predict long-term outcomes for neurocritical patients, with predictive accuracy improving from ICU admission to discharge. These findings support the feasibility of ML-based prediction of long-term outcomes at different neurocritical care time points. However, prospective studies are needed to determine their clinical utility. The findings underscore the potential of ML models to enhance individual patient management by improving prognostic assessments in neurocritical care.

Declarations

Ethics Approval

The study was approved at the University of Milano-Bicocca by the ethics committee Brianza ASST-Monza on 21 November 2017 and was carried out according to the Declaration of Helsinki and the International Conference on Harmonization Good Clinical Practice guidelines. National or local approvals at study sites were obtained by national coordinators and local principal investigators, according to local regulations.

Consent for publication

Not applicable.

Availability of data and material

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

Competing interests

KF, LB, VH and BU declare no competing interests. SJS received grants and non-financial support from Reactive Robotics GmbH (Munich, Germany), ASP GmbH (Attendorn, Germany), STIMIT AG (Biel, Switzerland), ESICM (Geneva, Switzerland), grants, personal fees, and non-financial support from Fresenius Kabi Deutschland GmbH (Bad Homburg, Germany), grants from the Innovationsfond of The Federal Joint Committee (G-BA), personal fees from Springer Verlag GmbH (Vienna, Austria) for educational purposes and Advanz Pharma GmbH (Bielefeld, Germany), non-financial support from national and international societies (and their

congress organizers) in the field of anesthesiology and intensive care medicine, outside the submitted work. Dr. Schaller holds stocks in small amounts from Alphabet Inc., Bayer AG, and Siemens AG; these holdings have not affected any decisions regarding his research or this study. GC reports institutional research grants from Integra and NeurOptics and received personal fees as a speakers' bureau member and advisory board member from Integra, NeurOptics, Natus, all outside the submitted work. Giuseppe Citerio participated in the manuscript preparation during their personal involvement in the Italian Minister of University MUR Dipartimenti di Eccellenza 2023-2027 (l. 232/2016, art. 1, commi 314 - 337) and in PRIN 2022SYXEH.

Funding

None.

Authors' contributions

KEF and LB are the principal investigators. GC and FG were responsible for the ethical approval and developed the study protocol. BU and FG are the study statisticians. GC was instrumental in developing the methodology. KF, LB and SJS were involved in the analysis and all authors in the interpretation of the data. KF and LB drafted the manuscript. All authors were substantial in revising the manuscript and approved its final version.

Acknowledgements

We thank the SYNAPSE-ICU team in Italy and especially Giuseppe Citerio for providing the data set and for helping to develop the sub-analysis.

Use of Generative AI and AI-assisted technologies

Generative AI and AI-assisted technologies were NOT used in the preparation of this work.

Journal Pre-proof

References

1. Bongiovanni D, Schreiner N, Gosetti R, Mayer K, Angiolillo DJ, Sibbing D, et al. Immature Platelet Fraction Predicts Adverse Events in Patients With Acute Coronary Syndrome: the ISAR-REACT 5 Reticulated Platelet Substudy. *Arterioscler Thromb Vasc Biol* 2023;43:e83–e93.
2. Abulhasan YB, Teitelbaum J, Al-Ramadhani K, Morrison KT, Angle MR. Functional Outcomes and Mortality in Patients With Intracerebral Hemorrhage After Intensive Medical and Surgical Support. *Neurology* 2023;100:e1985–e95.
3. Shah VA, Thompson RE, Yenokyan G, Acosta JN, Avadhani R, Dlugash R, et al. One-Year Outcome Trajectories and Factors Associated with Functional Recovery Among Survivors of Intracerebral and Intraventricular Hemorrhage With Initial Severe Disability. *JAMA Neurol* 2022;79:856–68.
4. Boltzmann M, Schmidt SB, Gutenbrunner C, Krauss JK, Höglinger GU, Rollnik JD. One-year outcome of brain injured patients undergoing early neurological rehabilitation: a prospective observational study. *BMC Neurol* 2022;22:30.
5. Bartolo M, Bargesles S, Castioni CA, Bonaiuti D, Antenucci R, Benedetti A, et al. Early rehabilitation for severe acquired brain injury in intensive care unit: multicenter observational study. *Eur J Phys Rehabil Med* 2016;52:90–100.
6. Einhüpl B, Götze D, Reichl S, Willacker L, Pletz R, Kohlmann T, et al. Long-term outcomes of community-based intensive care treatment following neurological early rehabilitation- results of a multicentric German study. *Neurol Res Pract* 2025;7:35.
7. Liu N, Cadilhac DA, Andrew NE, Zeng L, Li Z, Li J, et al. Randomized controlled trial of early rehabilitation after intracerebral hemorrhage stroke: difference in outcomes within 6 months of stroke. *Stroke* 2014;45:3502–7.
8. Cinotti R, Putegnât JB, Lakhal K, Desal H, Chenet A, Buffenoir K, et al. Evolution of neurological recovery during the first year after subarachnoid haemorrhage in a French university centre. *Anaesth Crit Care Pain Med* 2019;38:251–7.
9. Leitgeb J, Mauritz W, Brazinova A, Majdan M, Janciak I, Wilbacher I, et al. Glasgow Coma Scale score at intensive care unit discharge predicts the 1-year outcome of patients with severe traumatic brain injury. *Eur J Trauma Emerg Surg* 2013;39:285–92.
10. Robba C, Graziano F, Rebora P, Elli F, Giussani C, Oddo M, et al. Intracranial pressure monitoring in patients with acute brain injury in the intensive care unit (SYNAPSE-ICU): an international, prospective observational cohort study. *The Lancet Neurology* 2021;20:548–58.
11. Uparela-Reyes MJ, Villegas-Trujillo LM, Cespedes J, Velásquez-Vera M, Rubiano AM. Usefulness of Artificial Intelligence in Traumatic Brain Injury: A Bibliometric Analysis and Mini-review. *World Neurosurg* 2024;188:83–92.
12. Citerio G, Prisco L, Oddo M, Meyfroidt G, Helbok R, Stocchetti N, et al. International prospective observational study on intracranial pressure in intensive care (ICU): the SYNAPSE-ICU study protocol. *BMJ Open* 2019;9:e026552.
13. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. *J Neurotrauma* 1998;15:573–85.
14. Yuh EL, Jain S, Sun X, Pisica D, Harris MH, Taylor SR, et al. Pathological Computed Tomography Features Associated With Adverse Outcomes After Mild Traumatic Brain Injury: A TRACK-TBI Study With External Validation in CENTER-TBI. *JAMA Neurol* 2021;78:1137–48.
15. Lenell S, Nyholm L, Lewén A, Enblad P. Clinical outcome and prognostic factors in elderly traumatic brain injury patients receiving neurointensive care. *Acta Neurochir (Wien)* 2019;161:1243–54.

16. H2O.ai. H2O AutoML: Automatic Machine Learning [15 Dec 2023].
17. Poirier ELaS. H2O AutoML: Scalable Automatic Machine Learning. 7th ICML Workshop on Automated Machine Learning (AutoML)2020.
18. Kagerbauer SM, Ulm B, Podtschaske AH, Andonov DI, Blobner M, Jungwirth B, et al. Susceptibility of AutoML mortality prediction algorithms to model drift caused by the COVID pandemic. *BMC Med Inform Decis Mak* 2024;24:34.
19. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*: John Wiley & Sons; 2013.
20. de Beer Jde V, Winemaker MJ, Donnelly GA, Miceli PC, Reiz JL, Harsanyi Z, et al. Efficacy and safety of controlled-release oxycodone and standard therapies for postoperative pain after knee or hip replacement. *Can J Surg* 2005;48:277–83.
21. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011;45:1 – 67.
22. Lim MJR, Quek RHC, Ng KJ, Loh NW, Lwin S, Teo K, et al. Machine Learning Models Prognosticate Functional Outcomes Better than Clinical Scores in Spontaneous Intracerebral Haemorrhage. *J Stroke Cerebrovasc Dis* 2022;31:106234.
23. Arefan D, Pease M, Eagle SR, Okonkwo DO, Wu S. Comparison of machine learning models to predict long-term outcomes after severe traumatic brain injury. *Neurosurg Focus* 2023;54:E14.
24. Ballesterio M, de Souza LC, Levada ALM, Pongeluppi RI, Funo SN, Pineda FG, et al. Is artificial intelligence superior to traditional regression methods in predicting prognosis of adult traumatic brain injury? *Neurosurg Rev* 2025;48:355.
25. Dengler NF, Madai VI, Unteroberdörster M, Zihni E, Brune SC, Hilbert A, et al. Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores. *Neurosurg Rev* 2021;44:2837–46.
26. Mao B, Ling L, Pan Y, Zhang R, Zheng W, Shen Y, et al. Machine learning for the prediction of in-hospital mortality in patients with spontaneous intracerebral hemorrhage in intensive care unit. *Sci Rep* 2024;14:14195.
27. Yap XV, Tu KC, Chen NC, Wang CC, Chen CJ, Liu CF, et al. Developing a high-performance AI model for spontaneous intracerebral hemorrhage mortality prediction using machine learning in ICU settings. *BMC Med Inform Decis Mak* 2025;25:149.
28. Ikawa F, Ichihara N, Horie N, Shiokawa Y, Nakatomi H, Ohkuma H, et al. Machine learning validation of a simple prediction model for the correlation between advanced age and clinical outcomes in patients with aneurysmal subarachnoid hemorrhage. *Neurosurg Rev* 2025;48:44.
29. Maldaner N, Zeiltberger AM, Sosnova M, Goldberg J, Fung C, Bervini D, et al. Development of a Complication- and Treatment-Aware Prediction Model for Favorable Functional Outcome in Aneurysmal Subarachnoid Hemorrhage Based on Machine Learning. *Neurosurgery* 2021;88:E150–e7.
30. Moriya M, Karako K, Miyazaki S, Minakata S, Satoh S, Abe Y, et al. Interpretable machine learning model for outcome prediction in patients with aneurysmatic subarachnoid hemorrhage. *Crit Care* 2025;29:36.
31. Mohammadzadeh I, Niroomand B, Shahnazian Z, Ghanbarnia R, Nouri Z, Tajerian A, et al. Machine learning for predicting poor outcomes in aneurysmal subarachnoid hemorrhage: A systematic review and meta-analysis involving 8445 participants. *Clin Neurol Neurosurg* 2025;249:108668.
32. Savarraj JPJ, Hergenroeder GW, Zhu L, Chang T, Park S, Megjhani M, et al. Machine Learning to Predict Delayed Cerebral Ischemia and Outcomes in Subarachnoid Hemorrhage. *Neurology* 2021;96:e553–e62.
33. Zhou W, Liao X, Shi H, Wang M, Xiao Y, Yu X, et al. Prediction of the 180 day functional outcomes in aneurysmal subarachnoid hemorrhage using an optimized XGBoost

model. *Sci Rep* 2025;15:20833.

34. Arkes HR, Aberegg SK, Arpin KA. Analysis of Physicians' Probability Estimates of a Medical Outcome Based on a Sequence of Events. *JAMA network open* 2022;5:e2218804–e.
35. Van Deynse H, Cools W, De Deken VJ, Depreitere B, Hubloue I, Tisseghem E, et al. One-year employment outcome prediction after traumatic brain injury: A CENTER-TBI study. *Disabil Health J* 2025;18:101716.
36. Ding R, Deng M, Wei H, Zhang Y, Wei L, Jiang G, et al. Machine learning-based prediction of clinical outcomes after traumatic brain injury: Hidden information of early physiological time series. *CNS Neurosci Ther* 2024;30:e14848.
37. Huang J, Chen H, Deng J, Liu X, Shu T, Yin C, et al. Interpretable machine learning for predicting 28-day all-cause in-hospital mortality for hypertensive ischemic or hemorrhagic stroke patients in the ICU: a multi-center retrospective cohort study with internal and external cross-validation. *Front Neurol* 2023;14:1185447.
38. Zeng W, Chen J, Shen L, Xia G, Xie J, Zheng S, et al. Clinical, radiological, and radiomics feature-based explainable machine learning models for prediction of neurological deterioration and 90-day outcomes in mild intracerebral hemorrhage. *BMC Med Imaging* 2025;25:184.
39. Shu L, Yan H, Wu Y, Yan T, Yang L, Zhang S, et al. Explainable machine learning in outcome prediction of high-grade aneurysmal subarachnoid hemorrhage. *Aging (Albany NY)* 2024;16:4654–69.

Figure Legends

Figure 1: Flow Diagram

Abbr.: ICU Intensive Care Unit, GOSE Glasgow Outcome Scale Extended

Figure 2: AUROC and AUPRC for the respective time points

A Area under the receiver operating characteristic curve and **B** Area under the precision-recall at different timepoints from ICU admission until Hospital discharge. **C** Receiver operating characteristic curves and **D** Precision–recall curves comparing model performance. Here, the four models for the different timepoints (ICU admission, ICU day 3, ICU day 7 and Hospital discharge) are plotted. Abbr.: AUROC Area under the receiver operating characteristic curve AUPRC Area under the precision-recall curve CI Confidence Interval ICU Intensive Care Unit

Figure 3: Information gain and SHAP Contributions

A Machine Learning models indicating the importance of the influencing factors as well as the information gain at different time points (ICU admission, ICU day 3, ICU day 7 and Hospital discharge). Importance is presented as percentage. **B** Average Shapley values for the respective time points (Baseline ICU admission Day 3, 7 and Hospital discharge) representing the overall contribution of the influencing factors to the model's predictive power. Higher Shapley values indicate greater importance of the corresponding feature. The parameters “death”, “vegetative state”, “severe/moderate disability” and “good recovery” are the components of the GOSE (Glasgow Outcome Scale Extended), which were derived from the patients’ status at hospital discharge. The parameters “Rehabilitation unit”, “Other”, “Other hospital”, “Home” and “Nursing home” are the components of the discharge destination, which were derived from the patients’ medical file at hospital discharge. The parameters “TBI”, “SAH”, and “ICH” are the components of the admission diagnosis, which were determined at ICU admission. Abbr.: CSF Cerebrospinal Fluid Drainage; GCS Glasgow Coma Scale with the sub-components *M* Motor, *V* Verbal and *E* Eyes; GOSE Glasgow Outcome Scale Extended; *ICPm* Intracranial Pressure Monitoring; *ICH* Intracerebral Hemorrhage; *ICU* Intensive Care Unit; *SAH* Subarachnoid Hemorrhage; *TBI* Traumatic Brain Injury; *TIL* Therapy Intensity Level

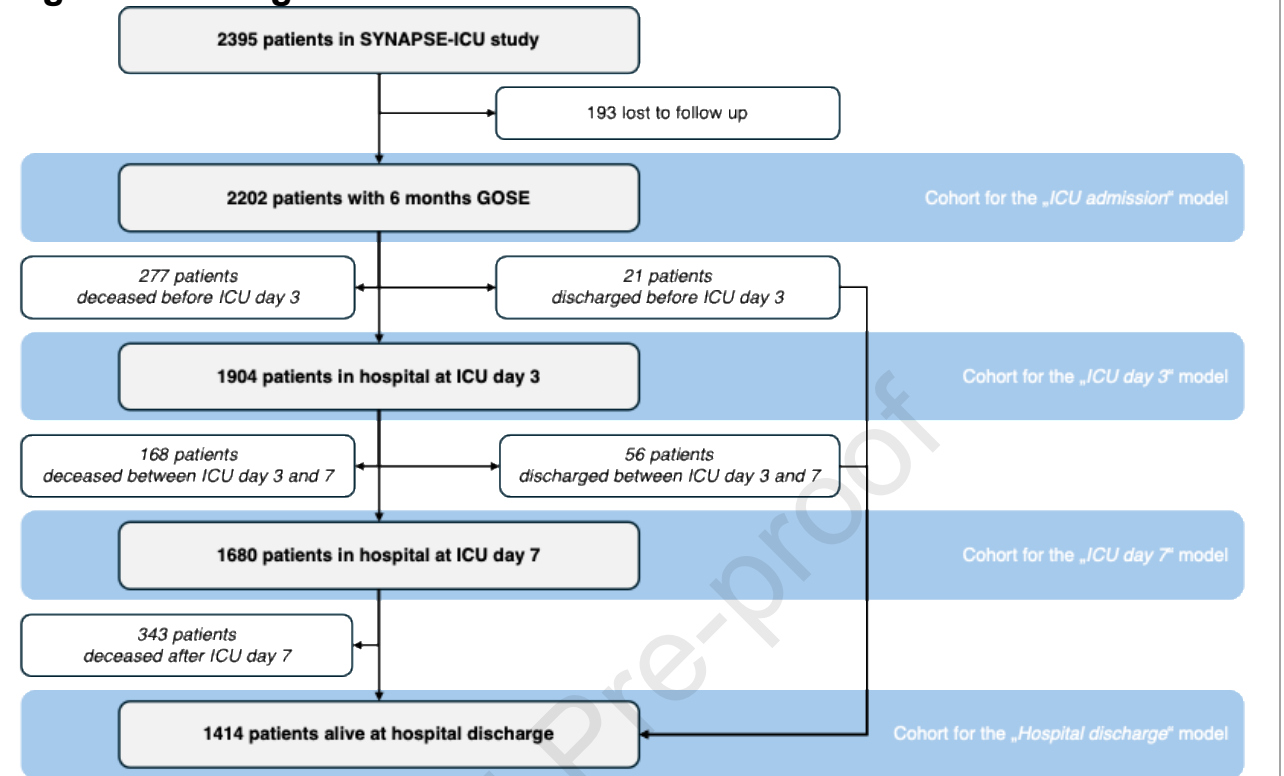
Figure 4: Probability of negative outcome according to ML models for the four time points

Figure 4 illustrates the trajectory of the predicted probability for an unfavorable long-term outcome (GOSE <5 at 6 months) across various time points. Each trajectory line represents an individual patient with the prediction generated by the respective machine learning models. The categories are color-coded according to the respective outcome: red for patients with unfavorable outcome, and purple for those with a favorable outcome. The two highlighted lines illustrate the prediction trajectories of two patients who had the same predicted risk upon ICU admission. Abbr.: GOSE Glasgow Outcome Scale Extended, ICU Intensive Care Unit, ML Machine Learning

Figure 5: Waterfall Plots

Waterfall plots illustrate explanations for individual predictions. The input is a vector of variables that are responsible for the prediction of an individual patient. The waterfall plot starts at the top at the baseline risk of our entire cohort at each time point. Red bars represent variables that increase risk, blue bars represent those that decrease risk. Contributions over 7% are shown in the graph. The patient's predicted probability of a negative outcome is shown on the x-axis. The baseline changes with the contribution of each value to the overall risk and ends at the individual risk estimate for the respective. The predictions of the models at the 4 different time points of 2 patients are shown. The prediction probability of a negative outcome for both patients at ICU admission is nearly identical. (37.6% vs 37.7%). On the left side a 28-year-old patient with a low predicted probability. Abbr.: *CPP* Cerebral Perfusion Pressure; *GCS* Glasgow Coma Scale with the sub-components *M* Motor, *V* Verbal and *E* Eyes; *GOSE* Glasgow Outcome Scale Extended; *ICU* Intensive Care Unit; *LOS* Length of Stay; *ML* Machine Learning; *TBI* Traumatic Brain Injury; *TIL* Therapy Intensity Level

Figures

Fig. 1 Flow Diagram

Abbr.: ICU Intensive Care Unit, GOSE Glasgow Outcome Scale Extended

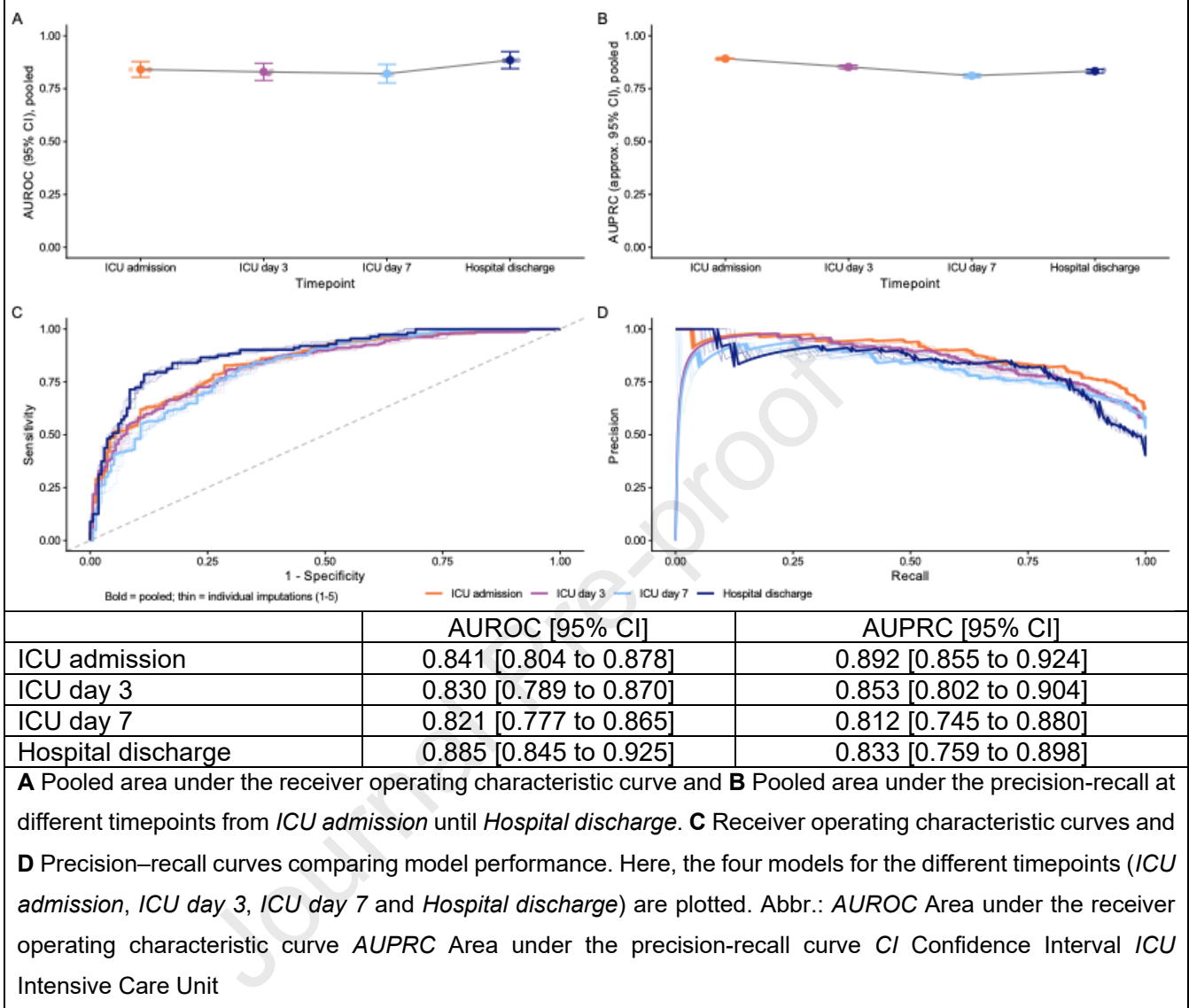
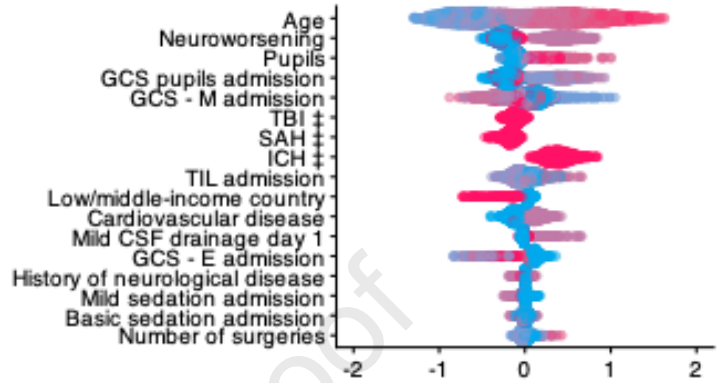
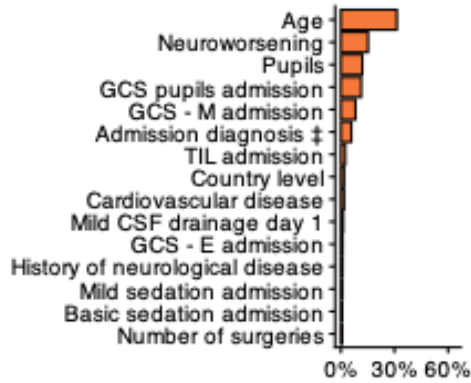
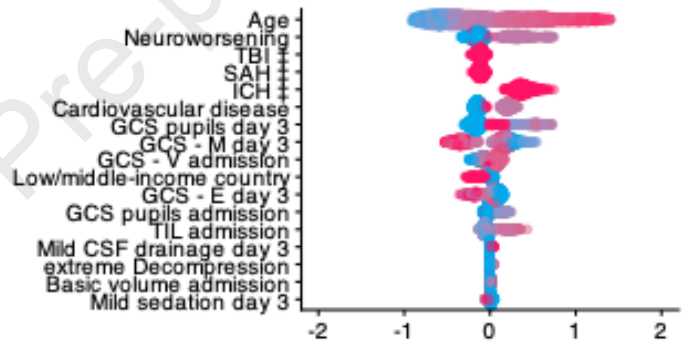
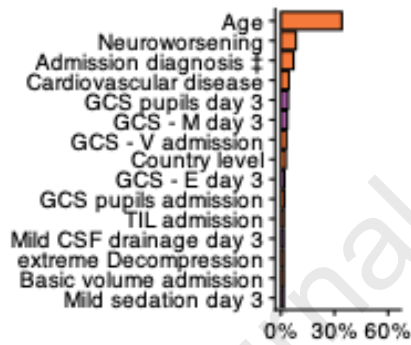
Fig 2. Pooled AUROC and pooled AUPRC for the respective time points

Fig. 3 Information gain and SHAP Contributions**A****B**

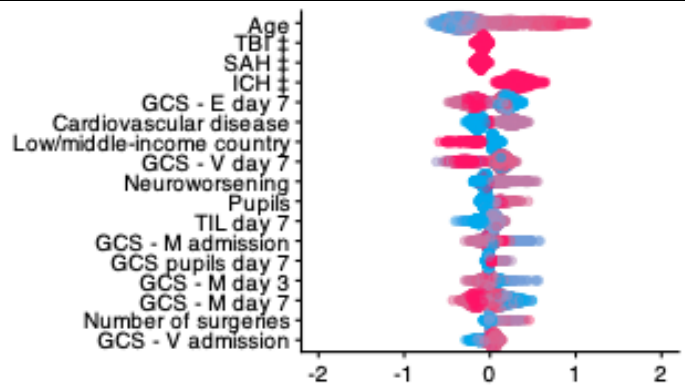
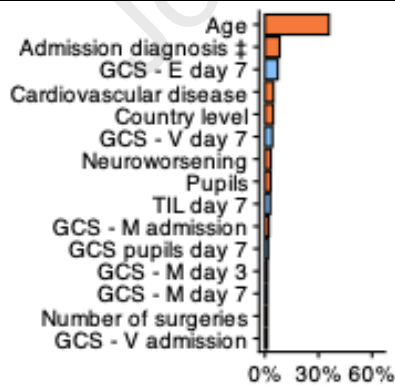
ICU admission



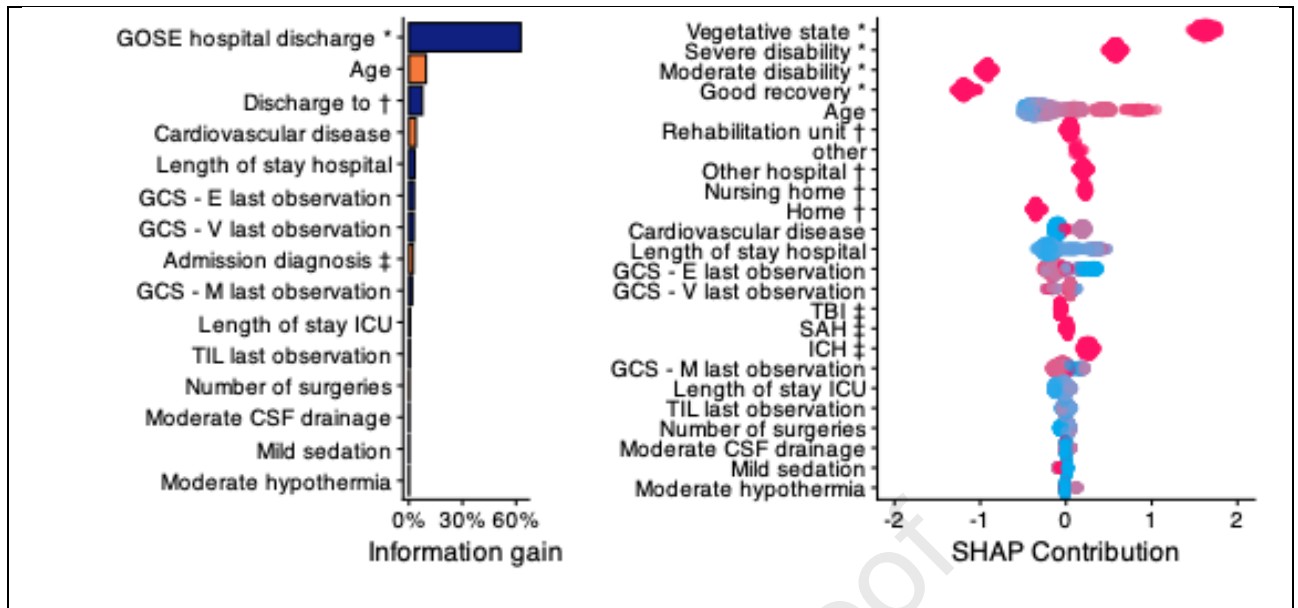
ICU day 3



ICU day 7



Hospital discharge



A Machine Learning models indicating the importance of the influencing factors as well as the information gain at different time points (ICU admission, ICU day 3, ICU day 7 and Hospital discharge). Importance is presented as percentage. **B** Average Shapley values for the respective time points (Baseline ICU admission Day 3, 7 and Hospital discharge) representing the overall contribution of the influencing factors to the model's predictive power. Higher Shapley values indicate greater importance of the corresponding feature. The parameters “death”, “vegetative state”, “severe/moderate disability” and “good recovery” are the components of the GOSE (Glasgow Outcome Scale Extended), which were derived from the patients’ status at hospital discharge. The parameters “Rehabilitation unit”, “Other”, “Other hospital”, “Home” and “Nursing home” are the components of the discharge destination, which were derived from the patients’ medical file at hospital discharge. The parameters “TBI”, “SAH”, and “ICH” are the components of the admission diagnosis, which were determined at ICU admission. Abbr.: CSF Cerebrospinal Fluid Drainage; GCS Glasgow Coma Scale with the sub-components *M* Motor, *V* Verbal and *E* Eyes; GOSE Glasgow Outcome Scale Extended; *ICPm* Intracranial Pressure Monitoring; *ICH* Intracerebral Hemorrhage; *ICU* Intensive Care Unit; *SAH* Subarachnoid Hemorrhage; *TBI* Traumatic Brain Injury; *TIL* Therapy Intensity Level

Fig. 4 Probability of negative outcome according to ML models for the four time points

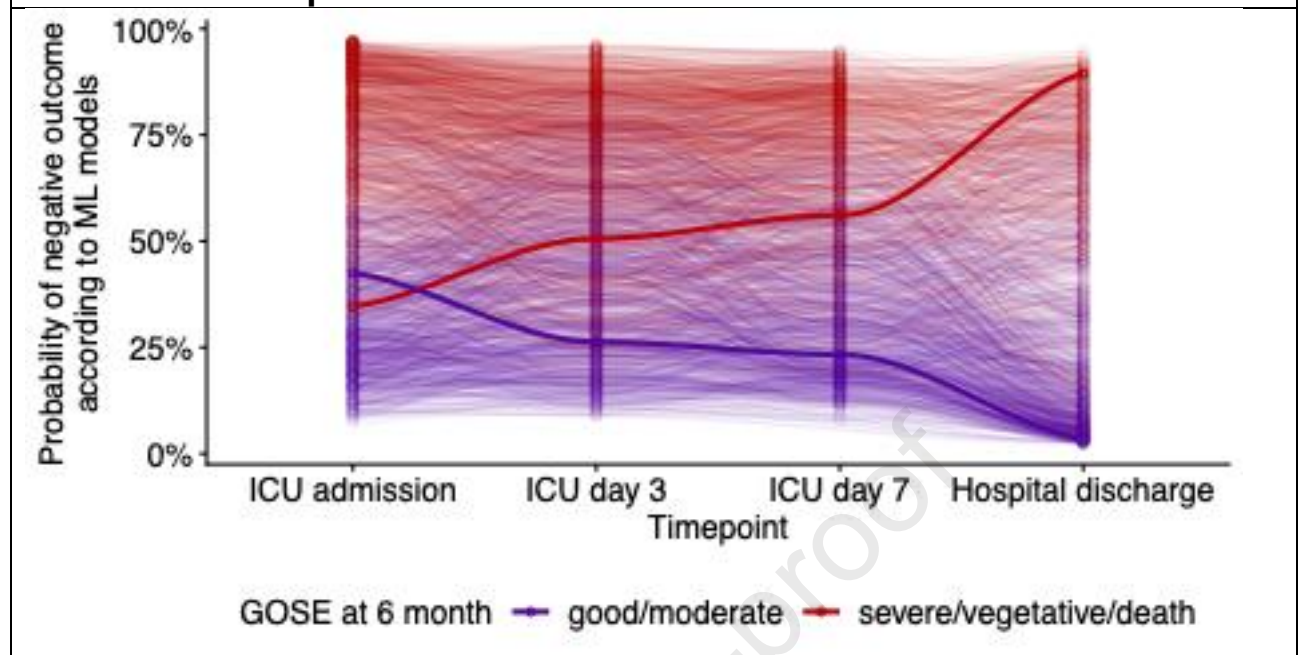
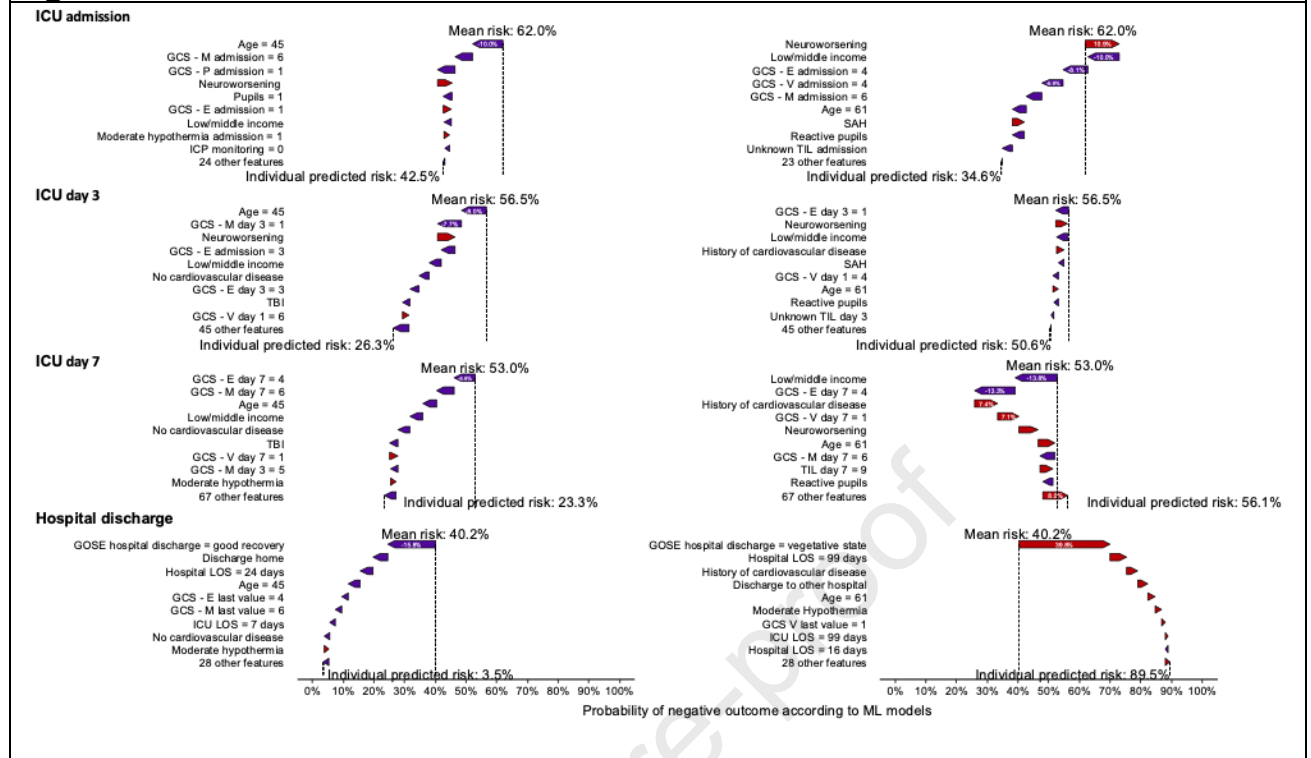


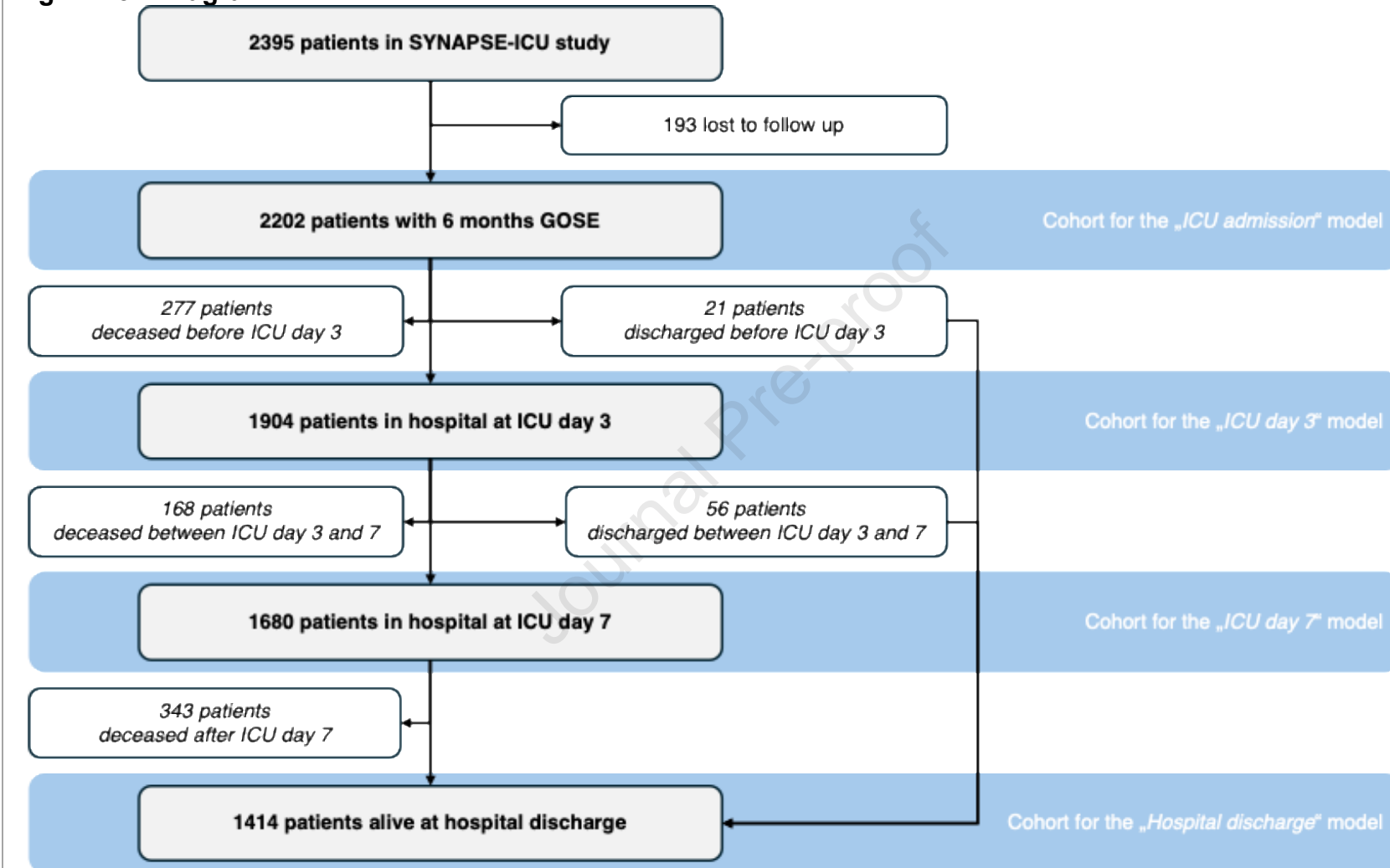
Figure 4 illustrates the trajectory of the predicted probability for an unfavorable long-term outcome (GOSE <5 at 6 months) across various time points. Each trajectory line represents an individual patient with the prediction generated by the respective machine learning models. The categories are color-coded according to the respective outcome: red for patients with unfavorable outcome, and purple for those with a favorable outcome. The two highlighted lines illustrate the prediction trajectories of two patients who had the same predicted risk upon ICU admission. Abbr.: GOSE Glasgow Outcome Scale Extended, ICU Intensive Care Unit, ML Machine Learning

Fig. 5 Waterfall Plots

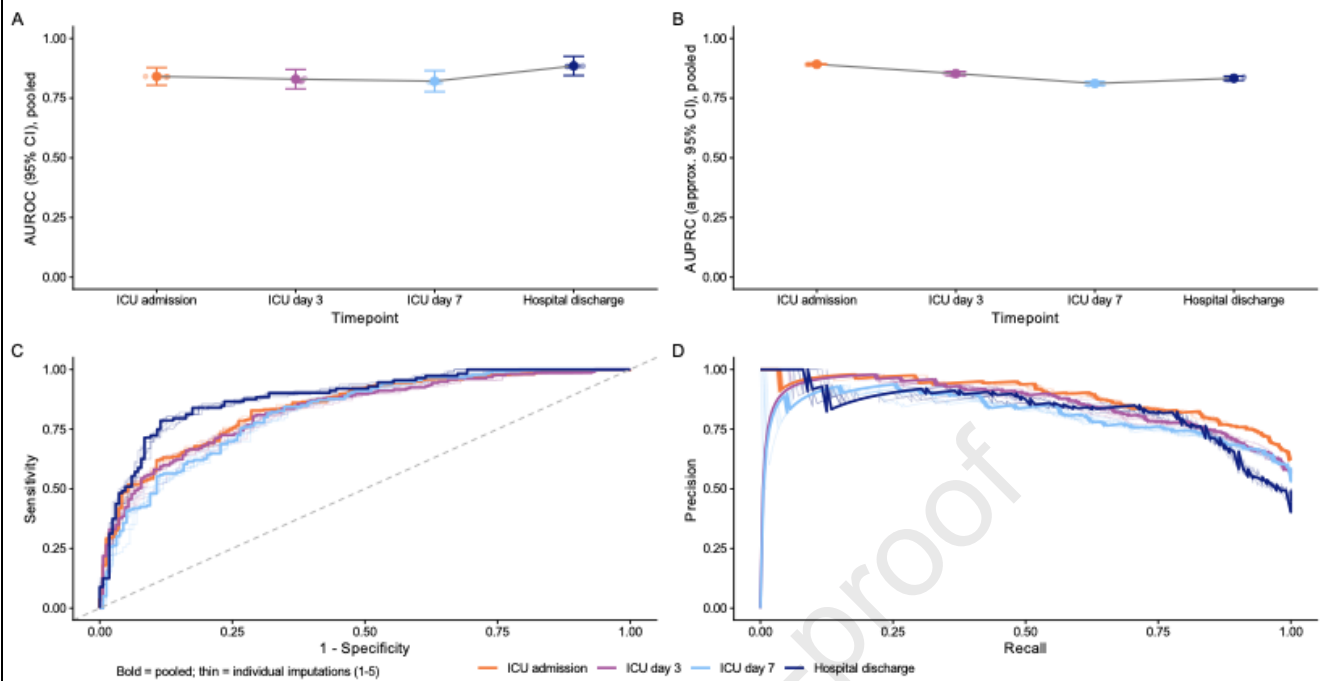
Waterfall plots illustrate explanations for individual predictions. The input is a vector of variables that are responsible for the prediction of an individual patient. The waterfall plot starts at the top at the baseline risk of our entire cohort at each time point. Red bars represent variables that increase risk, blue bars represent those that decrease risk. Contributions over 7% are shown in the graph. The patient's predicted probability of a negative outcome is shown on the x-axis. The baseline changes with the contribution of each value to the overall risk and ends at the individual risk estimate for the respective. The predictions of the models at the 4 different time points of 2 patients are shown. The prediction probability of a negative outcome for both patients at ICU admission is nearly identical. (37.6% vs 37.7%). On the left side a 28-year-old patient with a low predicted probability. Abbr.: CPP Cerebral Perfusion Pressure; GCS Glasgow Coma Scale with the sub-components M Motor, V Verbal and E Eyes; GOSE Glasgow Outcome Scale Extended; ICU Intensive Care Unit; LOS Length of Stay; ML Machine Learning; TBI Traumatic Brain Injury; TIL Therapy Intensity Level

Tables

Table 1 Patient characteristics				
	Overall, n = 2,202	Good recovery / moderate disability after 6 months, n = 836	Severe disability/ vegetative state/ death after 6 months, n = 1,366	p-value
Age in years - median [IQR]	56 [40, 69]	46 [32, 59]	62 [48, 74]	<0.001
Female - n (%)	771 (35%)	272 (33%)	499 (37%)	0.059
Admission diagnosis				<0.001
Traumatic brain injury - n (%)	1,173 (53%)	544 (65%)	629 (46%)	
Subarachnoid hemorrhage - n (%)	485 (22%)	187 (22%)	298 (22%)	
Intracerebral hemorrhage - n (%)	544 (25%)	105 (13%)	439 (32%)	
ICP monitoring - n (%)	1,220 (55%)	487 (58%)	733 (54%)	0.038
History of cardiovascular disease - n (%)	922 (43%)	216 (27%)	706 (53%)	<0.001
History of neurological disease - n (%)	269 (13%)	74 (9%)	195 (15%)	<0.001
ICU length of stay in days – median [IQR]	11 [4, 21]	12 [6, 21]	10 [3, 20]	<0.001
Hospital length of stay in days – median [IQR]	19 [8, 38]	25 [15, 43]	14 [4, 36]	<0.001
GCS on admission - median [IQR]	5 [3, 7]	6 [3, 8]	4 [3, 7]	<0.001
GOSE at hospital discharge				<0.001
Death - n (%)	788 (36%)	0 (0%)	788 (59%)	
Vegetative state - n (%)	174 (8%)	17 (2%)	157 (12%)	
Severe disability - n (%)	561 (26%)	228 (27%)	333 (25%)	
Moderate disability - n (%)	253 (12%)	208 (25%)	45 (3%)	
Good recovery - n (%)	399 (18%)	377 (45%)	22 (2%)	
Continuous variables were described as median [IQR], and categorical data were expressed as frequency (%). Abbr.: GCS Glasgow Coma Scale, GOSE Glasgow Outcome Score Extended, ICP Intracranial Pressure, ICU Intensive Care Unit, IQR Interquartile Range				

Fig. 1 Flow Diagram

Abbr.: ICU Intensive Care Unit, GOSE Glasgow Outcome Scale Extended

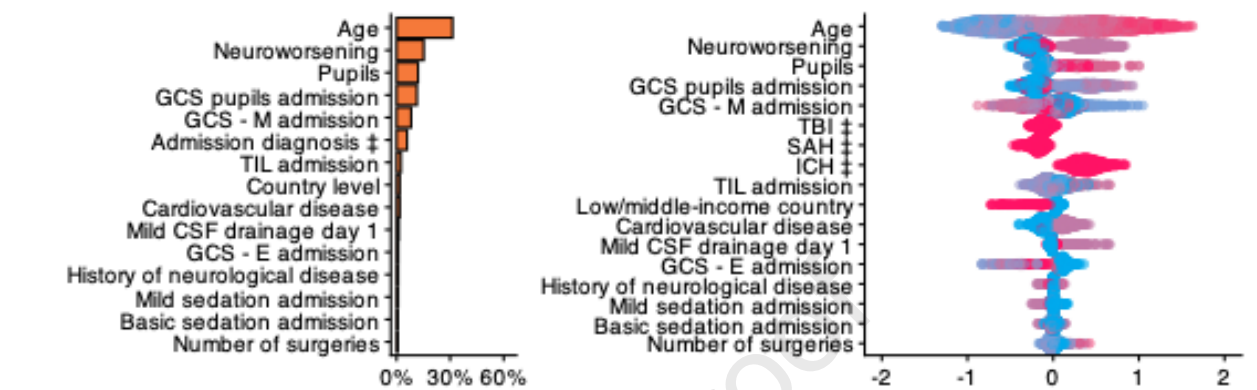
Fig 2. Pooled AUROC and pooled AUPRC for the respective time points

	AUROC [95% CI]	AUPRC [95% CI]
ICU admission	0.841 [0.804 to 0.878]	0.892 [0.855 to 0.924]
ICU day 3	0.830 [0.789 to 0.870]	0.853 [0.802 to 0.904]
ICU day 7	0.821 [0.777 to 0.865]	0.812 [0.745 to 0.880]
Hospital discharge	0.885 [0.845 to 0.925]	0.833 [0.759 to 0.898]

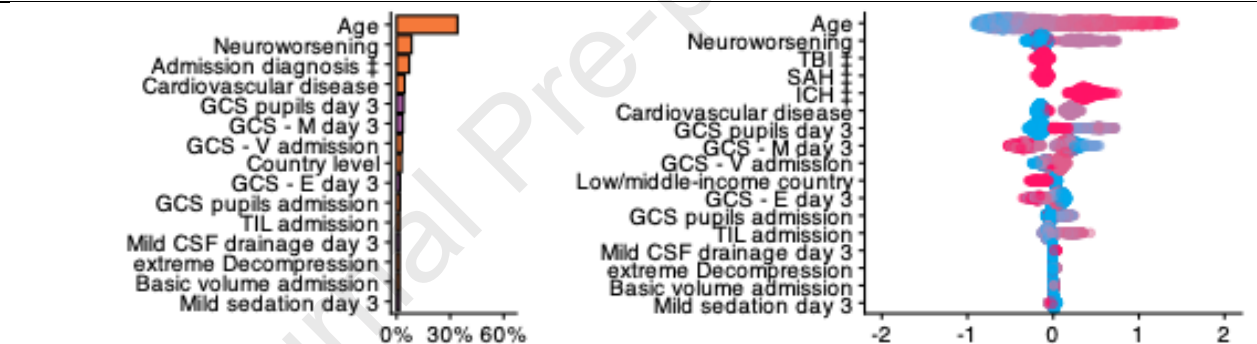
A Pooled area under the receiver operating characteristic curve and **B** Pooled area under the precision-recall at different timepoints from *ICU admission* until *Hospital discharge*. **C** Receiver operating characteristic curves and **D** Precision-recall curves comparing model performance. Here, the four models for the different timepoints (*ICU admission*, *ICU day 3*, *ICU day 7* and *Hospital discharge*) are plotted. Abbr.: *AUROC* Area under the receiver operating characteristic curve *AUPRC* Area under the precision-recall curve *CI* Confidence Interval *ICU* Intensive Care Unit

Fig. 3 Information gain and SHAP Contributions**A****B**

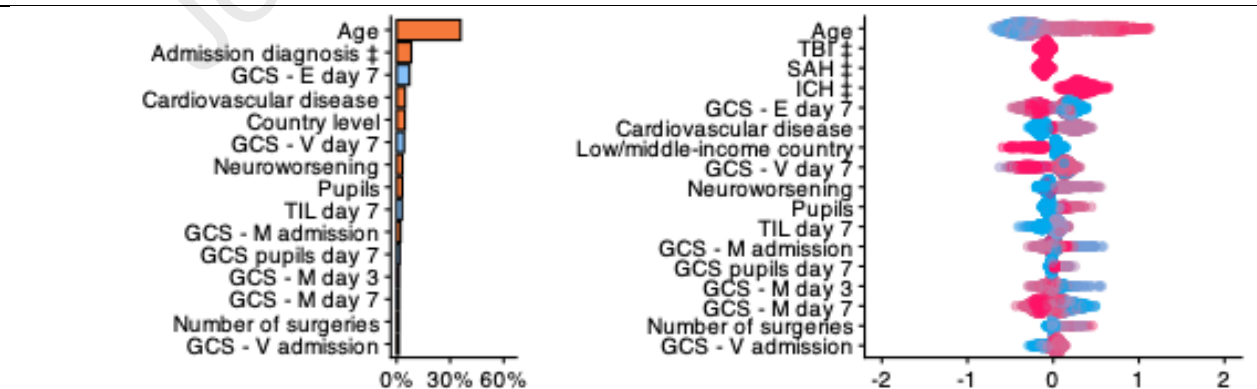
ICU admission



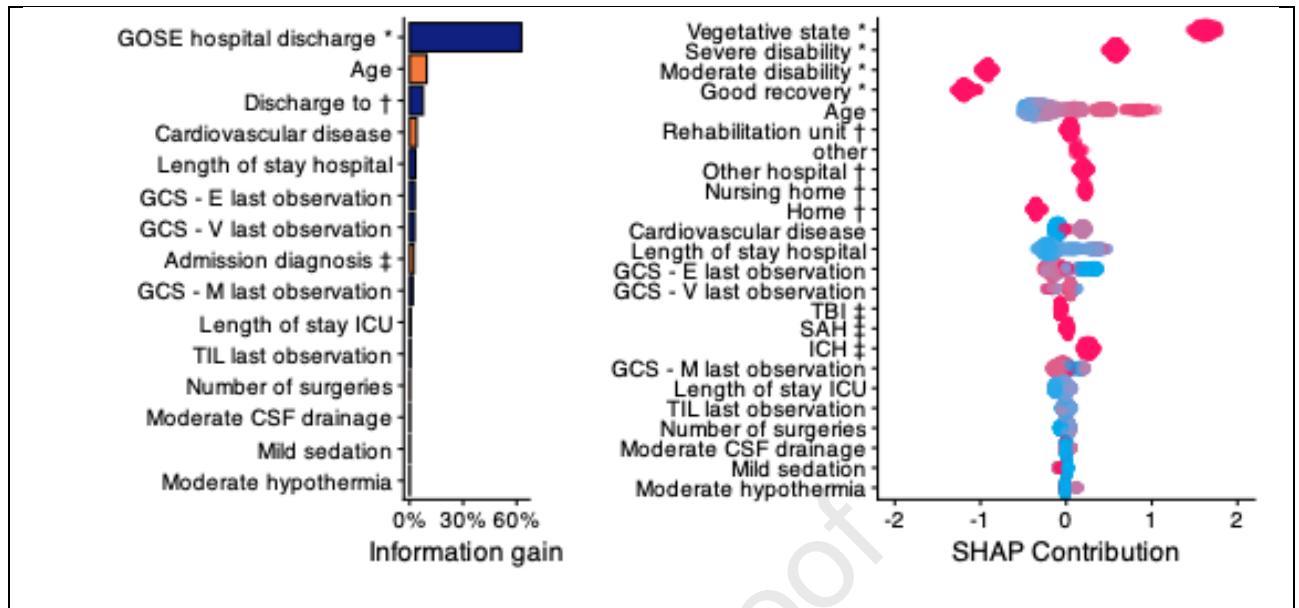
ICU day 3



ICU day 7



Hospital discharge



A Machine Learning models indicating the importance of the influencing factors as well as the information gain at different time points (ICU admission, ICU day 3, ICU day 7 and Hospital discharge). Importance is presented as percentage. **B** Average Shapley values for the respective time points (Baseline ICU admission Day 3, 7 and Hospital discharge) representing the overall contribution of the influencing factors to the model's predictive power. Higher Shapley values indicate greater importance of the corresponding feature. The parameters “death”, “vegetative state”, “severe/moderate disability” and “good recovery” are the components of the GOSE (Glasgow Outcome Scale Extended), which were derived from the patients’ status at hospital discharge. The parameters “Rehabilitation unit”, “Other”, “Other hospital”, “Home” and “Nursing home” are the components of the discharge destination, which were derived from the patients’ medical file at hospital discharge. The parameters “TBI”, “SAH”, and “ICH” are the components of the admission diagnosis, which were determined at ICU admission. Abbr.: CSF Cerebrospinal Fluid Drainage; GCS Glasgow Coma Scale with the sub-components *M* Motor, *V* Verbal and *E* Eyes; GOSE Glasgow Outcome Scale Extended; *ICPm* Intracranial Pressure Monitoring; *ICH* Intracerebral Hemorrhage; *ICU* Intensive Care Unit; *SAH* Subarachnoid Hemorrhage; *TBI* Traumatic Brain Injury; *TIL* Therapy Intensity Level

Fig. 4 Probability of negative outcome according to ML models for the four time points

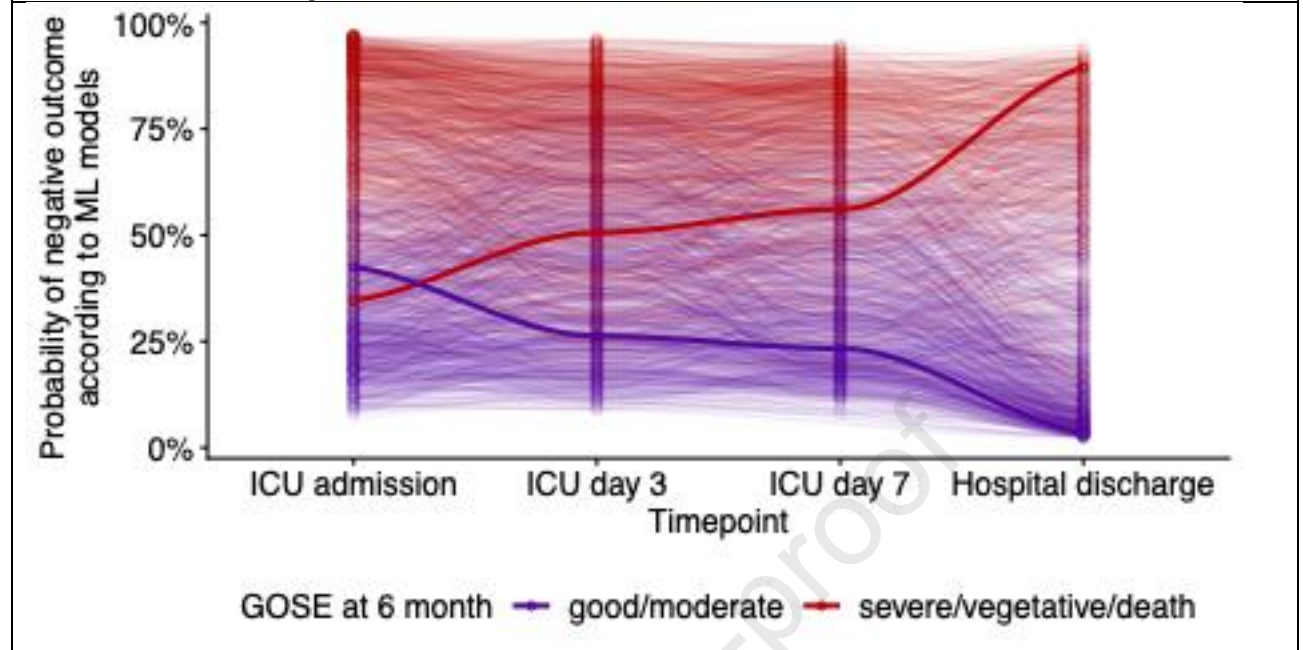
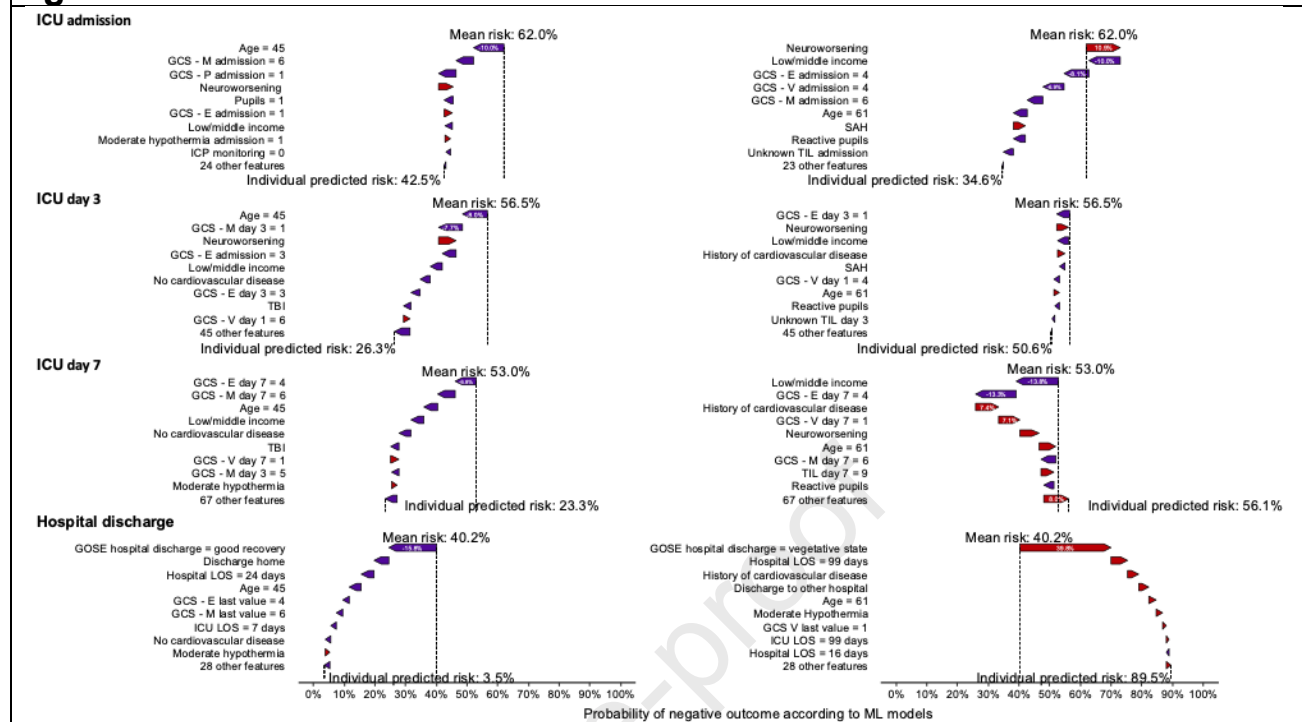


Figure 4 illustrates the trajectory of the predicted probability for an unfavorable long-term outcome (GOSE <5 at 6 months) across various time points. Each trajectory line represents an individual patient with the prediction generated by the respective machine learning models. The categories are color-coded according to the respective outcome: red for patients with unfavorable outcome, and purple for those with a favorable outcome. The two highlighted lines illustrate the prediction trajectories of two patients who had the same predicted risk upon ICU admission. Abbr.: GOSE Glasgow Outcome Scale Extended, ICU Intensive Care Unit, ML Machine Learning

Fig. 5 Waterfall Plots

Waterfall plots illustrate explanations for individual predictions. The input is a vector of variables that are responsible for the prediction of an individual patient. The waterfall plot starts at the top at the baseline risk of our entire cohort at each time point. Red bars represent variables that increase risk, blue bars represent those that decrease risk. Contributions over 7% are shown in the graph. The patient's predicted probability of a negative outcome is shown on the x-axis. The baseline changes with the contribution of each value to the overall risk and ends at the individual risk estimate for the respective. The predictions of the models at the 4 different time points of 2 patients are shown. The prediction probability of a negative outcome for both patients at ICU admission is nearly identical. (37.6% vs 37.7%). On the left side a 28-year-old patient with a low predicted probability. Abbr.: *CPP* Cerebral Perfusion Pressure; *GCS* Glasgow Coma Scale with the sub-components *M* Motor, *V* Verbal and *E* Eyes; *GOSE* Glasgow Outcome Scale Extended; *ICU* Intensive Care Unit; *LOS* Length of Stay; *ML* Machine Learning; *TBI* Traumatic Brain Injury; *TIL* Therapy Intensity Level

Abbreviations to:**Machine Learning for Predicting Long-Term Outcomes in****Neurocritical Care:****A sub-analysis from the SYNAPSE-ICU Study**

AUROC Area under the receiver operating characteristic curve

AUPRC Area under the precision-recall curve

CPP Cerebral Perfusion Pressure

CSF Cerebrospinal Fluid Drainage

DRF Distributed Random Forest

GBM Gradient Boosting Machine

GCS Glasgow Coma Scale

GLM Generalized Linear Model with regularization

GOSE Glasgow Outcome Scale Extended

ICH Intracerebral Hemorrhage

ICP Intracranial Pressure

ICPm Intracranial Pressure Monitoring

ICU Intensive Care Unit

IQR Interquartile Range

LOS Length of Stay

SAH Subarachnoid Hemorrhage

SHAP Shapley Additive Explanations Plots

TBI Traumatic Brain Injury

TIL Therapy Intensity Level

XGBoost eXtreme Gradient Boosting

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Stefan J Schaller reports a relationship with Reactive Robotics GmbH (Munich, Germany) that includes: funding grants and non-financial support. Stefan J Schaller reports a relationship with ASP GmbH (Attendorn, Germany) that includes: funding grants and non-financial support. Stefan J Schaller reports a relationship with STIMIT AG (Biel, Switzerland) that includes: funding grants and non-financial support. Stefan J Schaller reports a relationship with European Society of Intensive Care Medicine that includes: funding grants and non-financial support. Stefan J Schaller reports a relationship with Fresenius Kabi Germany that includes: consulting or advisory, funding grants, and non-financial support. Stefan J Schaller reports a relationship with Innovationsfond of The Federal Joint Committee (G-BA) that includes: funding grants. Stefan J Schaller reports a relationship with Springer-Verlag GmbH that includes: speaking and lecture fees. Stefan J Schaller reports a relationship with Advanz Pharma GmbH (Bielefeld, Germany) that includes: consulting or advisory. Stefan J Schaller reports a relationship with Alphabet Inc that includes: equity or stocks. Stefan J Schaller reports a relationship with Bayer AG that includes: equity or stocks. Stefan J Schaller reports a relationship with Siemens AG that includes: equity or stocks. Giuseppe Citerio reports a relationship with Integra that includes: consulting or advisory and funding grants. Giuseppe Citerio reports a relationship with NeurOptics Inc that includes: consulting or advisory and funding grants. Giuseppe Citerio reports a relationship with Natus that includes: consulting or advisory. Giuseppe Citerio participated in the manuscript preparation during their personal involvement in the Italian Minister of University MUR Dipartimenti di Eccellenza 2023-2027 (l. 232/2016, art. 1, commi 314 - 337) and in PRIN 2022SYXEH. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.