

Dynamic prediction of mortality in trauma patients: Development of a hybrid neural network model for risk assessment

Andreas Skov Millarch, Haytham Kaafarani, Ibrahim Chamseddine, Fredrik Folke,
Søren Steeman Rudolph, Martin Sillesen

Submitted to: Journal of Medical Internet Research
on: October 20, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 28

..... 28

Multimedia Appendixes 29

Multimedia Appendix 1..... 29



Dynamic prediction of mortality in trauma patients: Development of a hybrid neural network model for risk assessment

Andreas Skov Millarch¹ Msc; Haytham Kaafarani² MD; Ibrahim Chamseddine² Phd; Fredrik Folke³ MD; Søren Steeman Rudolph¹ MD; Martin Sillesen¹ MD

¹ Rigshospitalet Copenhagen DK

² Massachusetts General Hospital Boston US

³ Gentofte Hospital Gentofte DK

Corresponding Author:

Martin Sillesen MD

Rigshospitalet
Blegdamsvej 9
Copenhagen
DK

Abstract

Background: The trajectory of a trauma patient is often complex and non-linear. Real-time estimation of the mortality risk from prehospital care to discharge is critical for point-of-care decision-making and for benchmarking the quality of care. Using data automatically extracted from electronic health records, this study aims to create an artificial intelligence risk prediction model for 90-day mortality in trauma patients capable of performing predictions at any time point through treatment phases from prehospital to discharge.

Objective: To develop risk prediction models based on transformer architectures for dynamic mortality prediction

Methods: Comprehensive data on pre- and in-hospital care of trauma patients treated in Denmark, Capital Region between 2017 and 2024 was used to train, validate, and evaluate the AI model. Demographic, comorbidity, and injury specific variables were structured as tabular data. Temporal data on vital signs, laboratory test results, and medications were structured as sequences with temporally determined dynamic bin sizes. The model architecture combines both tabular and sequential input data. A unified input is processed by multiple layers of transformer encoders before being passed into a neural network for binary classification.

Model performance was assessed using area under the receiver operating characteristic curve (AUROC) and area under precision-recall curve (AP) on a holdout dataset comprising 20% of the total data.

Results: A total of 8,372 patients were included and the 90-day mortality rate was 4.55%. The model achieved an AUROC of 0.963 (95% CI: 0.934-0.993) and an AUPRC of 0.800 (95% CI: 0.712-0.870) predicting 90-day mortality in the holdout evaluation set using full-length trajectories, comprising 1675 patients.

Conclusions: We designed a dynamic, automated, and highly accurate model that allows 90-day mortality risk prediction at any point in time during the complex and non-linear trajectory of the trauma patient. The model could be used as decision support for triage, patient deterioration alerts, bedside decision-making and family counselling. Clinical Trial: NA

(JMIR Preprints 20/10/2025:86202)

DOI: <https://doi.org/10.2196/preprints.86202>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org>



Original Manuscript

Dynamic prediction of mortality in trauma patients: Development of a hybrid neural network model for risk assessment

Andreas Skov Millarch MSc^{1,2}, Haytham M. Kaafarani MD MPH³, Ibrahim Chamseddine PhD⁴, Fredrik Folke MD PhD^{5,6,8}, Søren S. Rudolph MD⁷, and Martin Sillesen MD PhD^{1,2,8}

1. Department of Organ Surgery and Transplantation, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
2. Center for Surgical Translational and Artificial Intelligence Research (CSTAR), Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
3. Division of Trauma, Emergency Surgery and Surgical Critical Care, Massachusetts General Hospital and Harvard Medical School, Boston, USA
4. Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
5. Copenhagen Emergency Medical Services, University of Copenhagen, Ballerup, Denmark
6. Department of Cardiology, Herlev Gentofte University Hospital, Denmark.
7. Department of Anesthesia, Center of Head and Orthopedics, Rigshospitalet, Copenhagen, Denmark.
8. Department of Clinical Medicine, University of Copenhagen, Denmark

Corresponding author:

Martin Sillesen MD, PhD
Assoc. Professor of Surgery
Copenhagen University Hospital, Rigshospitalet
Blegdamsvej 9
2100 Copenhagen Ø
Denmark
Mail: martin.sillesen@regionh.dk

Funding: The study was supported by grants from the Danish Victims foundation (Grant #21-610-00127) and the Novo Nordisk foundation (Grant #NNF19OC0055183) from the Novo Nordisk Foundation to MS.

Manuscript category: Original article

Conflicts of interest: Author MS has co-founded by the company Aiomic, a company developing AI models for the healthcare sector. The present work does not relate to any commercial activities and is for research only.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability statement: Due to confidentiality issues, we are not at liberty to share the Electronic Health Record (pre- and in-hospital) data used in this study. Access to Danish health data is granted through The Danish Patient Safety Authority. EHR data can then be obtained by reasonable request through the relevant region.

Source code availability statement: Project source code is available at the following GitHub repository <https://github.com/a-millarch/dynamic-risk-prediction>.

ABSTRACT

BACKGROUND

The trajectory of a trauma patient is often complex and non-linear. Real-time estimation of the mortality risk from prehospital care to discharge is critical for point-of-care decision-making and for benchmarking the quality of care. Using data automatically extracted from electronic health records, this study aims to create an artificial intelligence risk prediction model for 90-day mortality in trauma patients capable of performing predictions at any time point through treatment phases from prehospital to discharge.

METHODS

Comprehensive data on pre- and in-hospital care of trauma patients treated in Denmark, Capital Region between 2017 and 2024 was used to train, validate, and evaluate the AI model. Demographic, comorbidity, and injury specific variables were structured as tabular data. Temporal data on vital signs, laboratory test results, and medications were structured as sequences with temporally determined dynamic bin sizes. The model architecture combines both tabular and sequential input data. A unified input is processed by multiple layers of transformer encoders before being passed into a neural network for binary classification. Model performance was assessed using area under the receiver operating characteristic curve (AUROC) and area under precision-recall curve (AP) on a holdout dataset comprising 20% of the total data.

RESULTS

A total of 8,372 patients were included and the 90-day mortality rate was 4.55%. The model achieved an AUROC of 0.963 (95% CI: 0.934-0.993) and an AUPRC of 0.800 (95% CI: 0.712-0.870) predicting 90-day mortality in the holdout evaluation set using full-length trajectories, comprising 1675 patients.

CONCLUSION

We designed a dynamic, automated, and highly accurate model that allows 90-day mortality risk prediction at any point in time during the complex and non-linear trajectory of the trauma patient. The model could be used as decision support for triage, patient deterioration alerts, bedside decision-making and family counselling.

INTRODUCTION

The trajectory of a trauma patient is often complex and non-linear, creating a highly dynamic data flow that has the potential to predict clinically relevant outcomes^{1,2}. This data flow starts from the prehospital setting, continues through to hospital treatment and discharge, encompassing a wide range of parameters, including physiological measurements, clinical assessments, treatment interventions, and patient responses. Real-time estimation of the mortality risk from prehospital care to discharge is critical for decision-making and benchmarking the quality of care. Firstly, it informs crucial decision-making processes, allowing healthcare providers to adapt treatment strategies promptly and allocate resources efficiently. Secondly, it serves as a valuable tool for benchmarking the quality of care across different trauma centres and systems. As healthcare systems increasingly leverage big data and machine learning technologies, the integration of dynamic risk assessment models into clinical practice holds promise for significantly improving patient outcomes in trauma care³.

Conventionally, mortality risk estimation in trauma only relies on summarized information about the patient state at a certain timepoint, e.g. first value of systolic blood pressure, or by aggregation, e.g. lowest systolic blood pressure. A prime example is the widely used Trauma Score and Injury Severity Score⁴ (TRISS), which is calculated by trauma type (blunt or penetrating), age group, Revised Trauma Score⁵ (RTS) and Injury Severity Score⁶ (ISS). The RTS, a component of TRISS, is calculated using the initial values of Glasgow Coma Scale⁷ (GCS), systolic blood pressure and respiratory rate. While these established scoring systems have demonstrated clinical utility, they have significant limitations. By reducing complex time-series data to single values, they inherently discard a wealth of potentially valuable information about the patient's evolving condition. This approach fails to capture the dynamic nature of trauma progression and response to treatment, potentially overlooking critical trends and patterns that could inform more accurate risk assessments and guide clinical decision-making. As healthcare technology advances, there is a growing opportunity to leverage the continuous stream of patient data being collected, potentially enhancing the precision and timeliness of mortality risk predictions in trauma care.

Machine learning (ML) models offer significant advantages in predicting trauma outcomes, particularly in their ability to handle complex data relationships and numerous input features. These models can capture non-linear relationships between variables that are often challenging to detect using classical statistical approaches. This capability is especially valuable in trauma care, where patient outcomes may depend on intricate interactions between multiple factors. Previous studies demonstrated superior predictive performance using ML compared to TRISS in predicting mortality^{8,9} and the potential to predict post-trauma complications¹⁰. In contrast to most previous studies which use static information, the model architecture used in this study combines tabular and time-series data through a sequential data fusion approach. This adds a sense of timing and context to the model, which currently can be harvested using attention, enhancing its ability to capture dynamic patterns and relationships.¹¹. Using a modelling approach that can capture intricate interactions between multiple factors with a sense of timing and context of changes in treatment and patient states may improve predictive performance.

The overall concept is depicted in Figure 1. Using data automatically extracted from prehospital and in-hospital Electronic Health Record (EHR) for ease of use in an acute clinical setting, this study aims to create a dynamic artificial intelligence (AI) risk prediction model for 90-day mortality in trauma patients, capable of performing dynamic predictions at any time point through treatment phases from prehospital to discharge fully utilizing the array of data generated as patients advance through treatment trajectories.

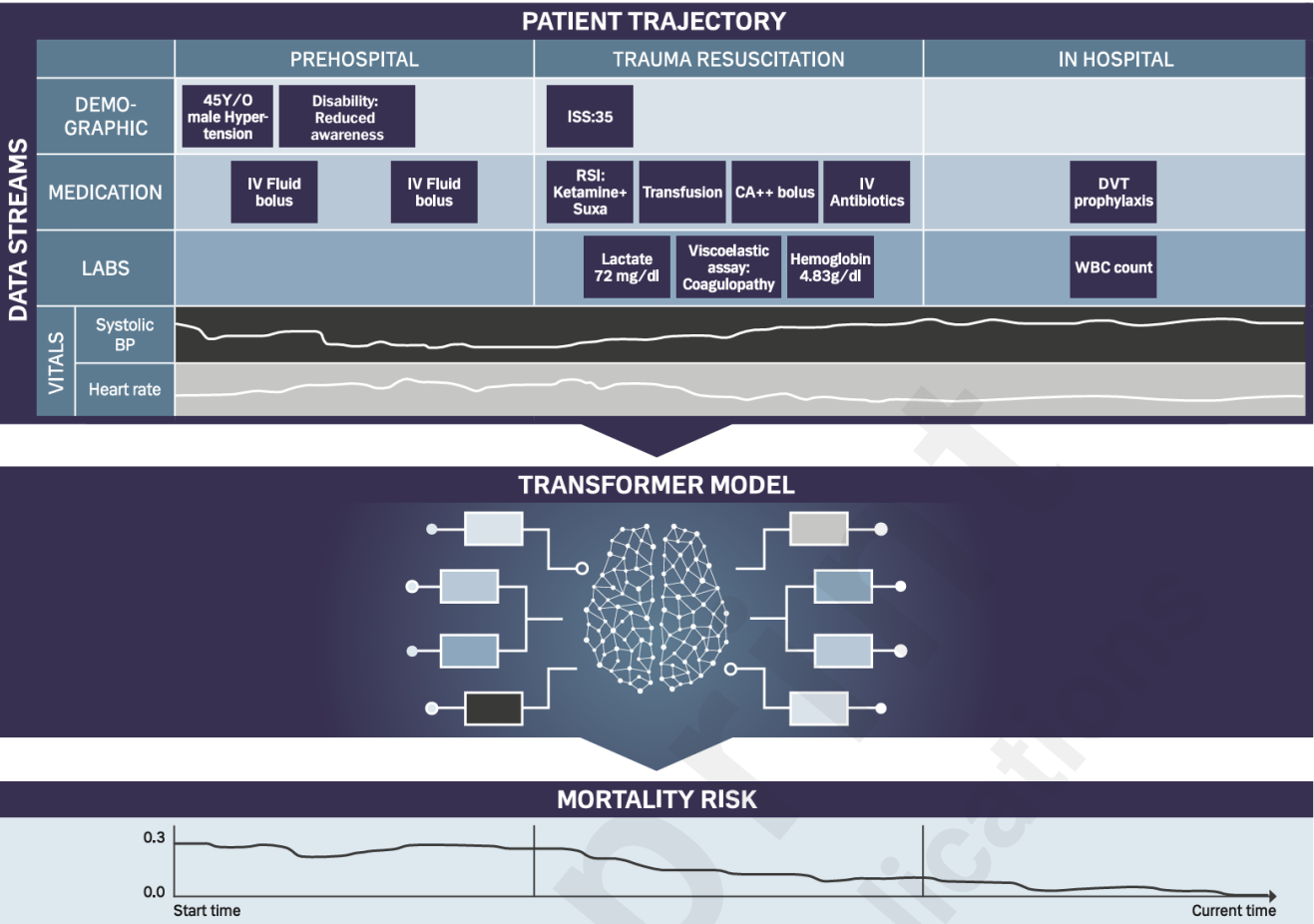


Figure 1. An example of a patient trajectory consisting of several phases and different data streams feeding into our model as they are created. Thus, displaying the clinical evolution of a patient and the dynamic development in mortality risk estimation over time from pre-hospital till discharge

METHODS

Data sources and inclusion

The study used retrospective data from pre- and in-hospital electronic health records (EHR) from trauma patients treated at any of the 6 acute care hospital in the Capital Region of Denmark from 2017 to June 2024. This region covers 1.9 million inhabitants, where healthcare (including acute care services) is offered free of charge for all citizens. Prehospital triage criteria to either local acute care hospital or the region's single level I trauma centre (Rigshospitalet) are based on clinical guidelines (Figure S1).

Prehospital EHR data was obtained through the Emergency Medical Services, Capital Region of Denmark. In-hospital EHR data was provided through the Capital Region Research Data Analysis Platform.

Patients were identified by treatment through an in-hospital EHR trauma call registration (Danish procedure code BWST1F). We included patients of all ages that were present in both the pre- and in-hospital EHR. Patients declared dead on scene were excluded from the dataset. Of note, trauma team activation criteria follow local hospital guidelines, and is thus not standardized.

Outcome

The model predicts 90-day mortality calculated from the trauma date. The date of death was obtained from the EHR which is updated with information of both in and out-of-hospital deaths using the Danish Central Person Registry. As such, the EHR data also provides information on post-discharge mortality.

Dataset construction and pre-processing

Trauma cases were identified as defined above (all patients meeting trauma team activation criteria at any acute care hospital in the Capital Region of Denmark). Demographics, comorbidity, and injury-specific variables were structured as tabular data and obtained from the EHR. Vital signs, laboratory test results, and medication were structured as sequential time series data. Prehospital data encompassed continuous measurements of vital signs, i.e. blood pressure; heart rate and oxygen saturation; Glasgow Coma Score (GCS) and ABCD assessments. In-hospital EHR data consisted of vital signs; GCS; medication groupings, i.e. opioids, antibiotics, neurological drugs, antithrombotic drugs, cardiovascular drugs, infusions, anaesthetics, diuretics, insulin, haemostatics and blood transfusions; lab results, i.e. lactate, base excess, haemoglobin, leukocyte count and thromboelastogram (TEG, Hemosol, Mount Juliet, TN, USA) measures. The underlying ATC codes for each medication group are presented in Table S1. Laboratory test groupings and underlying variables are described in Table S2.

The trajectory beginning was marked using the first available timestamp from the prehospital EHR for each trauma case. The in-hospital starting and ending time points were marked for each case using in-hospital data on admission, discharge, and transfer events. Using temporally determined binning frequencies for the data segments with increasing bin size over time up to 90 days allowed using a higher granularity of data at the beginning of patient trajectories, where the patient state is also more dynamic, while allowing longer sequences within model architectural limits. Binning frequencies for each segment are presented in Table S3.

The segments were used to map vital signs, laboratory test results, and medications across the entire patient trajectory when available. Vital signs, GCS, and laboratory tests were represented at their absolute values, and medication groups were binarily indicated in each segment. Vital signs were summarized in each segment as the mean and standard deviation of the binned values. Laboratory tests were summarized using the maximum value in each segment.

Diagnosis codes, based on the International Classification of Diseases (ICD), prior to the trauma event were used to calculate the Elixhauser Comorbidity Index¹² using the comorbidity R package¹³. Diagnosis codes registered during the hospital stay were used for estimating ISS using ICDPIC-R¹⁴, as previously demonstrated a feasible approach for a Danish trauma population¹⁵. Both were used as tabular data. Sex and age were also added to the tabular dataset.

Initial prehospital assessments of airway, breathing, circulation, and disability (ABCD)¹⁶ were also treated as tabular data using the predetermined values available to emergency service providers nationally on their EHR interface. The classifications are presented in Table S4. In the case of level of affectedness, these categories are not objectively defined but are based on the judgment of the prehospital provider using a mix of established reference ranges, patient-specific factors, and clinical context to decide whether a particular category is abnormal. Therefore, this model input feature should be regarded as a proxy for the clinician's level of concern regarding each of the ABCD categories.

The complete trauma dataset was split into three parts: A training dataset, a validation dataset, and a holdout dataset. The holdout dataset comprised 20 percent of the complete data and was only used after model training was completed to ensure proper model generalization. The remaining 80 percent of the complete dataset was further split into 80 percent training and 20 percent validation datasets. This process is illustrated in Figure S2. The records in training and validation were interchanged in a

5-fold training process as described in the methods section “Training process”. All splits were random with outcome stratification and seeding for reproducibility.

For the tabular dataset, additive binary variables indicating missing values were added for categorical and continuous features. Missing values were replaced with the most frequent value (mode). Categorical features were categorised, and continuous features were standardised per variable by mean and standard deviation. The training and validation dataset was standardised separately from the holdout dataset to avoid data leakage.

Training process

The training and validation dataset was split in an 80/20% split during cross-validation¹⁷ using 5 splits. For 50 trials, the folds were randomised but stratified on outcome prevalence. During this process, a range of hyper-parameters (Figure 2) were optimised against maximizing the sum of area under precision-recall curve across all folds. Using the optimised hyper-parameters found by sweeps, we trained the model using the full training and cross-validation dataset combined before evaluating model performance on the holdout test dataset.

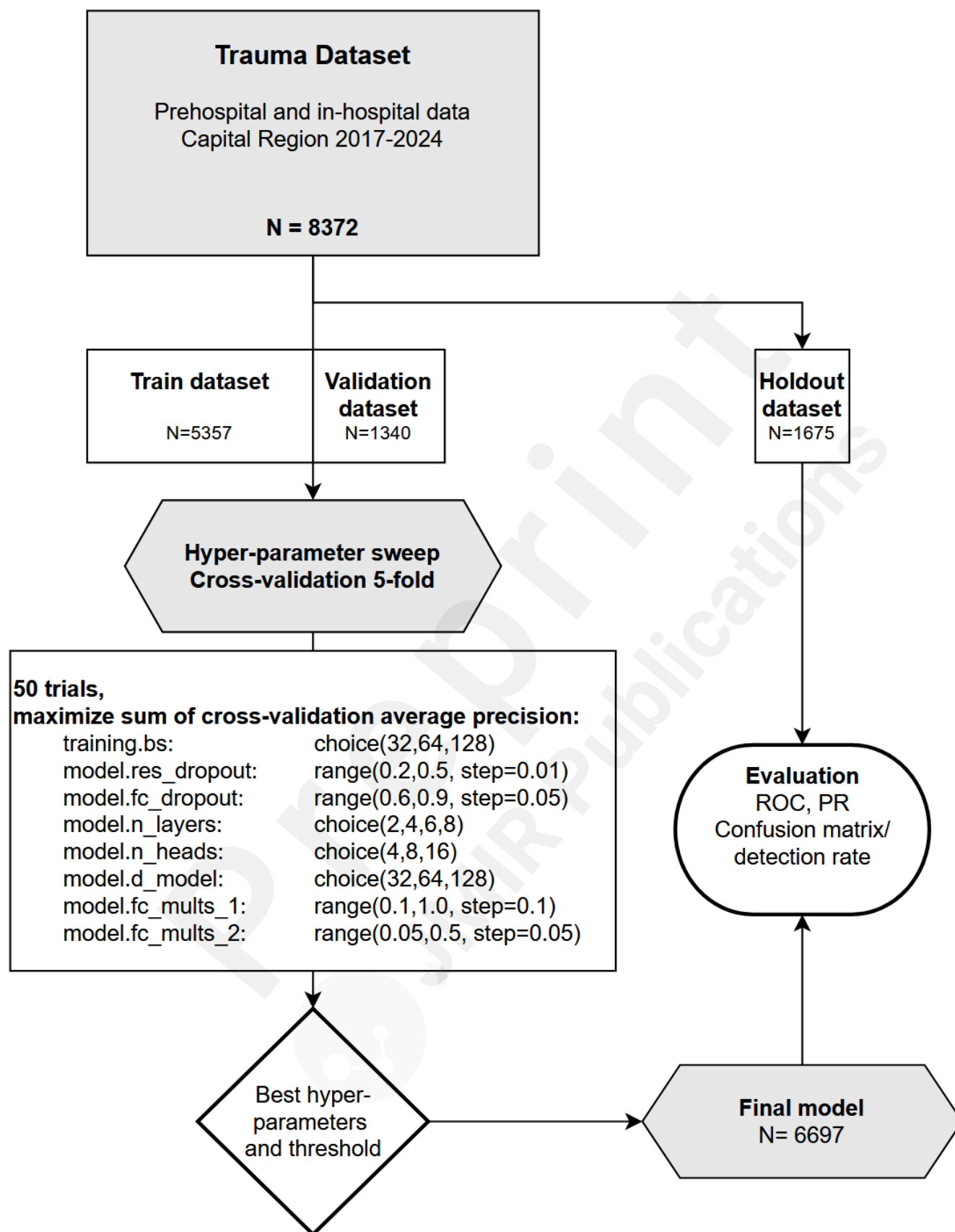


Figure 2. The overall training process displaying how optimal hyper-parameters and thresholds were determined through hyper-parameter sweeping over 50 trials each of 5-fold cross-validation optimizing for the sum of area under precision-recall curve for the internal folds and finally evaluating performance on the holdout dataset using receiver operator characteristics (ROC), precision and recall curve (PR) and confusion matrices. Each trial optimized: Batch size (bs),

residual dropout fraction (res_dropout), the fully connected neural network head dropout fraction (fc_dropout), number of transformer encoder layers (n_layers), number of attention heads (n_heads), model dimension (d_model) and fully connected neural network hidden layer multiplier (fc_mults_1 and fc_mults_2).

Model architecture

Our model utilized a combination of static contextual features and sequential input data. We employed the timeseries-tabular-transformer¹⁸, an adaptation of the TabTransformer¹⁹ architecture designed for time-series and tabular data fusion. This hybrid neural network approach capitalizes on the attention mechanism's strengths for processing sequential data while incorporating tabular patient information to contextualize the representation. The model architecture is elaborated in Appendix S1. Implementations are listed in Appendix S2.

Performance evaluation metrics

To evaluate performance, we employed receiver operating characteristic (ROC) curve and a precision-recall (PR) curve. The ROC-curve was summarized with area under the curve (AUROC), and for PR we calculated the area under precision-recall curve (AUPRC). Confidence intervals were calculated by DeLong method for AUROC and bootstrapping for AUPRC. Dynamic risk prediction performance was evaluated by masking sequences continuously at each step, equivalent to binning frequencies, to plot AUROC and AUPRC as a function of time. Moreover, we selected the top 5, 10, 15 and 20 and 25 percentage of predictions at each timestep, marked them as positive, calculated the sensitivity/recall and plotted with confidence intervals over time. This demonstrates the ability to capture positive cases with different tolerances to false positives, that may differ according to use-case scenarios.

To further assess the reliability of the model's probability estimates, we created calibration plots and calculated the Brier score²⁰ using full length sequences. Furthermore, we optimized classification thresholds using the F-beta score, allowing us to balance precision and recall according to the specific applications of our model. We applied beta=5 for an analysis prioritizing sensitivity and beta=1 for the harmonic mean between the two metrics. These optimized thresholds were then used to construct confusion matrices, which were used for calculating sensitivity, specificity and positive predictive value at a specific threshold.

RESULTS

The study included 8,372 trauma cases with a mean patient age of 43 years (SD: 22.2); 66.4% were male. The mean Elixhauser Comorbidity Index was 1.0 (SD: 3.7), Injury Severity Score mean was 3.2 (SD: 7.3), and Glasgow Coma Score was 11.1 (SD: 4.4). The 90-day mortality rate was 4.55%. Sequential data encompassed 6,759,517 vital sign measurements, 499,266 medication administrations, and 639,707 lab test results over a mean patient trajectory of 8.0 (SD: 24.3) days. Patient characteristics are available in Table 1. The training process used 6,697 (80%) trauma cases, and 1,675 cases (20%) were held-out for independent validation.

				Grouped by 90-day mortality		
		Missin g	Overall	0	1	P- Value
n			8372	7991	381	
Age, mean (SD)		0	42.6 (22.2)	41.4 (21.6)	69.5 (18.0)	<0.001
Sex, n (%)	Female	0	2811 (33.6)	2688 (33.6)	123 (32.3)	0.623
	Male		5561 (66.4)	5303 (66.4)	258 (67.7)	
Trajectory duration in days, mean (SD)		0	8.0 (24.3)	8.0 (24.8)	8.4 (67.7)	0.557
Elixhauser Comorbidity Index, mean (SD)		2063	1.0 (3.7)	0.8 (3.4)	4.6 (6.0)	<0.001
Injury Severity Score, mean (SD)		145	3.2 (7.3)	3.0 (7.0)	8.4 (11.3)	<0.001
Glasgow Coma Score, mean (SD)		975	12.9 (3.9)	13.3 (3.6)	6.4 (4.6)	<0.001
Heart rate (agg. mean), mean (SD)		0	83.9 (15.0)	83.7 (14.9)	87.5 (16.8)	<0.001
Respiratory rate (agg. mean), median [Q1,Q3]		1572	16.5 [15.6,18.0]	16.4 [15.6,18.0]	17.4 [15.0,19.5]	<0.001
Systolic blood pressure (agg. mean), mean (SD)		0	127.9 (17.9)	128.1 (17.7)	125.2 (21.9)	0.012
Oxygen saturation (agg. mean), median [Q1,Q3]		1	97.7 [96.5,98.6]	97.8 [96.6,98.7]	95.8 [94.1,97.5]	<0.001
Temperature (agg. mean), median [Q1,Q3]		2168	37.0 [36.7,37.2]	37.0 [36.7,37.2]	37.0 [36.5,37.4]	0.322
Anastethics (count of prescriptions), mean [min,max]		0	3.3 [0.0,199.0]	3.2 [0.0,199.0]	6.2 [0.0,70.0]	<0.001
Antibiotics (count of prescriptions), mean [min,max]		0	11.1 [0.0,910.0]	10.6 [0.0,910.0]	21.3 [0.0,329.0]	<0.001

Anti-thrombotic agents (count of prescriptions), mean [min,max]		0	5.0 [0.0,577.0]	5.0 [0.0,577.0]	6.0 [0.0,202.0]	0.199
Blood products, mean [min,max]		0	0.1 [0.0,29.0]	0.1 [0.0,21.0]	0.7 [0.0,29.0]	<0.001
Cardiovascular drugs (count of prescriptions), mean [min,max]		0	4.7 [0.0,837.0]	4.5 [0.0,837.0]	9.8 [0.0,194.0]	<0.001
Diuretics (count of prescriptions), mean [min,max]		0	2.0 [0.0,492.0]	1.7 [0.0,492.0]	7.9 [0.0,315.0]	<0.001
Hemostatics (count of prescriptions), mean [min,max]		0	0.3 [0.0,83.0]	0.3 [0.0,83.0]	1.1 [0.0,22.0]	<0.001
Infusion (count of prescriptions), mean [min,max]		0	3.6 [0.0,476.0]	3.3 [0.0,476.0]	10.4 [0.0,204.0]	<0.001
Insulin (count of prescriptions), mean [min,max]		0	1.4 [0.0,604.0]	1.1 [0.0,604.0]	5.9 [0.0,236.0]	<0.001
Local anesthetics (count of prescriptions), mean [min,max]		0	0.6 [0.0,57.0]	0.6 [0.0,57.0]	0.5 [0.0,20.0]	0.728
Neurological drugs (count of prescriptions), mean [min,max]		0	10.2 [0.0,1009.0]	10.2 [0.0,1009.0]	9.5 [0.0,326.0]	0.587
Opioids (count of prescriptions), mean [min,max]		0	11.6 [0.0,574.0]	11.6 [0.0,574.0]	12.8 [0.0,243.0]	0.356
Base Excess, mean (SD)		3725	1.2 (3.9)	1.2 (3.5)	1.5 (6.8)	0.442
Lactate, median [Q1,Q3]		2119	2.1 [1.5,3.1]	2.1 [1.5,3.0]	3.5 [2.2,7.4]	<0.001
Hemoglobin, mean (SD)		657	8.7 (1.1)	8.7 (1.0)	8.0 (1.4)	<0.001
Leukocytes count, median [Q1,Q3]		764	10.1 [7.6,13.9]	9.9 [7.5,13.5]	15.7 [11.1,20.7]	<0.001
TEG-Reaction time, median [Q1,Q3]		7367	5.3 [4.6,6.2]	5.3 [4.6,6.2]	5.8 [4.8,7.2]	0.001
TEG-Maximum amplitude, mean (SD)		7367	65.3 (6.7)	65.3 (6.5)	65.1 (8.3)	0.829

TEG-Fibrinolysis, median [Q1,Q3]		7468	0.2 [0.0,1.0]	0.3 [0.0,1.0]	0.1 [0.0,0.7]	0.031
A: Airway, n (%)	Blocked	703	25 (0.3)	6 (0.1)	19 (5.8)	<0.001
	Clear		7288 (95.0)	7070 (96.3)	218 (66.1)	
	Threatened		356 (4.6)	263 (3.6)	93 (28.2)	
B: Breathing, n (%)	Lightly affected	690	1039 (13.5)	957 (13.0)	82 (24.8)	<0.001
	Normal		6329 (82.4)	6185 (84.1)	144 (43.6)	
	Respiratory arrest (ceased)		62 (0.8)	23 (0.3)	39 (11.8)	
	Very affected		252 (3.3)	187 (2.5)	65 (19.7)	
C: Circulation, n (%)	Cardiac arrest	697	64 (0.8)	24 (0.3)	40 (12.2)	<0.001
	Lightly affected		1050 (13.7)	956 (13.0)	94 (28.6)	
	Normal		6314 (82.3)	6173 (84.0)	141 (42.9)	
	Very affected		247 (3.2)	193 (2.6)	54 (16.4)	
D: Disability, n (%)	Awake	709	6029 (78.7)	5911 (80.6)	118 (36.1)	<0.001
	Reduced awareness		1221 (15.9)	1134 (15.5)	87 (26.6)	
	Unconscious		413 (5.4)	291 (4.0)	122 (37.3)	

Table 1: Patient characteristics. TEG: Thrombelastography

Performance

Predicting 90-day mortality for the holdout dataset using full length sequences, the model achieved an area under the receiver operator characteristics curve (AUROC) of 0.963 (95% CI: 0.934-0.993) and an area under precision-recall curve (AP) of 0.800 (95% CI: 0.712-0.870). ROC and PR curves at selected trajectory timepoints is presented in Figure 3.

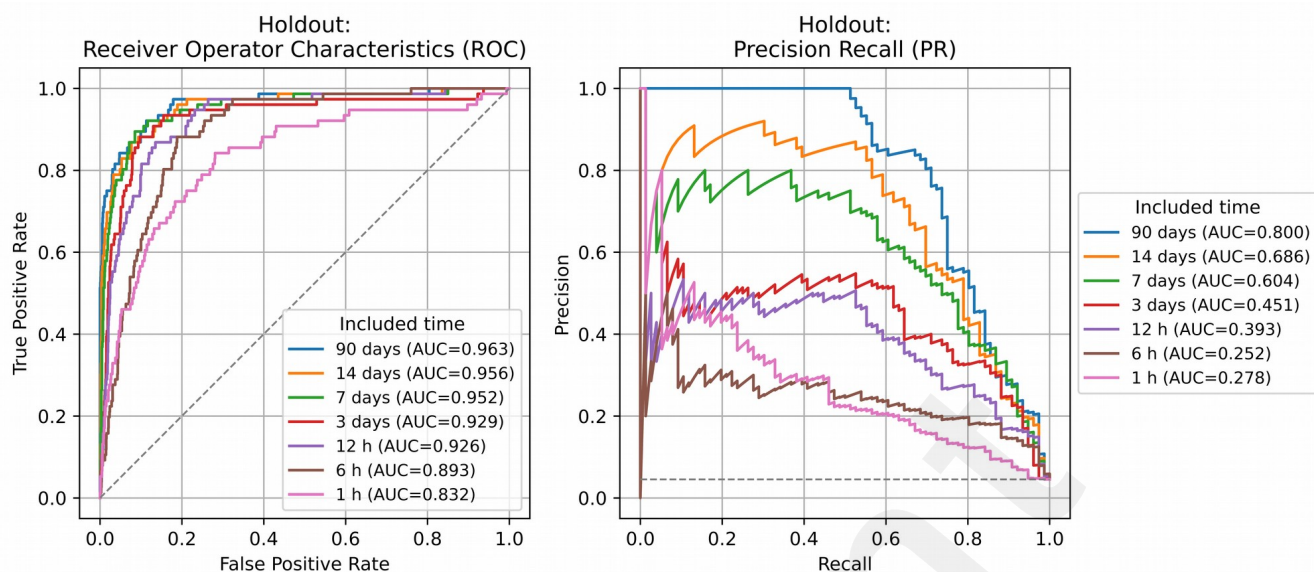


Figure 3. Receiver operator characteristics curve and precision recall curve for the holdout evaluation summarized by area under the curve (AUC) and area under precision-recall curve (AP).

The model showed good calibration on the holdout dataset with slight overconfidence in predictions with predicted probabilities consistently falling below the line, indicating perfect calibration (Figure 4). The Brier score was 0.0207.

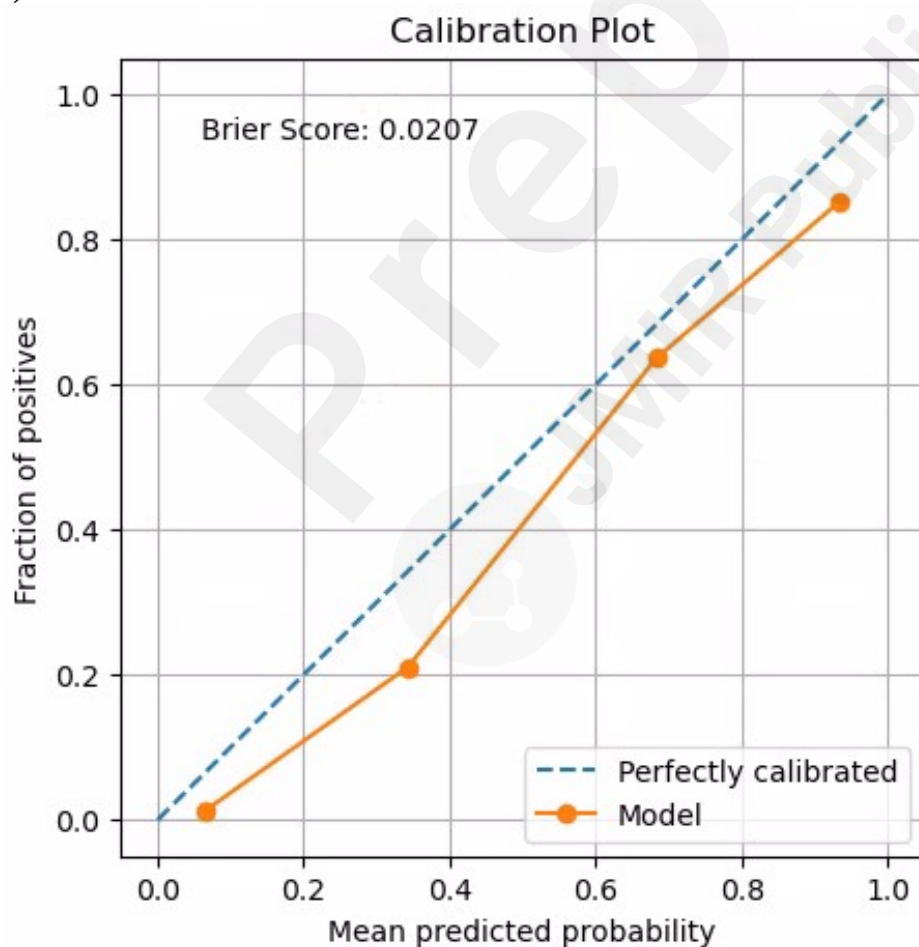


Figure 4. Calibration plot for the holdout dataset presented with Brier score.

The model performance varied depending on prediction time in the patient trajectory with AUROC interval 0.832-0.926 and AUPRC interval 0.214-0.393 in the first 12 hours, from 12 to 72 hours AUROC was 0.926-0.934 and AUPRC interval 0.381-0.457, from 72 hours to 7 days AUROC was 0.940-0.952 and AUPRC 0.463-0.604, from 7 days to 30 days AUROC was 0.953-0.962 and AUPRC 0.615-0.796, and stabilizing over 0.962 AUROC and over 0.796 AUPRC after 30 days (visualized in Figure 5, presented with CI's in Table S5).

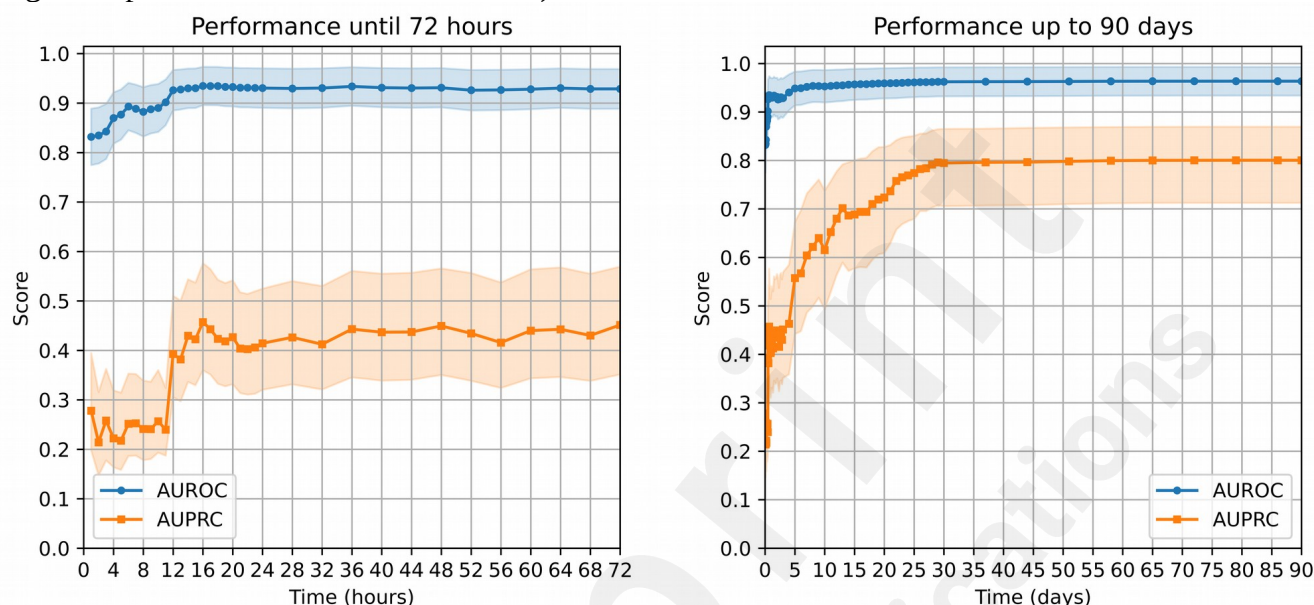


Figure 5. Area under the receiver operator characteristics curve (AUROC) and area under the precision recall curve (AUPRC) for each timestep included in prediction for the holdout dataset.

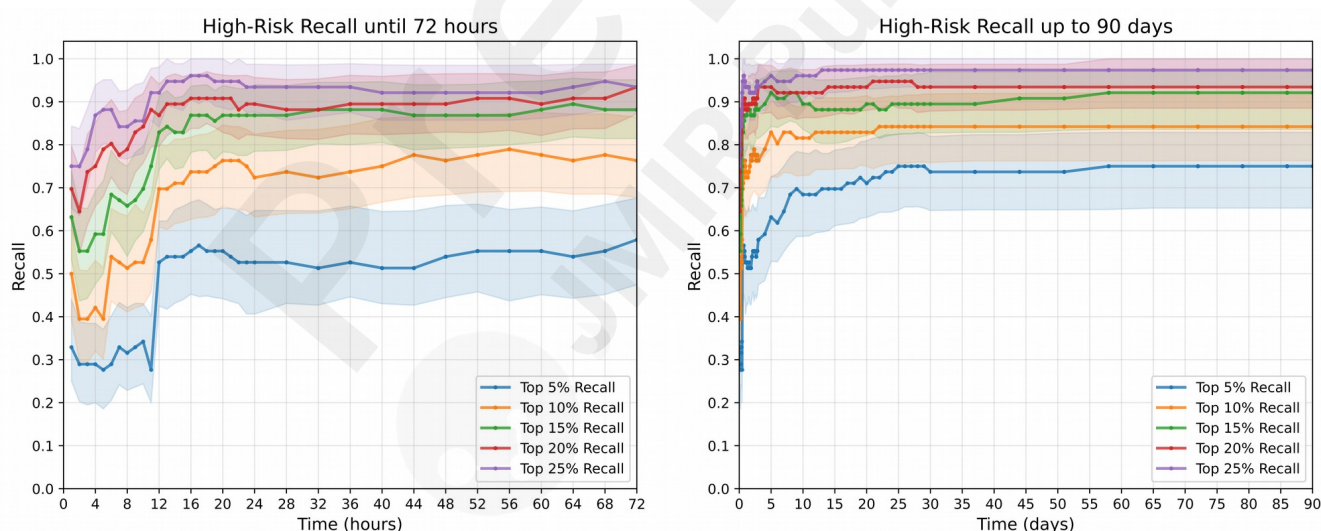


Figure 6. Recall for each timestep by predicting mortality for top 5, 10, 15, 20 and 25 percent of probabilities.

The model demonstrated sensitivity ranging from 0.28 to 0.90 within the first 12 hours with 5 to 25 top percent predictions marked as positives. Sensitivities across the trajectory for top 5 to 25 percent high-risk patients are presented in Figure 6.

F-beta optimized thresholds calculated from internal cross-validation with the final hyper-parameters, where using beta of 5 resulted in 84% sensitivity, 94% specificity (Figure S4) and a positive predictive value (PPV) of 48%. Using the harmonic mean of precision (beta = 1), the model achieves 74% sensitivity, 98% specificity and a PPV of 94% (Figure S5).



DISCUSSION

This study focused on constructing trauma patient trajectories automatically from available electronic health records (EHR) for dynamic risk prediction of 90-day mortality. We found that using a hybrid neural network for predicting 90-day mortality achieved an excellent area under the operator characteristics curve (AUROC) of 0.963 and an area under precision-recall curve (AP) of 0.800 on a holdout dataset using full length sequences. Model calibration analysis revealed probabilities consistently falling below the perfectly calibrated diagonal line, exhibiting overconfidence in predictions by systematically assigning higher probabilities to outcomes than warranted by observed frequencies. The Brier score for our binary classification analysis was 0.0207, which is substantially lower than the maximum possible score of 0.043 for a non-informative model with this class distribution. The dynamic evaluation of performance by masking sequences demonstrated varying predictive performance over time, increasing as trajectories progress or end. Since our modelling approach was treating mortality as a binary outcome rather than time-to-event, the interpretation of our dynamic evaluation should account for using all patients in all evaluations. Moreover, the patient trajectory duration followed a positive skewed distribution with most patients having less than 30 days duration (Figure S3).

For most use cases, the model output would present to clinicians as a score or an index with reference levels. In other cases, such as selecting cases for examination and evaluation in quality of care, a threshold may be determined. When optimizing for sensitivity using F-beta score (beta = 5) on the validation dataset, the model achieved 84% sensitivity while maintaining 94% specificity. In absolute numbers, this means identifying 64 of 76 deaths while wrongly labelling 94 cases. For live risk scoring in a clinical setting, e.g. as deterioration detection, it is important to avoid alarm fatigue. Using the F1-score to optimize the threshold for positive predictive value (PPV), the model achieved 94% PPV. In absolute numbers, we identified 56 of 76 deaths with only 25 false positives. In our dynamic sensitivity analysis, we demonstrated that selecting high-risk predictions at various thresholds the model can achieve a high sensitivity, if prioritized. Tolerance to false positives and necessity of high sensitivity may vary largely depending on use-case and timepoint.

Our model's performance compares with recent advancements in trauma mortality prediction. The Parkland Trauma Index of Mortality (PTIM)²¹ is a machine learning algorithm predicting 48-hour mortality during the first 72 hours of hospitalization for polytrauma patients, reporting an AUROC of 0.94 for 516 cases. PTIM utilizes 23 variables, including vital signs, laboratory values, clinical scores, and demographics. One notable advantage of the PTIM the integration into the EHR system, allowing for real-time, hourly updates of mortality predictions. This feature enhances its clinical utility, as demonstrated in a survey-based study where 77% of providers reported that the PTIM assisted in their treatment decisions²².

Moreover, Epic Deterioration Index (EDI)²³ also demonstrated impressive results, with an AUROC of 0.98 for predicting in-hospital mortality within 24 hours of admission for 1325 level 1 trauma centre patients. Both the PTIM study and the EDI validation study operate on shorter timeframes compared to our study, which focuses on 90-day mortality prediction. Moreover, our study included a larger and more heterogeneous population, whereas PTIM only included polytrauma patients, and the EDI validation included only level 1 trauma centre patients. Particularly, these differences limit direct performance comparisons but demonstrate the feasibility of machine learning-based mortality prediction models in trauma care.

Traditionally, risk estimation often required manual data entry into separate interfaces, with information gathered from the EHR and then entered into a distinct system for risk calculation, such as the Predictive Optimal Trees in Emergency Surgery Risk calculator²⁴. In our approach, we have concentrated on input features that can be automatically collected, ensuring that risk predictions can be generated with minimal manual effort and made readily available. While difficult, this strategy

reduces the burden on healthcare providers and enables real-time risk assessment and reduces the potential for human error²⁵.

As a limitation to our study, the population consists of patients in a single region in eastern Denmark, thus not fully representing the entire demographical spectrum of trauma patients. Moreover, the data might reflect local practices and guidelines, such as in the pre-defined categories for ABCD assessment. Despite these specific categories being implemented and used nationally, there may be regional differences in how clinicians are instructed to conduct subjective assessments. If our model is to be used in other countries or regions, retraining with local data is highly advisable^{9,26}. Further studies should investigate whether similar modelling approaches yield comparable predictive performance when using different structures of the same information. This research should focus on varied data representations (e.g., structured vs. unstructured data), alternative categorization schemes, and alternative encoding methods.

While our model demonstrates strong predictive performance, there are several avenues for potential improvement and future research. This study focused on 90-day mortality, where predicting additional outcomes such as post-trauma complications, readmissions, or functional status could provide a more comprehensive risk assessment. Another application of this modelling approach was demonstrated by our research group using prehospital EHR to predict surgical needs stratified into specialties²⁷. Adapting the model to predict several outcomes relevant at different decision-making branching points could significantly impact patient care from the earliest stages of trauma management to hospital discharge.

Incorporating additional data types could also enhance the model's predictive power. However, careful consideration must be given to avoid data leakage, particularly when including, e.g. diagnoses and procedures that might be directly related to the outcome. Future studies should explore integrating such information while maintaining the integrity of the prediction task. Another area for improvement lies in exploring different data representations and aggregations. For instance, medications and laboratory results could be grouped or aggregated in various ways to capture more nuanced patterns. These alternative representations might reveal additional insights and improve the model's performance. Research into optimal grouping strategies for different data types could yield significant advancements in predictive accuracy.

Next-phase evaluation would assess the effect of our model in the clinic by randomized control trials prescribing our model, using well-established quality indicators as outcomes such as used in the Trauma Quality Improvement Program (TQIP) by the American College of Surgeons (ACS)²⁸. This would quantify clinical benefits while identifying implementation barriers and care gaps. Our study demonstrates the feasibility and potential of constructing patient trajectories from electronic health records for dynamic risk prediction, showcased by the model's strong performance in predicting 90-day mortality. While there are clear areas for future improvement and expansion, this approach advances towards more data-driven, personalized trauma care. As we continue to refine and expand these models, their integration into clinical practice could support trauma care from prehospital triage to long-term management. While these tools enhance clinical judgment rather than replace it, their integration could improve trauma outcomes through timely, insight-driven decision support.

ETHICS STATEMENT

The study was approved by the Danish Patient Safety Board (Styrelsen for Patientsikkerhed, approval #31-1521), and the Danish Capital Regions Data Safety Board (Videncenter for Dataanmeldelser, approval #P-2020-180). In compliance with Danish legislation on healthcare

research, patient consent was not required due to the retrospective and de-identified nature of the dataset and thus not obtained.

AUTHOR CONTRIBUTIONS

AM, MS, HK, IC, SR AND FF conceived the study. AM and IC performed data extraction, setup and analysis. AM drafted the manuscript. Remaining authors performed critical revisions. All authors contributed to the article and approved the submitted version.

FUNDING

The study was supported by grants from the Danish Victims foundation (Grant #21-610-00127) and the Novo Nordisk foundation (Grant #NNF19OC0055183) to MS.

CONFLICTS OF INTEREST

Author MS has co-founded Aiomic, a company developing AI models for the healthcare sector. The present work does not relate to any commercial activities and is for research only.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

Due to confidentiality issues, we are not at liberty to share the Electronic Health Record (pre- and in-hospital) data used in this study. Access to Danish health data is granted through The Danish Patient Safety Authority. EHR data can then be obtained by reasonable request through the relevant region.

SOURCE CODE AVAILABILITY STATEMENT

Project source code is available at the following GitHub repository²⁹

ABBREVIATIONS

Abbreviation	Full Term
AI	Artificial Intelligence
EHR	Electronic Health Record
AUROC	Area Under the Receiver Operating Characteristic Curve
AUPRC	Area Under the Precision-Recall Curve
ROC	Receiver Operating Characteristic
PR	Precision-Recall
PPV	Positive Predictive Value
CI	Confidence Interval
TRISS	Trauma and Injury Severity Score
RTS	Revised Trauma Score
ISS	Injury Severity Score
GCS	Glasgow Coma Scale
ABCD	Airway, Breathing,

	Circulation, Disability
ML	Machine Learning
ICD	International Classification of Diseases
TEG	Thromboelastogram
ATC	Anatomical Therapeutic Chemical (classification system)
R	Statistical programming language R
TQIP	Trauma Quality Improvement Program
ACS	American College of Surgeons
PTIM	Parkland Trauma Index of Mortality
EDI	Epic Deterioration Index
SD	Standard Deviation

REFERENCES

- 1 Salim A, Stein DM, Zarzaur BL, Livingston DH. Measuring long-term outcomes after injury: current issues and future directions. *Trauma Surg Acute Care Open* 2023; **8**. DOI:10.1136/tsaco-2022-001068.
- 2 Lapp L, Roper M, Kavanagh K, Bouamrane M-M, Schraag S. Dynamic Prediction of Patient Outcomes in the Intensive Care Unit: A Scoping Review of the State-of-the-Art. *J Intensive Care Med* 2023; **38**: 575–91.
- 3 Preti LM, Ardito V, Compagni A, Petracca F, Cappellaro G. Implementation of Machine Learning Applications in Health Care Organizations: Systematic Review of Empirical Studies. *J Med Internet Res* 2024; **26**: e55897.
- 4 Boyd CR, Tolson MA, Copes WS. Evaluating Trauma Care: The TRISS Method. *Journal of Trauma and Acute Care Surgery* 1987; **27**: 370.
- 5 Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A Revision of the Trauma Score. *Journal of Trauma and Acute Care Surgery* 1989; **29**: 623.
- 6 Baker SP, O'Neill B, Haddon WJ, Long WB. THE INJURY SEVERITY SCORE: A METHOD FOR DESCRIBING PATIENTS WITH MULTIPLE INJURIES AND EVALUATING EMERGENCY CARE. *Journal of Trauma and Acute Care Surgery* 1974; **14**: 187.
- 7 Teasdale G, Jennett B. ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS: A Practical Scale. *The Lancet* 1974; **304**: 81–4.
- 8 Maurer LR, Bertsimas D, Bouardi HT, et al. Trauma outcome predictor: An artificial intelligence interactive smartphone tool to predict outcomes in trauma patients. *J Trauma Acute Care Surg* 2021; **91**: 93–9.

- 9 Millarch AS, Bonde A, Bonde M, *et al.* Assessing optimal methods for transferring machine learning models to low-volume and imbalanced clinical datasets: experiences from predicting outcomes of Danish trauma patients. *Frontiers in Digital Health* 2023; **5**. <https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1249258> (accessed Feb 21, 2024).
- 10 Bonde A, Bonde M, Troelsen A, Sillesen M. Assessing the utility of a sliding-windows deep neural network approach for risk prediction of trauma patients. *Sci Rep* 2023; **13**: 5176.
- 11 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is All you Need. .
- 12 Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998; **36**: 8–27.
- 13 Gasparini A. comorbidity: An R package for computing comorbidity scores. *Journal of Open Source Software* 2018; **3**: 648.
- 14 Black A, Clark D. icdpicr: ‘ICD’ Programs for Injury Categorization in R. 2022; published online March 12. <https://cran.r-project.org/web/packages/icdpicr/index.html> (accessed Dec 16, 2024).
- 15 Eskesen TO, Sillesen M, Rasmussen LS, Steinmetz J. Agreement Between Standard and ICD-10-Based Injury Severity Scores. *Clin Epidemiol* 2022; **14**: 201–10.
- 16 Thim T, Krarup NHV, Grove EL, Rohde CV, Løfgren B. Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012; **5**: 117–21.
- 17 Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society Series B (Methodological)* 1974; **36**: 111–47.
- 18 Oguiza I. tsai - A state-of-the-art deep learning library for time series and sequential data. 2023. <https://github.com/timeseriesAI/tsai>.
- 19 Huang X, Khetan A, Cvitkovic M, Karnin Z. TabTransformer: Tabular Data Modeling Using Contextual Embeddings. 2020; published online Dec 11. DOI:10.48550/arXiv.2012.06678.
- 20 Brier GW. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. *Monthly Weather Review* 1950; **78**: 1–3.
- 21 Starr AJ, Julka M, Nethi A, *et al.* Parkland Trauma Index of Mortality: Real-Time Predictive Model for Trauma Patients. *Journal of Orthopaedic Trauma* 2022; **36**: 280.
- 22 Park C, Loza-Avalos SE, Harvey J, *et al.* A Real-Time Automated Machine Learning Algorithm for Predicting Mortality in Trauma Patients: Survey Says it’s Ready for Prime-Time. *Am Surg* 2024; **90**: 655–61.
- 23 Mou Z, Godat LN, El-Kareh R, Berndtson AE, Doucet JJ, Costantini TW. Electronic health record machine learning model predicts trauma inpatient mortality in real time: A validation study. *J Trauma Acute Care Surg* 2022; **92**: 74–80.
- 24 Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical Risk Is Not Linear: Derivation and Validation of a Novel, User-friendly, and Machine-learning-based Predictive OpTimal Trees in

- Emergency Surgery Risk (POTTER) Calculator. *Ann Surg* 2018; **268**: 574–83.
- 25 Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. *Annals of the Rheumatic Diseases* 2023; **82**: 306–11.
- 26 Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *npj Digit Med* 2022; **5**: 1–8.
- 27 Millarch AS, Folke F, Rudolph SS, Kaafarani HM, Sillesen M. Prehospital triage of trauma patients: predicting major surgery using artificial intelligence as decision support. *British Journal of Surgery* 2025; **112**: znaf058.
- 28 Nathens AB, Cryer HG, Fildes J. The American College of Surgeons Trauma Quality Improvement Program. *Surg Clin North Am* 2012; **92**: 441–54, x–xi.
- 29 Millarch A. Project source code. 2025; Available from: <https://github.com/a-millarch/dynamic-risk-prediction>.

Preprint
JMIR Publications

Preprint
JMIR Publications

Supplementary Files

Untitled.

URL: <http://asset.jmir.pub/assets/6574a2b0716b449cf28c38c8fa574a4b.pdf>

Multimedia Appendixes

Supplementary data.

URL: <http://asset.jmir.pub/assets/669aba713d4c3cde56bd79c05816267b.pdf>