

Research and Applications

1 A deep learning model for clinical outcome 2 prediction using longitudinal inpatient electronic 3 health records

4 Ruichen Rong, PhD^{1,#}, Zifan Gu, MS^{1,#}, Hongyin Lai, MS¹, Tanna L. Nelson, RN, PhD²,
5 Tony Keller², Clark Walker², Kevin W. Jin¹, Catherine Chen, MD³, Ann Marie Navar,
6 MD, PhD³, Ferdinand Velasco, MD², Eric D. Peterson, MD, MPH³, Guanghua Xiao,
7 PhD^{1,4,5}, Donghan M. Yang, PhD^{1*}, Yang Xie, PhD^{1,4,5*}

8
9 ¹ Quantitative Biomedical Research Center, Peter O'Donnell Jr. School of Public Health,
10 The University of Texas Southwestern Medical Center, Dallas, Texas, 75390, USA

11 ² Texas Health Resources, Arlington, Texas

12 ³ Department of Internal Medicine, The University of Texas Southwestern Medical
13 Center, Dallas, TX 75390, USA

14 ⁴ Department of Bioinformatics, The University of Texas Southwestern Medical Center,
15 Dallas, Texas, 75390, USA

16 ⁵ Simmons Comprehensive Cancer Center, The University of Texas Southwestern
17 Medical Center, Dallas, Texas, 75390, USA

18
19 # Equal contribution as first authors.

20 *Co-corresponding authors:

21 Donghan M. Yang, PhD

22 The University of Texas Southwestern Medical Center

23 Danciger Research Building

24 5323 Harry Hines Blvd. Ste H9.124

25 Dallas, TX 75390-8821

26 Phone: 214-648-8920

27 donghan.yang@utsouthwestern.edu

28

29 Yang Xie, PhD

30 The University of Texas Southwestern Medical Center

31 Danciger Research Building

32 5323 Harry Hines Blvd. Ste H9.124

33 Dallas, TX 75390-8821

34 Phone: 214-648-4003

35 yang.xie@utsouthwestern.edu

36

37

Research and Applications

38 **Keywords**

39 EHR, continuous monitoring data, Transformer, ICU, COVID-19

40

41 Word count: 3994

42 Structured abstract: 242

43 Tables: 2

44 Figures: 3

Research and Applications

45 **ABSTRACT**

46 **Objective**

47 Recent advances in deep learning show significant potential in analyzing continuous
48 monitoring electronic health records (EHR) data for clinical outcome prediction. We aim
49 to develop a Transformer-based, Encounter-level Clinical Outcome (TECO) model to
50 predict mortality in the intensive care unit (ICU) using inpatient EHR data.

51 **Materials and Methods**

52 TECO was developed using multiple baseline and time-dependent clinical variables
53 from 2579 hospitalized COVID-19 patients to predict ICU mortality, and was validated
54 externally in an ARDS cohort (n=2799) and a sepsis cohort (n=6622) from the Medical
55 Information Mart for Intensive Care (MIMIC)-IV. Model performance was evaluated
56 based on area under the receiver operating characteristic (AUC) and compared with
57 Epic Deterioration Index (EDI), random forest (RF), and extreme gradient boosting
58 (XGBoost).

59 **Results**

60 In the COVID-19 development dataset, TECO achieved higher AUC (0.89–0.97) across
61 various time intervals compared to EDI (0.86–0.95), RF (0.87–0.96), and XGBoost
62 (0.88–0.96). In the two MIMIC testing datasets (EDI not available), TECO yielded higher
63 AUC (0.65–0.76) than RF (0.57–0.73) and XGBoost (0.57–0.73). In addition, TECO was
64 able to identify clinically interpretable features that were correlated with the outcome.

65 **Discussion**

66 TECO outperformed proprietary metrics and conventional machine learning models in
67 predicting ICU mortality among COVID-19 and non-COVID-19 patients.

Research and Applications

68 **Conclusions**

69 TECO demonstrates a strong capability for predicting ICU mortality using continuous
70 monitoring data. While further validation is needed, TECO has the potential to serve as
71 a powerful early warning tool across various diseases in inpatient settings.

72

Research and Applications

73 **LAY SUMMARY**

74 In intensive care units (ICUs), accurately estimating the risk of death is crucial for timely
75 and effective medical intervention. This study developed a new AI algorithm, TECO
76 (Transformer-based, Encounter-level Clinical Outcome model), which uses electronic
77 health records to continuously predict ICU mortality after admission, with the capability
78 to update predictions on an hourly basis. TECO was trained on data from over 2,500
79 COVID-19 patients and was designed to analyze multiple types of continuous
80 monitoring data collected during a patient's ICU stay. We tested TECO's performance
81 against a widely used proprietary tool, the Epic Deterioration Index (EDI), and other
82 machine learning methods, such as random forest and XGBoost, across three patient
83 groups: COVID-19, ARDS (acute respiratory distress syndrome), and sepsis. TECO
84 consistently showed better performance and was able to predict death risk earlier than
85 other methods. Additionally, TECO identified key health indicators associated with ICU
86 mortality, making its predictions more interpretable for clinicians. These findings suggest
87 that TECO could become a valuable early warning tool, helping doctors monitor patients'
88 health and take timely action in a range of critical care situations.

89

Research and Applications

90 **BACKGROUND AND SIGNIFICANCE**

91 Modern medical and information technologies increasingly produce massive amounts of
92 electronic health record (EHR) data. However, there still exists a gap between the rapid
93 digitization of healthcare and the development of analytic tools capable of informing and
94 guiding real-world clinical practices.[1, 2] Intensive care unit (ICU) is one area ripe for
95 improved predictive analytics.[3] The COVID-19 pandemic posed unprecedented
96 challenges to the provision of ICU care due to a lack of bed capacity, clinical care staff,
97 and necessary analytics for resource allocation.[4] While the contemporary ICU medical
98 devices and systems collect vast amounts of time-stamped data (e.g., vital signs, lab
99 tests, and medication administrations), these rich continuous data streams are often not
100 used to their full extent to develop analytics tools that assist clinicians to predict clinical
101 outcomes.

102 One reason for this failure has been the analytics applied to such data. Conventional
103 statistical models are limited in processing multivariate, time-dependent datasets and
104 analyzing relations between different variables and different timestamps. Commercial
105 analytics tools often lack technical transparency and interoperability across EHR
106 platforms. For instance, the Epic Deterioration Index (EDI)[5, 6], a proprietary machine
107 learning-based metric only available on Epic systems, was designed to quantify the
108 level of deterioration patients experience at a point in time. The EDI utilizes data such
109 as age, vital signs, laboratory tests, yet without considering patients' overall
110 comorbidities.[7, 8] The detailed design and parameters of EDI's model have not been
111 publicized, and EDI has not been validated in routine clinical practice.[8] Moreover, the

Research and Applications

112 threshold of EDI that calls for intervention has been differentially implemented among
113 healthcare systems, rendering its actual usage dependent on local standards.[8]
114 As an alternative, machine learning models have been proposed to predict clinical
115 outcomes or procedures in critically ill patients. For example, researchers employed a
116 random forest (RF) classifier to predict COVID-19 disease severity at hospital
117 admission,[9] applied an extreme gradient boosting (XGBoost) based algorithm to
118 predict invasive mechanical ventilation,[10] and implemented PICTURE (Predicting
119 Intensive Care Transfers and Other Unforeseen Events) to predict deterioration.[11]
120 However, these algorithms face limitations when dealing with long series of time-
121 dependent data with high dimensionality and irregular time intervals, as commonly
122 encountered in inpatient monitoring data.

123 Recent advancements in deep learning have demonstrated preliminary success in
124 managing multi-dimensional sequential EHR data, effectively capturing complex
125 temporal patterns and interrelated features. Among these advancements, transformer
126 models have shown promise by leveraging attention-based mechanisms to enhance
127 both performance and efficiency in clinical settings.[12] Transformers employ multi-head
128 attention and skip connections, removing the recurrent dependencies characteristic of
129 earlier architectures like recurrent neural networks (RNNs) and long short-term memory
130 networks (LSTMs),[13] while using positional embeddings to encode temporal
131 information directly. Transformer-based architectures have since been adapted for
132 various clinical applications. For example, Wu et al. developed a transformer model to
133 forecast influenza prevalence from public health data, [14] while ClinicalBERT predicts
134 30-day hospital readmissions from clinical notes.[15] Models such as BEHRT [16] and

Research and Applications

135 Med-BERT [17], pre-trained on sequences of diagnosis codes, have been used to
136 predict future diagnoses, with Antikainen et al. expanding this approach to incorporate
137 different types of medical events for long-term mortality prediction in cardiovascular
138 patients. [18]

139 In an ICU setting, MeTra integrates chest radiographs with clinical data for mortality
140 prediction, though it restricts data input to the first 48 hours and may pose a high
141 computational burden due to the Vision Transformer component.[19] Similarly, Cheng et
142 al. used transformers for image data to predict COVID-19 mortality, but did not apply
143 transformers to the clinical data component of their model.[20] Song et al. proposed the
144 SAnD architecture for ICU tasks, including mortality prediction; however, they restricted
145 data input to the final 24 hours of the ICU stay, requiring prior knowledge of the
146 outcome time, which limits the model's applicability to real-world scenarios.[21]

147 Furthermore, none of these ICU mortality prediction models have been benchmarked
148 against the widely-used commercial tool EDI, which limits the assessment of their
149 clinical utility and relevance as a trans-platform tool.

150 **OBJECTIVE**

151 In this study, we propose the Transformer-based Encounter-level Clinical Outcome
152 (TECO) model, which fully utilizes continuous ICU monitoring data alongside patient-
153 level baseline characteristics for mortality prediction after ICU admission, with the
154 capability to update predictions on an hourly basis. We developed TECO using EHR
155 data from a cohort of COVID-19 patients and validated the model on two external non-
156 COVID-19 cohorts. We benchmarked TECO against the EDI, RF, and XGBoost to
157 evaluate its performance in an ICU setting across different disease profiles.

Research and Applications

158

159 **MATERIALS AND METHODS**

160 **Study setting and design**

161 In this study, the model development dataset contained EHR data from Texas Health
162 Resources (THR), a large faith-based, nonprofit health system in North Texas,
163 operating 20 acute care hospitals and serving 7 million residents in 16 counties. The
164 THR cohort included 2579 adult patients with laboratory-confirmed COVID-19 (age \geq
165 18), who were admitted into ICUs during their first COVID-19 hospitalization. Dates of
166 patient hospitalization ranged from March 3, 2020 to August 13, 2021. Patients who had
167 more than one ICU admission during the hospitalization were excluded.

168 The external validation dataset was extracted from the Medical Information Mart for
169 Intensive Care (MIMIC-IV), a large, publicly available, de-identified clinical database for
170 critically ill patients admitted to the emergency department of the Beth Israel Deaconess
171 Medical Center in Boston, MA from 2008 to 2019.[22] We identified two patient cohorts
172 admitted to ICU, one diagnosed with acute respiratory distress syndrome (ARDS) and
173 the other with sepsis. If a patient had multiple ICU or hospital admissions, only the first
174 ICU visit of the first hospital admission was included. ARDS was defined in accordance
175 to the Berlin definition[23] with the MIMIC-specific positive identification method.[24]
176 Sepsis was defined in accordance with the Third International Consensus Definitions for
177 Sepsis and Septic Shock,[25] quantified by the Sequential Organ Failure Assessment
178 (SOFA) score.[26, 27] Detailed definitions and query methods for identifying these two
179 cohorts in the MIMIC-IV database are provided in the Supplementary Materials.

Research and Applications

180 In this study, the models were designed to predict a binary outcome: death versus non-
181 death in the ICU. In this context, the non-death outcome refers to conditions not at
182 immediate risk of death and includes a range of states, from requiring continued ICU
183 care to being ready for ICU discharge. We included two types of variables in the model:
184 baseline variables and time-dependent variables. Baseline variables include age at
185 hospital admission, sex, ethnicity, and race. Time-dependent variables include ICU
186 monitoring measures, each recorded at a different and irregular pace: body temperature,
187 respiration rate, pulse oximetry (SpO₂), mSOFA (modified Sequential Organ Failure
188 Assessment) overall score,[28] mSOFA respiratory sub-score, and SF ratio (SpO₂/FiO₂,
189 where FiO₂ is the fraction of inspired oxygen). Body mass index (BMI) was included as
190 a time-dependent variable in the COVID-19 dataset. In the MIMIC-IV dataset, where
191 time-dependent BMI was unavailable, the BMI recorded at hospital admission was used
192 as a baseline variable.

193 The institutional review boards at THR and UT Southwestern Medical Center approved
194 this study (Protocol #STU-2020-0786; activated on 8/24/2020). All patient identifiers
195 were removed before EHR data extraction.

196 **Development of TECO Model**

197 The TECO model employs a Transformer-encoder architecture. The overall algorithm
198 design and data processing are illustrated in Figure 1. First, we aligned the time-
199 dependent variables by taking the mean of each variable in every 15-minute interval. If
200 a 15-minute mean value was missing, the value from the previous interval was carried
201 forward. Then, we concatenated the aligned time-dependent variables with baseline
202 variables, creating a feature set for each 15-minute interval, and embedded these

Research and Applications

203 features into a 512-dimension vector. The baseline variable input values remained static
204 across the entire time range. The timestamps of these 15-minute intervals were
205 encoded into another representative vector by positional encoding. We applied relative
206 positional encoding to avoid imputing large amounts of default values for time points
207 without measurement. The variable value vector and timestamp vector were fed into a
208 feed-forward network with 6 Transformer-encoder layers. Each Transformer-encoder
209 layer was equipped with 8 multi-head attention modules. The model outputs probability
210 of death using a linear classification layer with the Softmax activation function.

211 We developed the TECO model using the THR COVID-19 dataset. We used the holdout
212 method by creating 20 different data splits, where in each split the COVID-19 dataset
213 was randomly split into a training set (80%) and a validation set (20%) on the patient
214 level. During the training phase, we included only the data ranges that led to the
215 eventual outcome at the ICU endpoint. This setup ensures a clear representation of
216 both death and survival cases, while maintaining a balanced distribution between the
217 two outcomes (36% death, Table 1). In this particular context, the non-death outcome
218 corresponds to ICU discharge. In total, we trained nine TECO sub-models, each
219 designed to predict the outcome at a specific future time point: 0, 12, 24, 36, 48, 60, 72,
220 84, and 96 hours. Each sub-model was provided with training data from a specific time
221 interval (Supplementary Table 1). The algorithm does not require data to be fully
222 available throughout the entire time interval. Records with incomplete time intervals
223 were included in model training to ensure the model's applicability in a clinical setting.

224 To select hyperparameters, we performed a grid search on the hyperparameters on one
225 training-validation split. The hyperparameter set yielding the highest area under the

Research and Applications

226 receiver operating characteristic (ROC) curve (AUC) from that split was selected,
227 resulting in a model with 6 encoder layers, 8 attention heads, an embedding dimension
228 of 512, and a total of 18,928,130 trainable parameters (Supplementary Table 2).
229 The Transformer model for developing TECO was implemented in PyTorch (Version
230 1.8.1)[29] and trained on an NVIDIA Tesla V100 Tensor Core GPU with 32 GB of
231 memory. TECO was allowed to train for a maximum of 500 epochs with a batch size of
232 32. The SGD optimizer with momentum (0.9) was used to update model parameters.
233 The learning rate was set to 0.01 and reduced by a factor of 2 every 50 epochs. We set
234 the dropout rate to 0 and use the Gaussian Error Linear Unit (GELU) as the activation
235 function in the transformer layers. The training process would stop early if the validation
236 loss did not change by more than 10^{-4} after 100 epochs. Gradient clipping was set to 1.0
237 to avoid gradient exploding (Supplementary Table 2).

238 **Development of Other Models for Comparison**

239 To develop RF and XGBoost models using the same THR COVID-19 dataset, we
240 followed the same data preparation procedure as described above for TECO. The time-
241 dependent variables were aligned and averaged in 15-minute intervals, and
242 concatenated with the baseline variables. The models were trained using the same 20
243 data splits as for TECO. Hyperparameters were selected through a grid search using
244 the same one training-validation split as for TECO (Supplementary Table 2). RF and
245 XGBoost were implemented in scikit-learn (Version 1.0.2).[30]
246 To use EDI for outcome prediction, we used the mean EDI value from the preceding 24
247 hours of each data input time interval. The EDI models predict the binary outcome
248 solely based on a threshold on the continuous EDI values so no hyperparameter tuning

Research and Applications

249 was involved.

250 **Internal and External Validations**

251 We evaluated the prediction performance of the TECO model based on AUC and
252 compared it with the EDI, RF, and XGBoost models. Performance was evaluated
253 separately for each of the nine TECO sub-models. For the internal validation based on
254 the THR COVID-19 dataset, we reported the median AUCs on the validation sets
255 across all 20 training-validation splits. To plot ROC for the EDI-based models, we used
256 all possible thresholds within the range of the EDI data.

257 For external validation based on the MIMIC-IV dataset, we reported the AUCs and real-
258 time probability of death on the ARDS and sepsis cohorts for each involved model,
259 respectively. The EDI was evaluated only in the internal validation due to unavailability
260 of EDI data in the non-Epic-based MIMIC-IV. To validate TECO externally in a manner
261 more akin to a clinical setting, we positioned each model at varying time points after ICU
262 admission and utilized a rolling window of the most recent available data to predict
263 patient outcomes at future time points, as defined by each TECO sub-model's task. For
264 example, the 24-hour sub-model, which utilizes data from the preceding 96 hours
265 (Figure 1), was employed to predict mortality at 120, 132, 144, and up to 240 hours after
266 ICU admission (Figure 2).

267 **Feature Importance and Ablation Study**

268 To determine the feature importance of the ICU monitoring measures, we performed a
269 feature elimination analysis on TECO with 20 random splits of training and validation.
270 For each feature, the importance score was determined using the AUC information
271 gained between the original model and the model without the feature. In addition, we

Research and Applications

272 calculated the impurity scores for RF and compared feature importance consistency
273 between TECO and RF.
274 To assess the impact of baseline variables on model performance, we conducted an
275 ablation study by comparing the AUCs of the full models to those of models using only
276 time-dependent variables.

277 RESULTS

278 A total of 2579 patients were included in the THR model development cohort. All
279 enrolled patients had COVID-19, and of these, 925 (35.9%) expired in the ICU. The
280 characteristics of the baseline variables are presented in Table 1. Among these patients,
281 the median age was 63.0 years, and a majority were male (1499, 58.1%), white (1909,
282 74.0%), non-Hispanic (1723, 66.8%), and overweight to obese (BMI \geq 25) (1657,
283 64.3%). The MIMIC ARDS validation cohort included 2799 patients, of whom 471
284 (16.8%) expired in the ICU. The MIMIC sepsis validation cohort included 6622 patients,
285 of whom 1031 (15.6%) expired in the ICU. These two external validation cohorts
286 presented similar trends in the distribution of baseline variables, with the majority being
287 elderly, male, and white (Table 1).

288 **Table 1. Baseline characteristics in the COVID-19, ARDS, and sepsis cohorts.**

□	COVID-19	MIMIC-ARDS	MIMIC-Sepsis
Number of patients/encounters, n	2579	2799	6622
Age (years)			
Median (Q1, Q3)	63.0 (51.0, 74.0)	66.0 (54.0, 76.0)	66.0 (54.0, 77.0)
Sex, n (%)			
Male	1499 (58.1)	1595 (57.0)	3782 (57.1)

Research and Applications

<i>Female</i>	1080 (41.9)	1204 (43.0)	2840 (42.9)
Race, n (%)			
<i>White</i>	1909 (74.0)	1749 (62.5)	4155 (62.7)
<i>Black</i>	401 (15.5)	217 (7.8)	530 (8.0)
<i>Other</i>	125 (4.8)	168 (6.0)	470 (7.1)
<i>Unknown</i>	144 (5.6)	665 (23.8)	1467 (22.2)
Ethnicity, n (%)			
<i>Hispanic</i>	713 (27.6)	-	-
<i>Non-Hispanic</i>	1723 (66.8)	-	-
<i>Unknown</i>	143 (5.5)	-	-
BMI, n (%)			
<i>Underweight</i>	53 (2.1)	53 (1.9)	160 (2.4)
<i>Normal</i>	379 (14.7)	535 (19.1)	1451 (21.9)
<i>Overweight</i>	698 (27.1)	652 (23.3)	1673 (25.3)
<i>Obese</i>	959 (37.2)	792 (28.3)	1540 (23.3)
<i>Unknown</i>	490 (19.0)	767 (27.4)	1798 (27.2)
Outcome, n (%)			
<i>Death</i>	925 (35.9)	471 (16.8)	1031 (15.6)
<i>Discharge</i>	1654 (64.1)	2328 (83.2)	5591 (84.4)

289 (MIMIC: Medical Information Mart for Intensive Care. ARDS: acute respiratory distress
 290 syndrome.)

291 **Internal Validation**

292 In the COVID-19 model development cohort, AUCs on the validation sets across the 20
 293 data splits are summarized in Supplementary Figure 1 and Supplementary Table 1. In
 294 general, all models' performance improved as the targeted prediction time window was
 295 shortened, with the median AUCs ranging from 0.86 to 0.97. The median AUC of TECO
 296 model, ranging from 0.89 to 0.97, was higher than that of EDI (0.86–0.95), RF (0.87–
 297 0.96), and XGBoost (0.88–0.96) at every prediction time window, demonstrating its
 298 overall advantages. On the other hand, the median AUC of EDI-based prediction was
 299 consistently lower than that from the other models at every prediction time window. It is

Research and Applications

300 noteworthy that the median AUC achieved by TECO when predicting 60-hour mortality
301 (0.93) matched with that based on EDI when predicting 12-hour mortality
302 (Supplementary Table 1), showcasing the advantage in early warning capability for
303 TECO.

304 External Validation

305 In the two external validation cohorts, similar trends of AUC were observed across
306 different time intervals for each model (Table 2; Supplementary Figure 2). Using the
307 sub-model for prediction of 24-hour mortality as a demonstration, for the ARDS cohort,
308 AUCs for all three models improved from the earliest time point of prediction (120 hours
309 since ICU admission), where the AUCs were 0.66 for TECO, 0.60 for RF, and 0.60 for
310 XGBoost, to the latest time point (240 hours), where the AUCs were 0.76 (TECO), 0.73
311 (RF), and 0.72 (XGBoost), respectively. Similarly, for the sepsis cohort, AUCs of the
312 same 24-hour sub-model improved from 0.65 (TECO), 0.57 (RF), and 0.57 (XGBoost)
313 at the earliest prediction time point to 0.75 (TECO), 0.73 (RF), and 0.72 (XGBoost) at
314 the latest time point (Table 2). Based on AUC, the TECO model consistently
315 outperformed RF and XGBoost throughout a 5-day monitoring period in both cohorts.
316 This advantage of TECO was particularly apparent at earlier lookout time points (e.g.
317 120 through 216 hours since ICU admission).

318 As expected, the performance of all models in the external validation was generally
319 lower than that in the internal validation. TECO's performance in the ARDS cohort was
320 slightly better than that in the sepsis cohort, especially at earlier lookout time points
321 (Supplementary Figure 2).

322 **Table 2. Model performance on the external validation cohorts to predict 24-hour**

Research and Applications

323 **mortality.** At each time point after ICU admission, the models use the most recent 96
 324 hours of data to predict mortality in the next 24 hours. For example, at 156 hours after
 325 ICU admission, the models use data from 60 to 156 hours to predict outcomes at 180
 326 hours after admission. Model performances are presented as AUC [95% CI].
 327 Confidence intervals (CIs) are estimated by bootstrapping with 500 iterations, sampling
 328 the whole dataset with replacement. Statistical significance was assessed using
 329 DeLong's test to compare RF or XGBoost with TECO (* $p < 0.05$, *** $p < 0.001$).

Cohort	Hours since ICU admission	TECO	RF	XGBoost
ARDS	120	0.66 [0.60,0.71]	0.60 [0.53,0.67] *	0.60 [0.54,0.67]
	132	0.70 [0.65,0.74]	0.62 [0.57,0.68] *	0.62 [0.57,0.68] *
	144	0.73 [0.69,0.76]	0.67 [0.63,0.71] ***	0.66 [0.62,0.70] ***
	156	0.75 [0.72,0.78]	0.71 [0.67,0.74] *	0.70 [0.66,0.74] *
	168	0.76 [0.72,0.79]	0.73 [0.69,0.77] *	0.72 [0.67,0.75] *
	180	0.76 [0.72,0.78]	0.71 [0.67,0.74] ***	0.70 [0.67,0.74] ***
	192	0.77 [0.74,0.80]	0.72 [0.69,0.76] ***	0.72 [0.68,0.75] ***
	204	0.76 [0.73,0.79]	0.71 [0.67,0.74] ***	0.70 [0.67,0.74] ***
	216	0.76 [0.74,0.79]	0.72 [0.69,0.75] ***	0.72 [0.68,0.75] ***
	228	0.76 [0.73,0.79]	0.73 [0.70,0.76] *	0.72 [0.69,0.75] ***
	240	0.76 [0.72,0.79]	0.73 [0.70,0.76] *	0.73 [0.69,0.76] *
	Sepsis	120	0.65 [0.62,0.70]	0.57 [0.53,0.61] ***
132		0.68 [0.65,0.71]	0.60 [0.57,0.63] ***	0.59 [0.56,0.62] ***
144		0.71 [0.68,0.73]	0.63 [0.60,0.66] ***	0.62 [0.59,0.65] ***
156		0.72 [0.70,0.75]	0.67 [0.65,0.69] ***	0.66 [0.64,0.69] ***
168		0.73 [0.71,0.75]	0.68 [0.66,0.71] ***	0.67 [0.65,0.70] ***
180		0.73 [0.70,0.75]	0.69 [0.66,0.71] ***	0.67 [0.64,0.69] ***
192		0.74 [0.72,0.76]	0.70 [0.68,0.72]	0.68 [0.66,0.70] ***

Research and Applications

	204	0.75 [0.73,0.77]	0.71 [0.68,0.73] ***	0.69 [0.67,0.71] ***
	216	0.76 [0.74,0.78]	0.72 [0.70,0.74] ***	0.71 [0.68,0.73] ***
	228	0.76 [0.74,0.78]	0.73 [0.71,0.75] ***	0.72 [0.70,0.74] ***
	240	0.75 [0.73,0.78]	0.73 [0.71,0.75] *	0.72 [0.70,0.74] ***

330 (AUC: area under the receiver operating characteristic curve. TECO: Transformer-
 331 based Encounter-level Clinical Outcome. RF: random forest. XGBoost: Extreme
 332 Gradient Boosting. ARDS: acute respiratory distress syndrome.)

333 **Monitoring of Patient Deterioration**

334 TECO can be used to provide real-time estimation of mortality probability throughout the
 335 ICU stay. An illustration of this feature is shown in Figure 2, where results from external
 336 validations of the 24-hour sub-model are presented. TECO demonstrated that, at the
 337 cohort level, patients who ultimately survived their ICU stay exhibited a consistently
 338 lower probability of mortality throughout the 5-day monitoring period compared to those
 339 who eventually died in the ICU. Moreover, TECO displayed a decreasing trajectory in
 340 mortality probability for the surviving patients, a trend consistently observed across two
 341 external validation cohorts. Similar findings were observed with all nine TECO sub-
 342 models (Supplementary Figure 3). For an illustration of TECO-generated deterioration
 343 monitoring at the individual patient level, the mortality probability projection of eight
 344 representative patients from the ARDS and sepsis cohorts are shown in Supplementary
 345 Figure 4.

346 **Feature Importance and Ablation Study**

347 In the feature importance analysis, SF ratio appeared to be the most important feature
 348 for TECO, especially at later time points closer to the outcome (Figure 3). The mSOFA

Research and Applications

349 demonstrated the second highest importance. Similarly, the mSOFA and SF ratio
350 showed higher importance for RF, where the importance of SF ratio gradually increased
351 towards the time of the outcome.

352 Removing the contribution of baseline variables led to a performance decrease across
353 all models (TECO, RF, and XGBoost) in the two external validation cohorts
354 (Supplementary Table 3). This performance drop was consistent across models and at
355 different time points when the predictions were made. Notably, the contribution of these
356 baseline variables appears to be independent of the timing of prediction. Importantly,
357 even without baseline variables, TECO remained the top-performing model, particularly
358 at earlier prediction time points, underscoring its intrinsic capability to effectively handle
359 dynamic, time-dependent data.

360 **DISCUSSION**

361 In this study, we developed and validated a novel deep learning algorithm, TECO, for
362 mortality prediction in the ICU. Some existing methods for ICU mortality prediction also
363 utilize transformer architecture and continuous monitoring data, particularly those
364 available from the MIMIC-III and -IV databases.[19, 21] In contrast to these approaches,
365 TECO is a lightweight transformer model specifically tailored to handle time-dependent,
366 irregularly recorded features and time-independent baseline features in a joint manner.

367 Unlike MeTra or Song et al. models, TECO does not presume a fixed time range of
368 either the input data or the outcome. [19, 21] Instead, it can leverage the most recent
369 ICU data to make predictions at future time points.

370 Our work benchmarks TECO against the commercially available EDI, a closed-source
371 metric exclusive to Epic platforms. The robust performance of TECO, in comparison to

Research and Applications

372 EDI, highlights its potential as a better outcome prediction tool that is not confined to a
373 single EHR system provider. The EDI was developed prior to the COVID-19 pandemic
374 but was widely used for clinical decision support and ICU triage during the pandemic.[6,
375 8] In this study, we demonstrate that the EDI had relatively lower predictive performance
376 beyond the 24-hour window prior to the outcome. In contrast, all three non-proprietary
377 models—RF, XGBoost, and TECO—showed advantages especially at time points
378 further from the ICU outcome. According to limited public information, the EDI model
379 does not appear to use SF ratio in its development. [5, 6] Our feature importance
380 analysis demonstrates that SF ratio could be of high importance, which may explain the
381 limited performance of EDI. Besides SF ratio, mSOFA also had a high impact on the RF
382 and TECO. This is consistent with the findings that SOFA is a reliable indicator for
383 mortality among COVID-19 patients,[31-33] and that mSOFA has an equivalent
384 performance in mortality prediction.[28]

385 Compared with RF and XGBoost in each external cohort, TECO outperformed the non-
386 transformer models throughout a 5-day monitoring period, especially on the earlier days
387 in the ICU. Calibration plots also reveal a stronger separation of outcomes based on
388 TECO-estimated mortality probabilities (Supplementary Figure 5). TECO's multi-head
389 attention modules overcomes a bottleneck in traditional recurrent neural networks to
390 learn long-range dependencies in sequences by linearly projecting the dimensions and
391 queries of the input embedding.[12] The validation of TECO in these non-COVID-19
392 external cohorts underscores its generalizability in various severe diseases. It is also
393 worth noting that TECO's performance is comparable to several well-established ICU
394 and in-hospital mortality prediction algorithms.[34-40] Some of these algorithms, while

Research and Applications

395 showing better predictive capabilities,[36, 37, 39], are limited by some essential factors
396 such as a lack of external validation, a significantly smaller development sample size,
397 challenges in real-time monitoring implementation, or a combination of these.

398 We demonstrate TECO's practical utility in the ICU setting by highlighting its ability to
399 monitor patient deterioration. The TECO-estimated mortality risk was consistently
400 elevated among patients who eventually expired in the ICU, reflecting their heightened
401 illness severity (Figure 2). To exclude the possibility that such trends were systemically
402 introduced due to model artifacts, we examined and compared the mortality probability
403 of individual patients (Supplementary Figure 4). Importantly, we observed different
404 patterns among these patients, especially in the earlier days in ICU. Some of these
405 patterns (e.g., Patients D, F, I) differ significantly from the aggregate trends observed at
406 the group level (Figure 2), highlighting TECO's ability to capture real-time, patient-
407 specific details.

408 Developing an algorithm that can leverage the longitudinal time course of inpatient EHR
409 data may improve prediction accuracy and enable earlier detection. In this study, the
410 median AUC achieved by TECO using data up to 60 hours before the outcome (0.93)
411 matched with that based on EDI at 12 hours before the outcome. With the successful
412 validation of the external cohorts, this may suggest that TECO could signal a
413 deterioration alert a full 48 hours before EDI. For ICUs with heavy workloads such as
414 those observed during the COVID-19 pandemic, this improvement could substantially
415 facilitate hospital resource planning, clinician communication with patient families, and
416 play a vital role in future public health emergencies.

Research and Applications

417 **LIMITATIONS AND FUTURE DIRECTIONS**

418 This study has several limitations. Firstly, the usage of COVID-19 data to develop the
419 TECO model was primarily motivated by the large sample size accumulated over the
420 pandemic in our EHR systems. In addition, this Epic-based dataset included the
421 proprietary EDI as a benchmark to evaluate our model—an option not available in other
422 public ICU data sources. Ideally, external validation could have been conducted on a
423 COVID-19 cohort from a different health system. Unfortunately, such data were not
424 available at the time of this study. When validating TECO in the two external, non-
425 COVID-19 cohorts, we found the model's performance decreased compared with that in
426 the COVID-19 cohort. While a performance drop from the training setting onto the
427 independent testing setting is frequently observed [41, 42], it is important to note that
428 the two external validation cohorts in this study represent two distinct diseases that
429 differ from COVID-19 and the models were not trained on these diseases. Another
430 potential contributor to the observed performance drop may stem from the fact that,
431 when training TECO using the COVID-19 cohort, we only utilized patient outcomes at
432 the ICU endpoint (i.e., death or discharge). This approach was intended to ensure
433 representation of the two extreme scenarios across a broad spectrum of patients'
434 physiological conditions in the ICU. However, when evaluating TECO in the external
435 cohorts over a moving time scale (120-240 hours since ICU admission), patients who
436 are not at immediate risk of death may not necessarily present a health status that is
437 ready for discharge. While TECO demonstrates proof of concept as a potential ICU
438 monitoring tool, further validation across a broader range of disease states and ICU
439 settings could significantly improve its performance and generalizability.

Research and Applications

440 Nevertheless, our study demonstrates the feasibility of directly using structured EHR
441 data, especially ICU monitoring data, to empower deep learning-based outcome
442 prediction. The current version of TECO incorporates only 11 variables among which 6
443 are typically measured in the ICU in real time. Future work will need to consider whether
444 more complex models using additional variables could influence the models'
445 performance and generalizability, especially given TECO's intrinsic advantages in
446 handling long-range, high-frequency data. However, operational costs associated with
447 more complex and computationally intensive models must be thoroughly evaluated
448 when considering these potential improvements. The TECO model in this study remains
449 a lightweight transformer, which should not present significantly greater implementation
450 challenges than the more conventional RF and XGBoost models. Moreover, with the
451 advancement of large language models, clinical notes may also be used as additional
452 features to provide valuable insights.[43-46]

453 Secondly, missing data and inconsistent data quality across different health systems or
454 sites may significantly limit the applicability of a data-intensive model like TECO. For
455 example, erroneous data due to instrument or human operations are frequently
456 captured in the ICU data. More comprehensive data quality screening and control could
457 benefit the implementation of TECO in real-world settings.

458 Lastly, it is worth noting that patients diagnosed with ARDS could potentially meet the
459 sepsis criteria. Despite the absence of these cohorts in the training data for TECO,
460 ensuring no information leakage, this scenario may still introduce a degree of bias and
461 pose challenges to the implementation of TECO in the ICU and the interpretation of its
462 outputs.

Research and Applications

463 **CONCLUSION**

464 We developed TECO, a transformer-based model, to analyze multi-dimensional,
465 continuous monitoring data for ICU mortality prediction. In internal validation, TECO
466 outperformed EDI-based prediction and other conventional machine learning methods.
467 In two external validation cohorts (where EDI was not available), TECO outperformed
468 other conventional machine learning methods. TECO may be further tailored as a
469 disease-generic early warning tool in the ICU or inpatient settings.

470 **DATA AVAILABILITY STATEMENT**

471 The COVID-19 dataset could not be shared publicly due to data and privacy protection
472 policies at Texas Health Resources and UT Southwestern Medical Center. The MIMIC
473 dataset is publicly available at <https://physionet.org/content/mimiciv/2.2/>.

474 **ACKNOWLEDGEMENTS**

475 **COMPETING INTERESTS**

476 The authors have no potential conflict of interest to disclose.

477 **FUNDING**

478 This study was supported in part by the National Institutes of Health under award
479 number R01GM140012 (GX), R01DE030656 (GX), R01GM115473 (GX),
480 U01CA249245 (GX), U01AI169298 (YX), R35GM136375 (YX), the Cancer Prevention
481 and Research Institute of Texas (CPRIT RP180805: YX; CPRIT RP230330: GX), and
482 the Texas Health Resources Clinical Scholars Program.

Research and Applications

483 **AUTHOR CONTRIBUTIONS**

484 Conceptualization: RR, ZG, GX, DMY, YX

485 Data curation and formal analysis: RR, ZG, HL, TN, TK, CW, CC, FV, GX, DMY, YX

486 Methodology: RR, ZG, CC, AMN, EDP, GX, DMY, YX

487 Writing: RR, ZG, KWJ, CC, AMN, FV, EDP, GX, DMY, YX

488

Research and Applications

489 **FIGURE LEGENDS**

490 **Figure 1. The TECO algorithm design.** This figure demonstrates a 24-hour mortality
491 prediction example, utilizing data from the preceding 96 hours to predict the binary
492 outcome (death vs. non-death). Time-dependent, ICU monitoring variables were aligned
493 and concatenated with baseline variables, then embedded into 512-dimensional feature
494 vectors. These feature vectors were combined with positional timestamp vectors and
495 fed into a multi-layer Transformer-encoder. (BMI: Body Mass Index. mSOFA: modified
496 Sequential Organ Failure Assessment score. mSOFA-resp: mSOFA respiratory sub-
497 score. Resp Rate: respiratory rate. SpO2: pulse oximetry, SF ratio: SpO2/FiO2 ratio.
498 FiO2: fraction of inspired oxygen.)

499
500 **Figure 2. TECO-based monitoring of mortality probability in two external**
501 **validation cohorts.** The green line represents patients who were eventually discharged
502 alive from ICU, while the orange line represents patients who died in ICU in the ARDS
503 (left) cohort and sepsis (right) cohort, respectively. Probability of mortalities are
504 aggregated over repeated hours since admission to show the mean and 95%
505 confidence interval.

506 **Figure 3. Feature importance analysis in the COVID-19 cohort.** The left panel shows
507 the information gain calculated through feature elimination for TECO. Each bar
508 represents the AUC loss from 20 different validation splits. The right panel shows the
509 random forest impurity importance for each feature with score > 0.05.

510

Research and Applications

511 REFERENCES

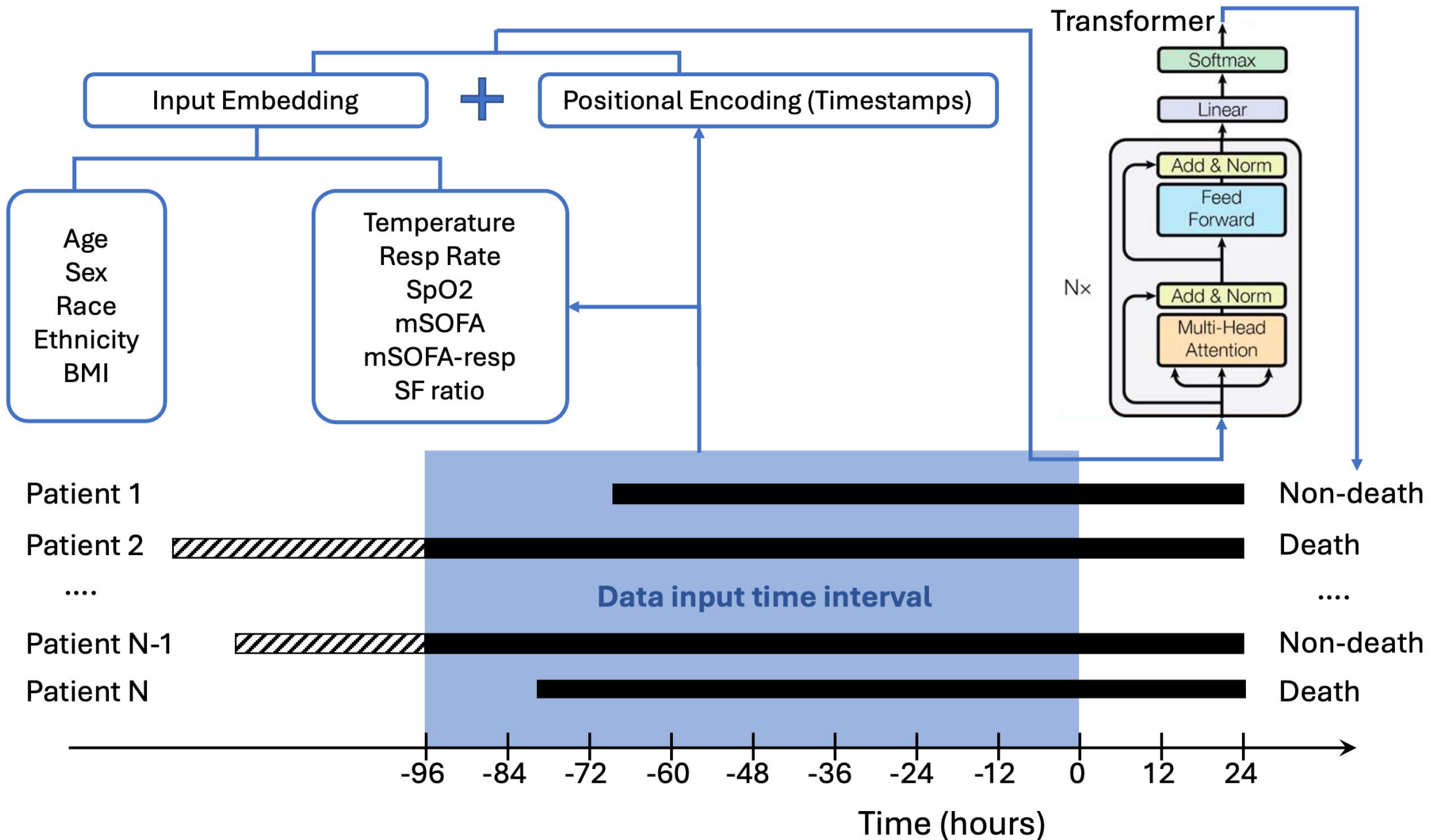
- 512
513 1. Awrahman, BJ, C Aziz Fatah, and MY Hamaamin A Review of the Role and Challenges
514 of Big Data in Healthcare Informatics and Analytics. *Comput Intell Neurosci*
515 2022;2022:5317760.
- 516 2. Peterson, ED Machine Learning, Predictive Analytics, and Clinical Practice: Can the Past
517 Inform the Present? *JAMA* 2019;322(23):2283-2284.
- 518 3. Dieteren, CM, MAJ van Hulsen, KIM Rohde, et al. How should ICU beds be allocated
519 during a crisis? Evidence from the COVID-19 pandemic. *PLoS One*
520 2022;17(8):e0270996.
- 521 4. Craxi, L, M Vergano, J Savulescu, et al. Rationing in a Pandemic: Lessons from Italy.
522 *Asian Bioeth Rev* 2020;12(3):325-330.
- 523 5. Systems, E, *Saving Lives with AI: Using the Deterioration Index Predictive Model to*
524 *Help Patients Sooner*. 2022: [https://www.epicshare.org/share-and-learn/saving-lives-](https://www.epicshare.org/share-and-learn/saving-lives-with-ai)
525 [with-ai](https://www.epicshare.org/share-and-learn/saving-lives-with-ai).
- 526 6. Systems, E, *Artificial Intelligence Triggers Fast, Lifesaving Care for COVID-19 Patients*.
527 2020: [https://www.epic.com/epic/post/artificial-intelligence-epic-triggers-fast-lifesaving-](https://www.epic.com/epic/post/artificial-intelligence-epic-triggers-fast-lifesaving-care-covid-19-patients/)
528 [care-covid-19-patients/](https://www.epic.com/epic/post/artificial-intelligence-epic-triggers-fast-lifesaving-care-covid-19-patients/).
- 529 7. Systems, E, *Cognitive Computing Model Brief: Deterioration Index*. 2023:
530 <https://galaxy.epic.com/Redirect.aspx?DocumentID=3883949&PrefDocID=99556>.
- 531 8. Singh, K, TS Valley, S Tang, et al. Evaluating a Widely Implemented Proprietary
532 Deterioration Index Model among Hospitalized Patients with COVID-19. *Ann Am*
533 *Thorac Soc* 2021;18(7):1129-1137.
- 534 9. Raman, G, B Ashraf, YK Demir, et al. Machine learning prediction for COVID-19
535 disease severity at hospital admission. *BMC Med Inform Decis Mak* 2023;23(1):46.
- 536 10. Bendavid, I, L Statlender, L Shvartser, et al. A novel machine learning model to predict
537 respiratory failure and invasive mechanical ventilation in critically ill patients suffering
538 from COVID-19. *Sci Rep* 2022;12(1):10573.
- 539 11. Cummings, BC, S Ansari, JR Motyka, et al. Predicting Intensive Care Transfers and
540 Other Unforeseen Events: Analytic Model Validation Study and Comparison to Existing
541 Methods. *JMIR Med Inform* 2021;9(4):e25066.
- 542 12. Vaswani, A, N Shazeer, N Parmar, et al. Attention is all you need. *Advances in neural*
543 *information processing systems* 2017;30.
- 544 13. Hochreiter, S and J Schmidhuber Long short-term memory. *Neural computation*
545 1997;9(8):1735-1780.
- 546 14. Wu, N, B Green, X Ben, et al. Deep transformer models for time series forecasting: The
547 influenza prevalence case. *arXiv preprint arXiv:2001.08317* 2020.
- 548 15. Huang, K, J Altsaar, and R Ranganath Clinicalbert: Modeling clinical notes and
549 predicting hospital readmission. *arXiv preprint arXiv:1904.05342* 2019.
- 550 16. Li, Y, S Rao, JRA Solares, et al. BEHRT: Transformer for Electronic Health Records. *Sci*
551 *Rep* 2020;10(1):7155.
- 552 17. Rasmy, L, Y Xiang, Z Xie, et al. Med-BERT: pretrained contextualized embeddings on
553 large-scale structured electronic health records for disease prediction. *NPJ Digit Med*
554 2021;4(1):86.

Research and Applications

- 555 18. Antikainen, E, J Linnosmaa, A Umer, et al. Transformers for cardiac patient mortality
556 risk prediction from heterogeneous electronic health records. *Sci Rep* 2023;13(1):3517.
- 557 19. Khader, F, JN Kather, G Muller-Franzes, et al. Medical transformer for multimodal
558 survival prediction in intensive care: integration of imaging and non-imaging data. *Sci*
559 *Rep* 2023;13(1):10666.
- 560 20. Cheng, J, J Sollee, C Hsieh, et al. COVID-19 mortality prediction in the intensive care
561 unit with deep learning based on longitudinal chest X-rays and clinical data. *Eur Radiol*
562 2022;32(7):4446-4456.
- 563 21. Song, H, D Rajan, J Thiagarajan, et al. *Attend and diagnose: Clinical time series analysis*
564 *using attention models*. in *Proceedings of the AAAI conference on artificial intelligence*.
565 2018.
- 566 22. Johnson, AEW, L Bulgarelli, L Shen, et al. MIMIC-IV, a freely accessible electronic
567 health record dataset. *Sci Data* 2023;10(1):1.
- 568 23. Force, ADT, VM Ranieri, GD Rubenfeld, et al. Acute respiratory distress syndrome: the
569 Berlin Definition. *JAMA* 2012;307(23):2526-33.
- 570 24. Yang, P, T Wu, M Yu, et al. A new method for identifying the acute respiratory distress
571 syndrome disease based on noninvasive physiological parameters. *PLoS One*
572 2020;15(2):e0226962.
- 573 25. Singer, M, CS Deutschman, CW Seymour, et al. The Third International Consensus
574 Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;315(8):801-10.
- 575 26. Vincent, JL, R Moreno, J Takala, et al. The SOFA (Sepsis-related Organ Failure
576 Assessment) score to describe organ dysfunction/failure. On behalf of the Working
577 Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine.
578 *Intensive Care Med* 1996;22(7):707-10.
- 579 27. Jones, AE, S Trzeciak, and JA Kline The Sequential Organ Failure Assessment score for
580 predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the
581 time of emergency department presentation. *Crit Care Med* 2009;37(5):1649-54.
- 582 28. Grissom, CK, SM Brown, KG Kuttler, et al. A modified sequential organ failure
583 assessment score for critical care triage. *Disaster Med Public Health Prep* 2010;4(4):277-
584 84.
- 585 29. Paszke, A, S Gross, F Massa, et al. Pytorch: An imperative style, high-performance deep
586 learning library. *Advances in neural information processing systems* 2019;32.
- 587 30. Pedregosa, F, G Varoquaux, A Gramfort, et al. Scikit-learn: Machine learning in Python.
588 *the Journal of machine Learning research* 2011;12:2825-2830.
- 589 31. Liu, S, N Yao, Y Qiu, et al. Predictive performance of SOFA and qSOFA for in-hospital
590 mortality in severe novel coronavirus disease. *Am J Emerg Med* 2020;38(10):2074-2080.
- 591 32. Yang, Z, Q Hu, F Huang, et al. The prognostic value of the SOFA score in patients with
592 COVID-19: A retrospective, observational study. *Medicine (Baltimore)*
593 2021;100(32):e26900.
- 594 33. Esmaeili Tarki, F, S Afaghi, FS Rahimi, et al. Serial SOFA-score trends in ICU-admitted
595 COVID-19 patients as predictor of 28-day mortality: A prospective cohort study. *Health*
596 *Sci Rep* 2023;6(5):e1116.
- 597 34. Ye, Z, S An, Y Gao, et al. The prediction of in-hospital mortality in chronic kidney
598 disease patients with coronary artery disease using machine learning models. *Eur J Med*
599 *Res* 2023;28(1):33.

Research and Applications

- 600 35. Li, F, H Xin, J Zhang, et al. Prediction model of in-hospital mortality in intensive care
601 unit patients with heart failure: machine learning-based, retrospective analysis of the
602 MIMIC-III database. *BMJ Open* 2021;11(7):e044779.
- 603 36. Pang, K, L Li, W Ouyang, et al. Establishment of ICU Mortality Risk Prediction Models
604 with Machine Learning Algorithm Using MIMIC-IV Database. *Diagnostics (Basel)*
605 2022;12(5).
- 606 37. Iwase, S, TA Nakada, T Shimada, et al. Prediction algorithm for ICU mortality and
607 length of stay using machine learning. *Sci Rep* 2022;12(1):12912.
- 608 38. Jamshidi, E, A Asgary, N Tavakoli, et al. Using Machine Learning to Predict Mortality
609 for COVID-19 Patients on Day 0 in the ICU. *Front Digit Health* 2021;3:681608.
- 610 39. Villar, J, JM Gonzalez-Martin, J Hernandez-Gonzalez, et al. Predicting ICU Mortality in
611 Acute Respiratory Distress Syndrome Patients Using Machine Learning: The Predicting
612 Outcome and STRatification of severity in ARDS (POSTCARDS) Study. *Crit Care Med*
613 2023;51(12):1638-1649.
- 614 40. Jeon, ET, HJ Lee, TY Park, et al. Machine learning-based prediction of in-ICU mortality
615 in pneumonia patients. *Sci Rep* 2023;13(1):11527.
- 616 41. Ramspek, CL, KJ Jager, FW Dekker, et al. External validation of prognostic models:
617 what, why, how, when and where? *Clin Kidney J* 2021;14(1):49-58.
- 618 42. Siontis, GC, I Tzoulaki, PJ Castaldi, et al. External validation of new risk prediction
619 models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*
620 2015;68(1):25-34.
- 621 43. Hao, B, S Sotudian, T Wang, et al. Early prediction of level-of-care requirements in
622 patients with COVID-19. *Elife* 2020;9.
- 623 44. Izquierdo, JL, J Ancochea, C-RG Savana, et al. Clinical Characteristics and Prognostic
624 Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective
625 Study Using Machine Learning and Natural Language Processing. *J Med Internet Res*
626 2020;22(10):e21801.
- 627 45. Adamidi, ES, K Mitsis, and KS Nikita Artificial intelligence in clinical care amidst
628 COVID-19 pandemic: A systematic review. *Comput Struct Biotechnol J* 2021;19:2833-
629 2850.
- 630 46. Ye, J, L Yao, J Shen, et al. Predicting mortality in critically ill patients with diabetes
631 using machine learning and clinical notes. *BMC Med Inform Decis Mak* 2020;20(Suppl
632 11):295.
633



ARDS cohort

Sepsis cohort

