

19 Tanshi Li, E-mail: lts301@163.com. Department of Emergency, Chinese PLA General Hospital,
20 Beijing, 100853, China.

21 Zhengbo Zhang, Email: zhengbozhang@126.com. Department of Biomedical Engineering and
22 Maintenance Center, Chinese PLA General Hospital, Beijing, 100853, China.

23 **Funding:**

24 There is no financial funding or interest to report.

25 **Take-home Message:**

26 A real-time early warning system for forecasting the onset of acute kidney injury (AKI) within
27 the first week of intensive care unit (ICU) stay in adult intensive care has been developed and
28 validated in two general populations, which considering both urine output and serum creatinine
29 diagnostic criteria and learned with machine learning models.

30 **Author Contributions:**

31 SH, CF designed the study, conducted the data collection, data analysis, data interpretation, and
32 wrote the manuscript. SH, CF conducted the data analysis, data interpretation and wrote the
33 manuscript. LC and WW conducted the data interpretation and reviewed the manuscript. ZZ and
34 TL designed the study and reviewed the manuscript. KL, LW, and XC conducted the data
35 interpretation and reviewed the manuscript.

36 **Conflicts of Interest**

37 The authors declare that they have no conflict of interest.

38 **Abstract**

39 **Background:** To improve the performance of early acute kidney injury (AKI) prediction in
40 intensive care unit (ICU), we developed and externally validated machine learning algorithms in
41 two large ICU databases.

42 **Methods:** Using eICU[®] Collaborative Research Database (eICU) and MIMIC-III databases, we
43 selected all adult patients (age ≥ 18). The detection of AKI was based on both the oliguric and
44 serum creatinine criteria of the KDIGO (Kidney Disease Improving Global Outcomes). We
45 developed an early warning system for forecasting the onset of AKI within the first week of ICU
46 stay, by using 6- or 12-hours as the data extraction window and make a prediction within a 1-
47 hour window after a gap window of 6- or 12-hours. We used 52 features which are routinely
48 available ICU data as predictors. eICU was used for model development, and MIMIC-III was
49 used for externally validation. We applied and experimented on eight machine learning
50 algorithms for the prediction task.

51 **Results:** 3,816 unique admissions in multi-center eICU database were selected for model
52 development, and 5,975 unique admissions in single-center MIMIC-III database were selected
53 for external validation. The incidence of AKI within the first week of ICU stay in eICU and
54 MIMIC-III cohorts was 52.1% (n=1,988) and 31.3% (n=1,870), respectively. In eICU cohort, the
55 performance of AKI prediction is better with shorter extraction window and gap window. We
56 found that the AdaBoost algorithm yielded the highest AUC (0.8859) on the model with 6-hours
57 data extraction window and 6-hours gap window (model 6-6) rather than other prediction models.
58 In MIMIC-III cohort, AdaBoost also performed well.

59 **Conclusions:** We developed the machine learning-based early AKI prediction model, which
60 considered clinical important features and has been validated in two datasets.

61 **Keywords:** Acute Kidney Injury; Intensive Care Unit; Artificial Intelligence; Machine Learning

62 **Background**

63 Acute kidney injury (AKI) is the major cause of multiple organ failure with an incidence more
64 than 50% in critically ill patients¹⁻⁵. Early prevention as one of the limited management options
65 could greatly improve the prognosis of AKI in the intensive care unit (ICU) and therefore the
66 early identification of critical patients with AKI becomes a critical issue, which could offer an
67 opportunity to develop strategies for early prevention and intervention of AKI in critical care⁶⁻⁹.
68 The available dataset of large electronic healthcare records (EHR) provide an opportunity of
69 predicting the early onset of AKI in critical care patients using machine learning algorithms,
70 which can provide a much earlier, applicable and cost-effective solution than current late
71 diagnostic criteria and biomarkers with limited utilities^{6, 10-18}.

72 Forecasting the onset of AKI was well-studied in previous. Some studies have reported
73 that their AKI prediction models were desirable. However, deploying such methods to the real
74 world would be relatively difficult due to several critical limitations¹⁹⁻²¹: (1) Some prediction
75 models were based on highly specific cohorts, such as cardiac disease and post-surgery cohorts,
76 which lack of the general hospital population background, thus need to be further explored²²⁻²⁷.
77 (2) The definitions of AKI in some studies were based on the serum creatinine (SCr) criteria
78 without oliguria criteria. Even though oliguria criteria is non-early and non-specific, some
79 studies have already confirmed that nearly one-third of AKI patients in ICU were identified by
80 oliguria criteria and without elevating SCr²⁶⁻³⁶. Some of the previous study such as Tomašev *et al.*
81 2019 were due to the lack of urine output information. (3) Those prediction tasks simply rely on
82 a single time point, which is the time of admission to hospital, due to data insufficiency^{26, 27, 31, 32,}
83 ³⁷⁻⁴⁰. (4) Many studies did not include physiological or laboratory parameters as predictors,
84 especially urine output, which has been confirmed as a critical role in predicting AKI and ICU

85 mortality^{26, 27, 34, 35, 41}. (5) The explored machine learning methods and selected features in the
86 studies were limited, which may not fully take the advantage of machine learning techniques
87 based on “big healthcare data” as well as investigating the best model performance^{26, 31-36, 39, 40, 42}.
88 For example, Tomašev *et al.* 2019 developed a deep learning approach for the continuous risk
89 prediction of AKI, but they did not compare their approach with other baseline algorithms such
90 as basic neural networks or other ensemble algorithms. (6) Many studies did not perform
91 external validation or not considering the real-world deployment^{26, 31-36, 40, 43-47}.

92 To overcome the above limitations, we utilized a large multi-center ICU database (eICU
93 Collaborative Research Database (v1.2)) and used both oliguria and SCr diagnostic criteria of
94 Kidney Disease Improving Global Outcomes (KDIGO)¹⁴ as the definition of AKI, and included
95 most of the routinely available physiological and laboratory parameters in ICU and clinical
96 interventions as predictors. We then adopted eight machine learning algorithms to conduct
97 experiments and develop the comparable models of predicting the early onset of AKI within the
98 first week of ICU stay. We also used another large non-overlapping single-center MIMIC-III
99 (Medical Information Mart for Intensive Care III (v1.4)) ICU database for external validation.
100 By fully utilizing the existed databases, we evaluated four time series models in an artificial
101 intelligence perspective. The most important clinical features are also identified by feature
102 importance analysis.

103

104 **Methods**

105 **Study Design and Databases**

106 We developed an early warning model for prediction the onset of AKI based on the retrospective
107 analysis of the multi-center eICU database. The database including 459 ICUs from 58 hospitals
108 during 2003–2016 in the United States^{48, 49}. The model was externally validated on a single-
109 center publicly available ICU database (MIMIC-III database), which contains 5 different ICUs in
110 a tertiary medical center (Beth Israel Deaconess Medical Center, BIDMC; Boston,
111 Massachusetts) during 2001–2012^{50, 51}.

112 **Study Cohorts**

113 All patients in the publically available eICU (v1.2) and MIMIC-III (v1.4) databases were
114 considered in this study. The inclusion criteria are as follows: 1) Age \geq 18 years old⁵²; 2) ICU
115 stay for at least 24 consecutive hours; and 3) The first ICU admission of the first hospitalization
116 were used⁵³. The exclusion criteria are as follows: 1) The primary diagnosis contained “Kidney
117 disease”, “Acute renal failure”, and “Renal obstruction”; 2) clinical history of end-stage renal
118 disease (ESRD) or their SCr was \geq 4 mg/dl within the first 24 hours of ICU admission²⁶; 3)
119 Received renal replacement therapy (RRT) within the first 24 hours of ICU admission⁵⁴; 4)
120 diagnosed as AKI based on oliguria diagnostic criteria within the first 24 hours of ICU admission
121^{54, 55}; 5) Withdrawal of treatment (which contains patients who did not receive mechanical
122 ventilation and died during ICU stay); 6) samples with any missing value after imputation⁵³.

123 **AKI Definition**

124 The detection of AKI in adults was based on the oliguric and SCr criteria of the KDIGO⁵⁶. The
125 “baseline” of SCr was defined as the ICU admission.^{32, 37, 38, 57, 58}. The oliguric criteria is based
126 on the urine output $<$ 0.5 ml/kg/h for more than 6 hours or 12 hours, or the urine output $<$ 0.3
127 ml/kg/h for 24 hours⁵⁶.

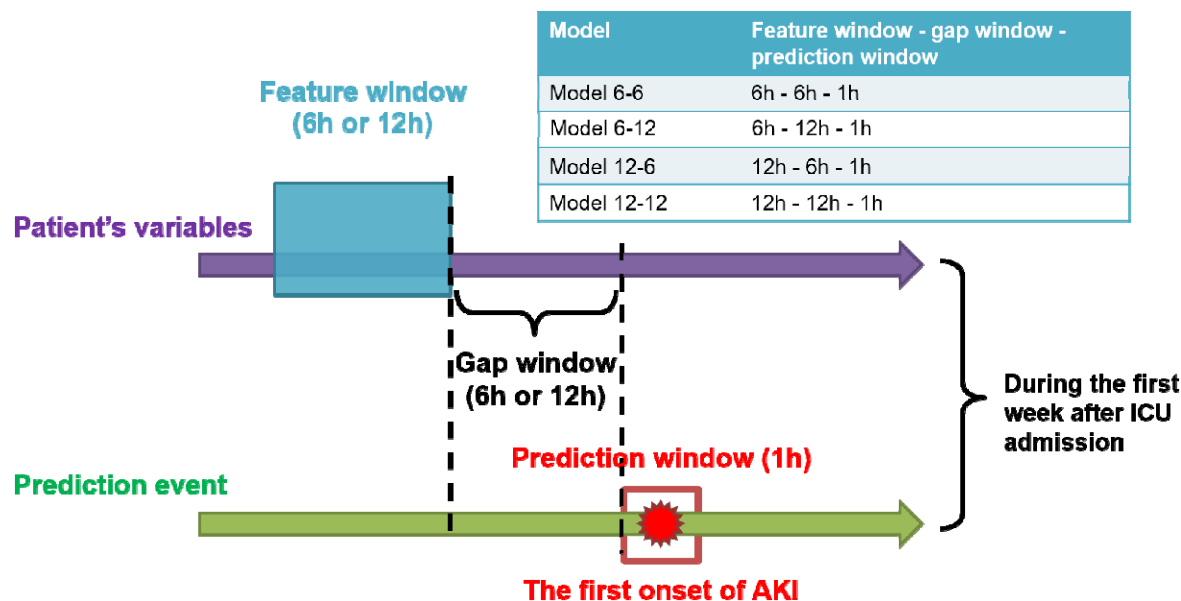
128 **Prediction Task and Outcomes**

129 The prediction target or the primary outcome of this study was the first onset of AKI at any stage
130 within the first week of ICU stay^{31, 52, 53}. We used 6 or 12 hours^{54, 55, 59} as the feature collecting
131 window, since the urine output based diagnostic criteria of KDIGO relies on 6 or 12 consecutive
132 hours, and we make a one-hour prediction window after a gap window of 6 or 12 hours. That is,
133 we used 6 or 12 hours of data to forecast the onset of AKI after 6 or 12 hours. All features and
134 labels are extracted at one-hour interval for time-series modeling (Figure 1). Although sampled
135 in one-hour, forecasting the future onset of AKI is a continuous prediction over the time.

136 In detail, we experimented four configurations of time-series models. The configurations
137 are described as follow with model name and the hours of feature collecting window – the hours
138 of gap window – the hours of prediction window, respectively.

- 139 1) Model 6-6: 6h-6h-1h;
- 140 2) Model 6-12: 6h-12h-1h;
- 141 3) Model 12-6: 12h-6h-1h;
- 142 4) Model 12-12: 12h-12h-1h.

143 Secondary outcomes of the study included ICU mortality, initial sequential organ failure
144 assessment (ISOFA) score, need for renal replacement therapy (RRT) and ICU length of stay
145 (LOS)^{31, 52, 53}. The prediction task and different models are shown in Figure 2.



146

147 Figure 1. Model settings and prediction for a single admission. Timestamp is at one-hour interval.

148 Feature collecting window are the discretized time-series features (6 hours or 12 hours) as inputs

149 of machine learning algorithms for predicting the label of AKI after certain gaps (6 hours or 12

150 hours). We experimented on four different model configurations: model 6-6, model 6-12, model

151 12-6, and model 12-12.

152

153 **Model Comparison and Feature Importance Ranking**

154 The violin plot⁶⁰ and pairwise paired t-test (two sided)⁶¹ were used to visualize and evaluate the

155 performances across four different configurations of our time-series models (model 6-12; model

156 6-6; model 12-12, and model 12-6). Gini importance⁶², also known as the mean decrease

157 impurity, is averaged over all trees of the ensemble model, which was used for identifying and

158 ranking feature importance. Gini importance is calculated by scikit-learn packages⁶³.

159

160 **Feature Examination and Secondary Outcomes**

161 We used “psych” package⁶⁴ to further examine the association between ICU mortality, length of
162 stay (LOS), need for renal replacement therapy (RRT), and the occurrence of AKI in the first
163 week to obtain the secondary outcomes, since these features were collected after the onset of
164 AKI and not showed in Table 1. Moreover, ISOFA score was also included into secondary
165 outcomes. Continuous features are shown as the mean with standard deviation (s.d.) or median
166 with interquartile range (IQR). Categorical features are expressed as absolute (n) and relative (%)
167 frequency. Selected features were analyzed by the Spearman ρ correlation test (two-sided),
168 where the Spearman ρ statistic is used to estimate a rank-based measure of association. The test
169 is a non-parametric and distribution-free test, thus the data do not necessarily follow a bivariate
170 normal distribution⁶⁵. The significance was considered when $P < 0.05$. Statistical analysis was
171 performed by Python 3.7 and R version 3.5.1.

172

173 **Data Extraction and Pre-processing**

174 **Predictors**

175 For each patient, we extracted most of the routinely available ICU data as predictors based on
176 methods of previous studies^{31, 42, 54, 59, 66} and the availability in the selected databases. We
177 included following features: 1) 18 static variables of demographic information, primary
178 diagnosis and comorbidities (*e.g.*, age, gender, sepsis, hypertension, etc.). The primary diagnosis
179 were identified at the ICU admission from the databases; 2) 31 time-varying variables of vital
180 signs, laboratory values and fluid balance-related parameters (*e.g.*, heart rate, oxygen saturation,
181 white blood cell count, blood creatinine, fluids intake, urine output, etc.); 3) two critical

182 interventions, which are usages of mechanical ventilation and vasopressors; 4) two other
 183 variables: “Hours” (hours stayed in ICU) and Initial SOFA (ISOFA). The complete feature list is
 184 shown in Table 1, detailed feature selection criteria are listed in Table S1.

185 It is worth to know that those variables also include urine output and serum creatinine
 186 (SCr), the same two variables which were used to define the onset of AKI based on KDIGO⁵⁶.
 187 In time series analysis, these two variables are known as endogenous variables while other
 188 variables are served as exogenous variables^{67, 68}. Though the definition of AKI is based on urine
 189 output and SCr, forecasting the onset of AKI after 6 or 12 hours gap window cannot base on the
 190 future unseen urine output and SCr. Thus, complex machine learning algorithms are involved
 191 into this forecasting problem.

Groups	Variables						
Demographic information	Age	Gender	Ethnicity	Height	Weight	BMI ¹	
Vital signs	Temperature (TEM)	Heart rate (HR)	Respiratory rate (RR)	Diastolic blood pressure (DIAS_BP)	Systolic blood pressure (SYS_BP)	Mean blood pressure (MEAN_BP)	Oxygen saturation (SPO2)
Laboratory values	Sodium (NA_ION)	Calcium (CA_ION)	Potassium (K_ION)	Chloride (CL_ION)	Anion gap (AG)	Glucose (GLU)	Bicarbonate (HCO3)
	White blood cell count	Red blood cell count (RBC)	Hemoglobin (HGB)	Hematocrit (HCT)	Platelets (PLT)	P_F_ratio ² (PF)	

	(WBC)					
	Glutamic pyruvic transaminase (ALT)	Glutamic oxalacetic transaminase (AST)	Total bilirubin (TBB)	Blood urea nitrogen (UN)	Serum creatinine (SCr)	
Fluid balance	Fluids intake last 6 hours (INPUT_6HR)	Fluids intake last 12 hours (INPUT_12HR)	Fluids intake last 24 hours (INPUT_24HR)	Urine output last 6 hours (OUTPUT_6HR)	Urine output last 12 hours (OUTPUT_12HR)	Urine output last 24 hours (OUTPUT_24HR)
Additional respiratory and hemodynamic support	Vasopressor (VASO)	Mechanical ventilation (MECH)				
Primary diagnoses	Sepsis, including pneumonia (SEP)	Cardiovascular (CAR)	Neurological (NEU)	Other Respiratory (RES)	Other diseases (OD)	
Comorbidities	Hypertension (HYP)	Diabetes (DIA)	Congestive heart failure (CHF)	Chronic pulmonary disease (CPD)	Chronic kidney disease (CKD)	Chronic liver disease (CLD)
Others	Hours stayed in ICU (HOURS)	Initial SOFA ³ (ISOFA)				

192 Table 1. Variables and categorized groups. ¹Body Mass Index: Weight (kg) / Height (meter) *

193 Height (meter). BMI then converted to classes according to following criteria: BMI < 18.5: class

194 = 0; $18.5 \leq \text{BMI} < 23$: class = 1; $23 \leq \text{BMI} < 25$: class = 2; $25 \leq \text{BMI} < 40$: class = 3; $\text{BMI} \geq 40$:
195 class = 4. ²P_F_ratio: PO₂/FiO₂. ³SOFA: Sequential Organ Failure Assessment.

196

197 **Dataset Partitioning and Missing Data Imputation**

198 eICU database was used for model development and the data was analyzed with five-fold
199 cross-validation: *i.e.*, randomly split the entire cohorts into 60%:20%:20% as training, validation
200 and testing sets, respectively. The training set (60%) is used for developing machine learning
201 models, the validation set (20%) helps algorithms find their best hyper-parameters learned by the
202 training set. The testing set (20%) is used for evaluating the performances of the models learned
203 and optimized by training and validation sets. To prevent bias introduced when different
204 admissions of one patient appears in separate dataset, data was splitted according to their unique
205 patient ID instead of the admission ID, and assigned into either train, validation, or test set, to
206 ensure patients are mutually exclusive across different datasets.

207 Missing values were imputed with linear imputation approach ⁶⁹ with respect to each
208 admission. Once the linear imputation is done, nearest-neighbor imputation ⁷⁰⁻⁷² was adopted
209 within each admission to fill the remained missing values outside of the linear imputation range.
210 All imputations were performed piece-wisely on each single admission. As linear imputation was
211 adopted, it is critical to avoid data leakage on the validation and testing datasets. In our scenario,
212 each variable on each admission in validation and testing sets are either partially missing or
213 complete, otherwise that admission will be dropped. We then let those admissions impute the
214 missing value by itself instead of using the training set. Moreover, the onset of AKI is occurred
215 at the last timestamp of each admission since the data after the onset of AKI or at the time of

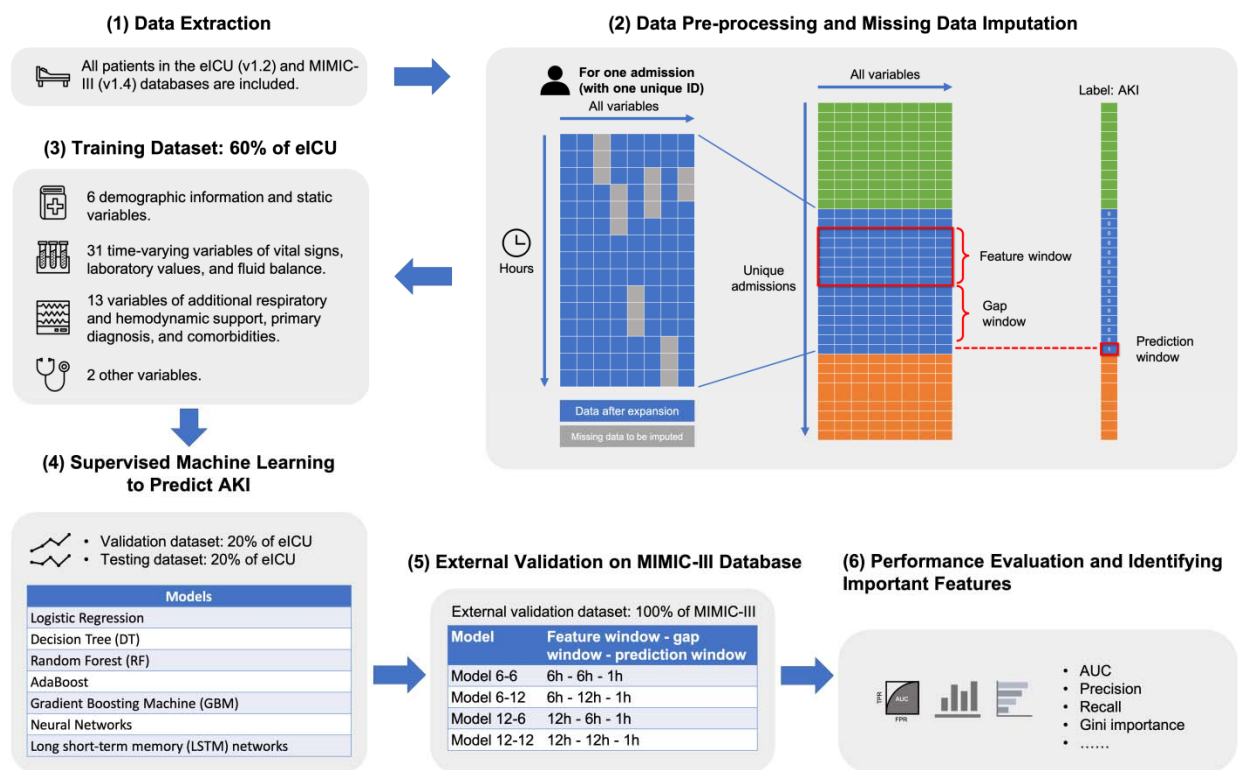
216 AKI were not used in model-building, thus, the experiment will not have the data leakage issue
217 on time and will not bias the prediction after the data imputation.

218 Data were coded as multi-dimensional time-series at one-hour interval⁶⁶. Static variables
219 (demographic information, primary diagnosis, comorbidities, and ISOFA) were replicated across
220 all timestamps for each patient. Categorical features with multiple categories but without order
221 information, such as “ethnicity”, were one-hot encoded by using “OneHotEncoder” function
222 from scikit-learn package⁶³. Vital signs and laboratory values with multiple measurements
223 within one-hour interval were pooled in average⁶⁶ and rounded to the nearest unit hour.

224

225 **Model Development and Validation**

226 Driven by the large population database, we used eight different machine learning
227 algorithms to develop our early AKI prediction model. The algorithms include: logistic
228 regression (baseline), random forest (RF)⁷³, AdaBoost (short for Adaptive Boosting, base
229 estimator: decision tree with maximum depth = 1)⁷⁴, decision tree (DT), Gradient Boost
230 Machine (GBM)⁷⁵, neural network (NN), L2 regularized neural network (NN-L2), and long
231 short-term memory (LSTM) neural networks⁷⁶. Then we used MIMIC-III database for external
232 validation. Time-series data were encoded and reshaped to 1-D vectors to feed the learning
233 algorithms except for LSTM. Hyper-parameters are fine-tuned for each algorithm on validation
234 set. The entire workflow is shown in Figure 2.



235

236 Figure 2. The study overview of the artificial intelligence-assisted AKI early warning system,
 237 including (1) Data extraction, (2) Data pre-processing and missing data imputation, (3) dataset
 238 splitting, (4) supervised machine learning method to predict AKI, (5) external validation, and (6)
 239 performance evaluation and important feature identification

240

241 Performance Evaluation and Statistical Analysis

242 As logistic regression was widely used in predicting time-series events for ICU-related problems
 243 ^{26, 54, 66}, we used logistic regression as the baseline for comparisons. Performances was evaluated
 244 by: 1) the area under the ROC curve (AUC) ⁷⁷; 2) Precision (macro) ⁷⁸; 3) Recall (macro) ⁷⁸; 4)
 245 F1 score (macro) ⁷⁹; 5) negative predictive value ⁸⁰; and 6) specificity ⁸¹. AUC, ranging from 0
 246 to 1, the higher the better, indicates the algorithm's performances. We considered the model with

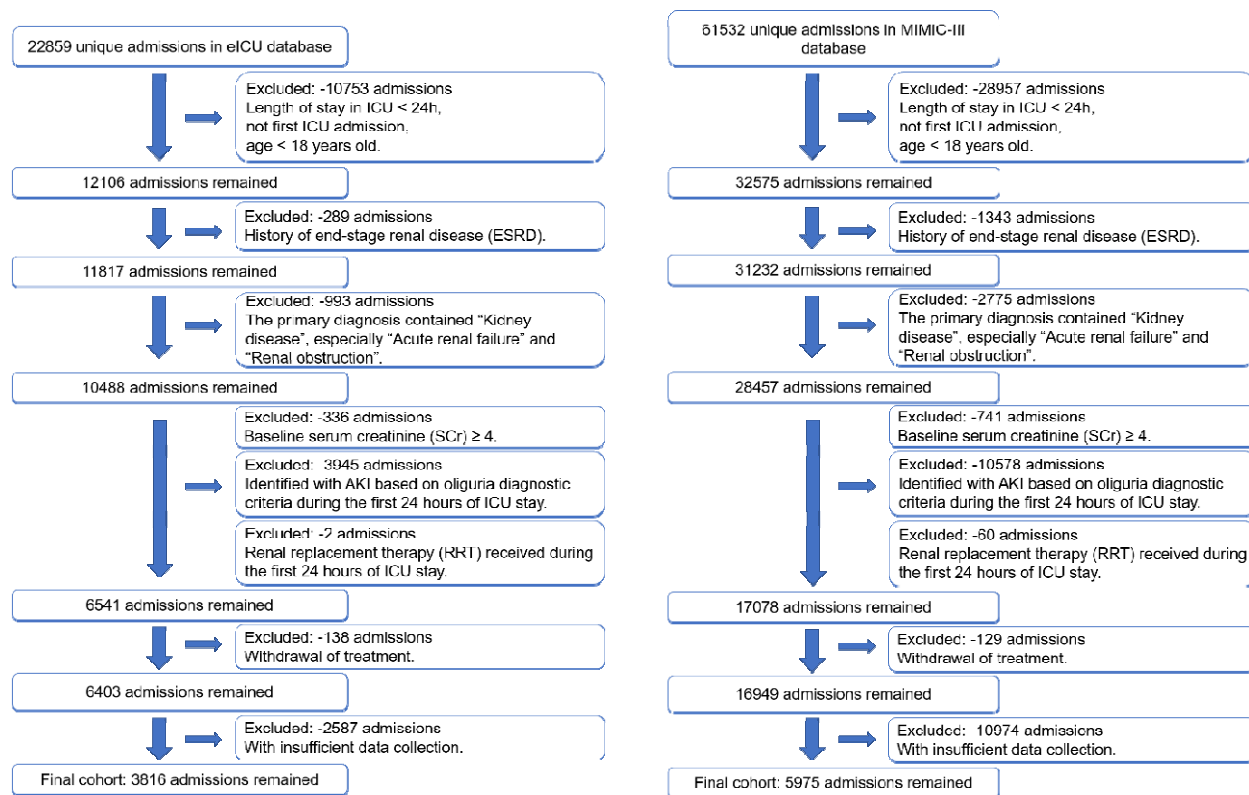
247 AUC above 0.85 as the adequate and well-performed model. Model with the highest AUC is
248 considered as the best model. Precision, also known as positive predictive value (PPV)⁸⁰, is the
249 fraction of true positive classification among the positive results classified by algorithm, a higher
250 precision indicates an algorithm's result is reliable. Recall, also known as sensitivity⁸¹, is the
251 fraction of true positive classification among all the true samples, describes the ability of
252 identifying true samples (AKI) among the whole population. F1 score is the harmonic average of
253 precision and recall, higher F1 score indicates better performance. As we reported PPV
254 (precision) and sensitivity (recall), we also took negative predictive value (NPV) and specificity
255 into account. NPV is the probability that an admission with a negative prediction truly don't have
256 the AKI. Specificity measures the proportion of actual negatives samples that are correctly
257 identified. Since five-fold cross-validation training scheme was adopted which has five different
258 mutually exclusive testing sets (each contains 20% of the entire cohort), thus all the performance
259 metrics are computed 5 times. Results are either presented by mean values in Tables or by
260 boxplot in Figures.

261

262 **Results**

263 **Study Cohorts: eICU and MIMIC-III**

264 A total of 3,816 unique admissions of eICU database and 5,975 unique admissions of MIMIC-III
265 database were included after data pre-processing and filtering by inclusion and exclusion criteria.
266 Detailed inclusion and exclusion process are presented in Figure 3.



267

A

B

268 Figure 3. Study cohort selection workflow of eICU (A) and MIMIC-III database (B) based on the
269 designed inclusion and exclusion criteria.

270

271 Descriptions and statistics of patients' demographics and clinical characteristics between
272 two cohorts are summarized in Table 2. Commonly missing features are also described in Table
273 S2. The eICU cohort contained 3,816 unique admissions with a mean age of 60 (standard
274 deviation=18.0); 56.8% (n=2,169) were male and 84.8% (n=3,234) were white (including
275 Eastern European, Brazilian, Russian, and other European). 5,975 unique admissions selected in
276 the final MIMIC-III cohort with a mean age of 60 (standard deviation=18.3); 57.0% (n=3,404)
277 were men and 69.3% (n=4,140) were white. The incidence of AKI within the first week in eICU

278 and MIMIC-III cohorts was 52.1% (n=1,988) and 31.3% (n=1,870), respectively. The AKI cases
 279 were determined based on urine output criteria in eICU and MIMIC-III cohorts were 45.1%
 280 (n=1,722) and 24.9% (n=1,486). Sepsis, including pneumonia, as the primary diagnosis
 281 accounted for 18.7% (n=713) in eICU cohort and 38.39% (n=2,290) in MIMIC-III cohort.
 282 Cardiovascular disease accounted for 35.8% (n=1367) in eICU cohort and 25.6% (n=1,531) in
 283 MIMIC-III cohort. The proportion of patients on mechanical ventilation and vasopressors were
 284 33.4% (n=1,276) and 26.4% (n=1,006) in the eICU, 47.4% (n=2,830) and 27.9% (n=1,666) in
 285 the MIMIC-III. The proportion of patients on cardiac surgery prior to AKI were 3.9% (n=149) in
 286 the eICU and 2.0% (n=122) in the MIMIC-III. The mean ISOFA score was 5.8 (standard
 287 deviation=3.3) in the eICU and 4.1 (standard deviation=3.0) in the MIMIC-III. The need for
 288 renal replacement therapy (RRT) was 1.8% (n=68) in the eICU and 0.95% (n=57) in the MIMIC-
 289 III. Median ICU length of stay was 2.6 days in eICU and 2.8 days in MIMIC-III. The proportion
 290 of ICU mortality was 10.0% in eICU and 8.6% in MIMIC-III.

Key Characteristics	eICU	MIMIC-III
Unique ICU types (n)	8	5
Final cohort (n)	3,816	5,975
The incidence of AKI within the first week (n(%))	1,988 (52.1%)	1,870 (31.3%)
AKI cases were determined based on urine output criteria (n(%))	1722 (45.1%)	1486 (24.9%)
Age, years (mean (s.d.))	60 (18.0)	60 (18.3)
Male gender (n(%))	2,169 (56.8%)	3,404 (57.0%)
Comorbidities (n(%)):		
Hypertension	1,625 (42.6%)	2,430 (40.7%)

Diabetes	527 (13.8%)	1,253 (21.0%)
Congestive heart failure	593 (15.5%)	1,172 (19.6%)
Chronic pulmonary disease	550 (14.4%)	795 (13.3%)
Chronic kidney disease	379 (10.0%)	395 (6.6%)
Chronic liver disease	201 (5.3%)	108 (1.8%)
Primary ICD-9 diagnosis (n(%)):		
Sepsis, including pneumonia	713 (18.7%)	2,290 (38.3%)
Cardiovascular	1,367 (35.8%)	1,531 (25.6%)
Other Respiratory	670 (17.6%)	642 (10.7%)
Neurological	401 (10.5%)	280 (4.7%)
Others	665 (17.4%)	1,232 (20.6%)
Ethnicity (n(%)):		
Black	264 (6.9%)	420 (7.0%)
Hispanic	66 (1.7%)	273 (4.6%)
Asian	38 (1.0%)	202 (3.4%)
White	3,234 (84.8%)	4,140 (69.3%)
Other	214 (5.6%)	940 (15.7%)
Initial SOFA (mean (s.d.))	5.8 (3.3)	4.1 (3.0)
Cardiac surgery (n(%))	149 (3.9%)	122 (2.0%)
Mechanical ventilation (n(%))	1,276 (33.4%)	2,830 (47.4%)
Vasopressors (n(%))	1,006 (26.4%)	1,666 (27.9%)
Need for renal replacement therapy (RRT) (n(%))	68 (1.8%)	57 (0.95%)
Length of stay, days (median, (IQR))	2.6 (1.6-4.8)	2.8 (1.8-5.0)
ICU mortality rate	10.00%	8.60%

291

292 Table 2. Description of the datasets. Ethnicity of Black includes African American, Cape
293 Verdean, Haitian, and African. Ethnicity of Hispanic includes Latino (central American), Latino
294 (Cuban), Latino (Puerto Rican), Latino (Honduran), Latino (Guatemalan), Latino (Mexican),
295 Latino (Dominican), Latino (Salvadoran), Latino (Colombian), and Portuguese. Ethnicity of
296 Asian includes Vietnamese, Thai, Asian Indian, middle Eastern, Korean, Chinese, Filipino,
297 Cambodian, Japanese, Asian other. White includes Eastern European, Brazilian, Russian, and
298 other European. Ethnicity of Other includes Unknown, not specified, multi race, patient declined
299 to answer, and unable to obtain. s.d.: Standard deviation in short. IQR: interquartile range in
300 short.

301

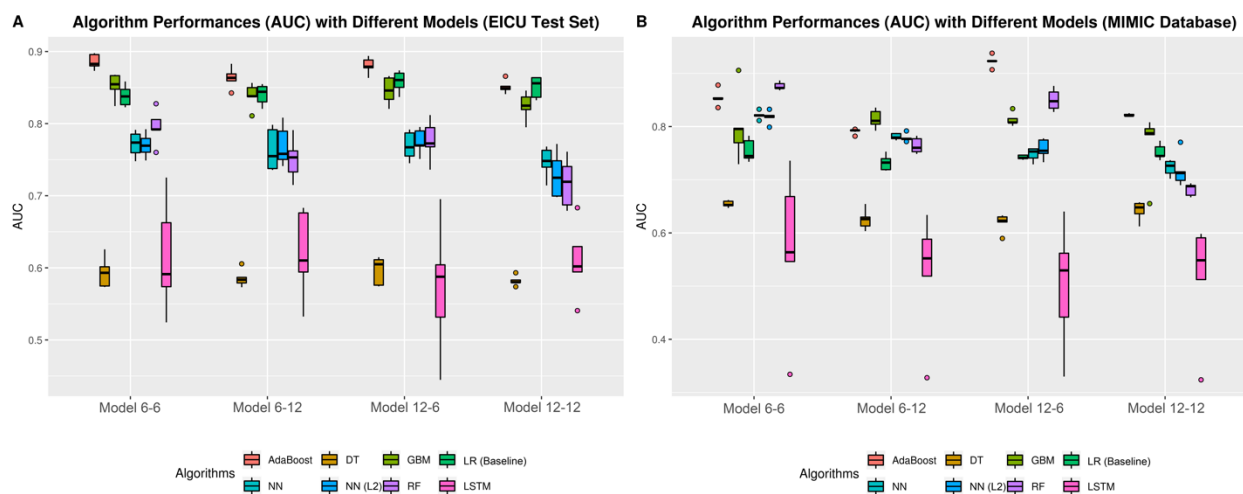
302 **Model Development and Evaluation in eICU Cohort**

303 Compared with the baseline algorithm (logistic regression algorithm) in eICU cohort, we found
304 that AdaBoost and Gradient Boosting Machine (GBM) achieved higher AUC than other
305 algorithms in predicting task of all four configurations of the time-series models. The model 6-6
306 configuration in general yielded the highest performance. Using model 6-6, the AUC of logistic
307 regression is 0.8385, while the AUC was higher in the AdaBoost (0.8859) and GBM (0.8522).
308 AUC results of all five-fold cross validation across four different configurations of the time-
309 series models and eight different machine learning algorithms are shown in Figure 4A. Precision
310 and Recall of the models are also demonstrated in Figure S1A and Figure S2A. The model
311 performance (mean values) on eICU cohort testing sets is shown in Table 3. Other performance

312 metrics on eICU cohort testing sets MIMIC-III such as F1 score, negative predictive value
313 (NPV), and specificity are also available at Table S3.

314 **Model External Validation and Evaluation in MIMIC-III Cohort**

315 After external validation on four configurations of the time-series models which developed by
316 eight different machine learning algorithms, we found model 12-6 developed by AdaBoost
317 achieved the highest AUC (0.9228) on external validation set. For some models, the higher
318 performance is noted on external validation than model development stage due to the differences
319 between eICU and MIMIC-III cohorts (for example, the onset of AKI in MIMIC-III is more
320 imbalanced), yet the evaluations are comparable within each database cohorts. AUC results of all
321 five-folds results across four different configurations of the time-series models and eight
322 different machine learning algorithms on MIMIC-III external validation set are presented in
323 Figure 4B. The model performances (mean values) on external validation cohort MIMIC-III of
324 AUC, precision, and recall are shown in Table 4. Other performance metrics on external
325 validation cohort MIMIC-III such as F1 score, negative predictive value (NPV), and specificity
326 are also available at Table S4.



327

328 Figure 4. Performance of four different configurations of the time-series models (model 6-6,
 329 model 6-12, model 12-6, model 12-12) and eight different machine learning algorithms on eICU
 330 for model development (A) and MIMIC-III for external validation (B). Abbreviations: DT:
 331 Decision Tree; GBM: Gradient Boosting Machine; LR: Logistic Regression; NN: Neural
 332 Networks; NN (L2): Neural Networks with L2 regularization; RF: Random Forest; LSTM: Long
 333 short-term memory networks.

334

Models and algorithms	Logistic Regression (Baseline)	Decision tree (DT)	AdaBoost	Gradient Boosting Machine (GBM)	Random forest (RF)	Neural Networks (NN)	Neural Networks (NN with L2)	Long short-term memory networks (LSTM)
Model 6-6								
AUC	0.8385	0.5937	0.8859	0.8522	0.7954	0.7716	0.7701	0.6154
Precision	0.6934	0.5209	0.7381	0.5430	0.5503	0.5328	0.5295	0.5056
Recall	0.5353	0.5937	0.5762	0.5307	0.5262	0.6016	0.5975	0.5492
Model 6-12								
AUC	0.8402	0.5856	0.8634	0.8385	0.7509	0.7636	0.7695	0.6192
Precision	0.7051	0.5185	0.7073	0.5707	0.5437	0.5377	0.5370	0.5002
Recall	0.5380	0.5856	0.5814	0.5551	0.5197	0.6014	0.6040	0.5094
Model 12-6								
AUC	0.8583	0.5962	0.8804	0.8459	0.7766	0.7692	0.7751	0.5727
Precision	0.7152	0.5229	0.7339	0.5742	0.5504	0.5269	0.5307	0.5022
Recall	0.5524	0.5962	0.5913	0.5584	0.5301	0.6147	0.6262	0.5151
Model 12-12								

AUC	0.8505	0.5820	0.8509	0.8244	0.7175	0.7467	0.7286	0.6100
Precision	0.6997	0.5187	0.7412	0.5746	0.5454	0.5249	0.5198	0.5005
Recall	0.5443	0.5820	0.5784	0.5496	0.5177	0.6076	0.5931	0.5098

335

336 Table 3. Comparison of model performance (mean values across five-fold validation results) on
 337 eICU internal validation. Abbreviations: AUC: area under the ROC curve. L2 stands for Ridge
 338 regularization.

339

Models and algorithms	Logistic Regression (Baseline)	Decision tree (DT)	AdaBoost	Gradient Boosting Machine (GBM)	Random forest (RF)	Neural Networks (NN)	Neural Networks (NN with L2)	Long short-term memory networks (LSTM)
Model 6-6								
AUC	0.7550	0.6544	0.8542	0.7993	0.8760	0.8213	0.8175	0.5696
Precision	0.5018	0.5123	0.5103	0.5059	0.5558	0.5073	0.5071	0.5638
Recall	0.5395	0.6544	0.5006	0.5648	0.5692	0.6857	0.6886	0.5267
Model 6-12								
AUC	0.7322	0.6253	0.7911	0.8142	0.7642	0.7811	0.7788	0.5241
Precision	0.5018	0.5080	0.5039	0.5108	0.5449	0.5064	0.5064	0.5268
Recall	0.5387	0.6253	0.5017	0.5987	0.6241	0.6690	0.6647	0.5136
Model 12-6								
AUC	0.7422	0.6192	0.9228	0.8131	0.8502	0.7479	0.7579	0.5006
Precision	0.5019	0.5077	0.5044	0.5076	0.5593	0.5047	0.5051	0.5343
Recall	0.5412	0.6192	0.5014	0.6005	0.6291	0.6695	0.6773	0.5077

Model 12-12								
AUC	0.7522	0.6419	0.8215	0.7662	0.6816	0.7224	0.7173	0.5147
Precision	0.5022	0.5095	0.5038	0.5066	0.5213	0.5044	0.5043	0.5250
Recall	0.5451	0.6419	0.5011	0.5792	0.6223	0.6512	0.6459	0.5140

340

341 Table 4. Comparison of model performance (mean values across five-fold validation results) on
342 MIMIC-III cohort for external validation. Abbreviations: AUC: area under the ROC curve. L2
343 stands for Ridge regularization.

344

345 **Secondary Outcomes**

346 Secondary outcomes of eICU and MIMIC-III databases are summarized in Table 5. P-value is
347 derived according to alternative hypothesis: true Spearman ρ is not equal to 0 when comparing
348 the variable with the onset of AKI. The patients who were diagnosed with AKI within the first
349 week of ICU were associated with higher ICU mortality (eICU: 261 (13.1%) vs 121 (6.6%),
350 spearman $\rho = 0.1083$, $p < 0.01$; MIMIC-III: 284 (15.2%) vs 230 (5.6%), spearman $\rho = 0.1585$,
351 $p < 0.01$), renal replacement therapy (RRT) (eICU: 63 (3.2%) vs 5 (0.3%), spearman $\rho = 0.1093$,
352 $p < 0.01$; MIMIC-III: 47 (2.5%) vs 10 (0.2%), spearman $\rho = 0.1083$, $p < 0.01$), longer ICU length
353 of stay (eICU: median=3.1 (2.2-3.9) vs median=2.0 (1.1-2.5), spearman $\rho = 0.2738$, $p < 0.01$;
354 MIMIC-III: median=4.8 (IQR: 3.5-5.8) vs median=2.2 (IQR: 1.2-2.6), spearman $\rho = 0.4185$,
355 $p < 0.01$), and larger ISOFA score (eICU: mean=6.2 (standard deviation=3.4) vs mean=5.3
356 (standard deviation=3.1), spearman $\rho = 0.1229$, $p < 0.01$; MIMIC-III: mean=4.9 (standard
357 deviation=3.3) vs mean=3.8 (standard deviation=2.8), spearman $\rho = 0.1623$, $p < 0.01$).

Groups and outcomes	eICU			MIMIC-III		
	AKI occurred within first week (n=1988)	AKI not occurred within first week (n=1828)	Spearman ρ (P-value*)	AKI occurred within first week (n=1870)	AKI not occurred within first week (n=4105)	Spearman ρ (P-value*)
ICU mortality (n(%))	261 (13.1%)	121 (6.6%)	0.1083 (p<0.01)	284 (15.2%)	230 (5.6%)	0.1585 (p<0.01)
Renal Replacement Therapy (RRT) (n(%))	63 (3.2%)	5 (0.3%)	0.1093 (p<0.01)	47 (2.5%)	10 (0.2%)	0.1083 (p<0.01)
Length of stay, days (median (IQR))	3.1 (2.2-3.9)	2.0 (1.1-2.5)	0.2738 (p<0.01)	4.8 (3.5-5.8)	2.2 (1.2-2.6)	0.4185 (p<0.01)
Initial SOFA (mean (s.d.))	6.2 (3.4)	5.3(3.1)	0.1229 (p<0.01)	4.9 (3.3)	3.8 (2.8)	0.1623 (p<0.01)

358

359 Table 5. Secondary outcomes of eICU and MIMIC-III databases. s.d.: standard deviation in short.

360 IQR: interquartile range in short. *P-value is derived according to alternative hypothesis: true

361 Spearman ρ is not equal to 0 when comparing the variable with the onset of AKI.

362

363

364 Relationships between Prediction Performances and Different Time-Series Models

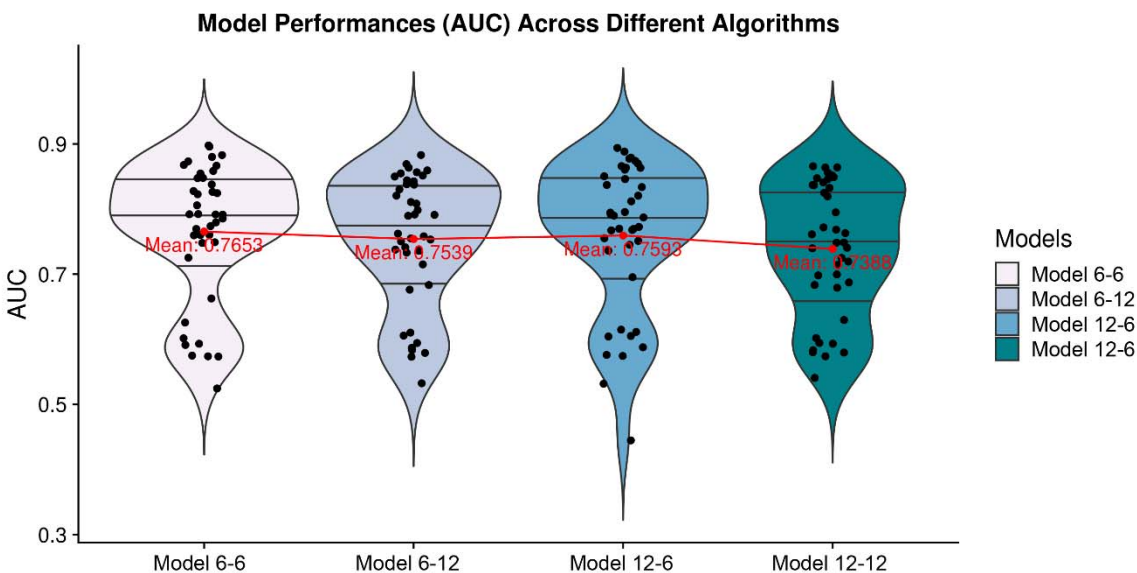
365 As eight different algorithms with five-fold cross-validation produces 40 results per each time

366 series model (model 6-6, model 6-12, model 12-6, model 12-12), violin plot (Figure 5) and

367 pairwise paired t-test (Table 6) are used to compare the overall performances across time series

368 models. In figure 5, we demonstrated that given fixed feature collecting window, models with

369 closer gap tend to have better AUC in average and in pairwise paired t-test (model 6-6: 0.7653 >
370 model 6-12: 0.7539, p-value: 0.0238; model 12-6: 0.7593 > model 6-12: 0.7388, p-value:
371 0.0218). Moreover, with fixed gap window, models with shorter feature collecting window also
372 tend to have better AUC in average and in pairwise paired t-test (model 6-6: 0.7653 > model 12-
373 6: 0.7593, p-value: 0.459; model 6-12: 0.7539 > model 12-12: 0.7388, p-value: 0.0227). Such
374 findings suggest that the prediction performance improved with closer gaps, while increasing the
375 lag of time-series data doesn't improve the model performances.



376
377 Figure 5. Model performances (AUC) violin plot of four different configurations of the models
378 (model 6-6, model 6-12, model 12-6, model 12-12) across eight different algorithms with five-
379 fold cross-validation. Black solid horizontal lines in each violin plot are the interquartile range
380 lines. Red points and lines indicate the mean performances (AUC) over four models.

381

	Set 2
--	-------

		Model 6-6		Model 6-12		Model 12-6		Model 12-12	
		t	P	t	P	t	P	t	P
Set 1	Model 6-6	-	-	2.35	2.38E-02*	0.75	4.59E-01	4.22	1.40E-04*
	Model 6-12	-2.35	2.38E-02*	-	-	-0.73	4.71E-01	2.37	2.27E-02*
	Model 12-6	-0.75	4.59E-01	0.73	4.71E-01	-	-	2.39	2.18E-02*
	Model 12-12	-4.22	1.40E-04*	-2.37	2.27E-02*	-2.39	2.18E-02*	-	-
Notes: t denotes the pairwise paired Student's t-test statistic, P denotes the p-value obtained. P-value < 0.05 are significant and indicated with * symbol.									

382

383 Table 6. Pairwise paired t-test comparison among the five-fold cross-validation AUC results over
384 four different time-series models.

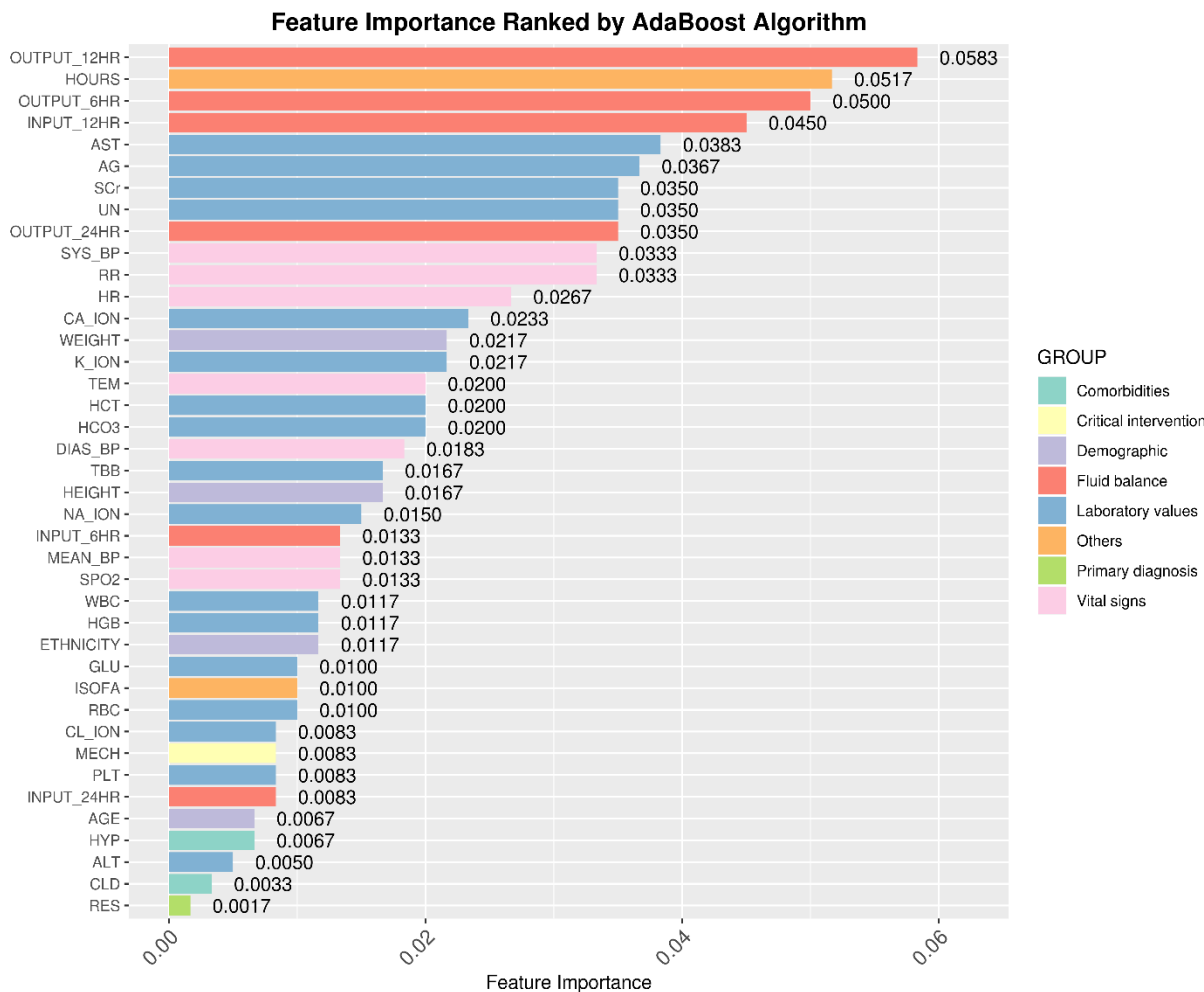
385

386 **Identifying Feature Importance**

387 The feature importance identified by the AdaBoost algorithm is determined by the average
388 feature importance over all ensembled trees. We categorized eight groups of features and
389 analyzed feature importance by Gini importance (mean decrease impurity). The list of
390 categorized groups is shown in Table 1.

391 To gain insights into the relevance of each feature, Figure 6 summarized and ranked the
392 most critical variables in model 6-6 based on the averaged Gini feature importance using
393 AdaBoost algorithm in model development (eICU database). Note that vitals, labs, topics and
394 static data are important for the prediction task, and the most important group of features was
395 vital signs. Specifically, fluid balance-related features (OUTPUT_12HR, OUTPUT_6HR,
396 INPUT_12HR, OUTPUT_24HR), AST, AG, SCr, UN, CA_ION, K_ION in the class of

397 laboratory values, SYS_BP, RR, HR in the class of vital signs, HOURS in the class of others,
 398 and WEIGHT in the class of Demographic, are the most important predictors for the early AKI
 399 prediction.



400
 401 Figure 6. Identified top important features in eICU database. Values and rankings are based on
 402 time-series model 6-6 and the averaged Gini importance from AdaBoost algorithm. Features are
 403 colored group-wisely according to the categories listed in Table 1. The higher value indicates the
 404 higher significance of the feature.

405

406 Discussion

407 Electronic healthcare systems (EHR) provides real-world clinical data not only for secondary
408 analysis, but also for developing the artificial intelligence-based platform to assist clinicians to
409 identify potential critical events, such as the onset of AKI⁸²⁻⁸⁴. Since there is no effective
410 treatment of AKI, the prevention of AKI becomes more critical^{2, 59, 85-89}, thus an early warning
411 system which may reduce the risk of exacerbating injury is necessary and in urgent demand.

412 Though some studies on early warning system have been successfully reduced the
413 nephrotoxic medication in AKI and prevented the contrast-induced AKI⁹⁰⁻⁹², several other
414 studies showed the failure of those methods⁹³⁻⁹⁵, which mainly due to the lack of external
415 validation. Therefore, the study with external validation are necessary before any developed
416 method be applied to the clinical setting, especially the externally validated prediction models
417 are relatively rare^{19, 42, 56}.

418 In this paper, we have integrated most of the routinely available ICU data based on two
419 large ICU databases with general population to developed and externally validated an ideal
420 artificial intelligence assisted early warning system for predicting the onset of AKI within the
421 first week of ICU stay. Our workflow and time-series models addressed several topics that have
422 been discussing over these days. For example, we used both oliguric and SCr as diagnostic
423 criteria of the KDIGO to identify the onset of AKI and included most of physiological and
424 laboratory parameters as predictors, which contained different clinical commonly used urine
425 output trends. As a result, we have got an ideal incidence of AKI in both eICU and MIMIC-III
426 cohort, which would minimize the potential mistakes and maximize the accuracy and improving
427 of predicting model performance in clinical situation. Moreover, we fully made use of all

428 routinely available ICU and clinical features as predictors, to optimize and improve the model
429 performances.

430 Using multiple types of ICU and clinical features to predict AKI, a complex clinical
431 syndrome, is very challenging and difficult as any feature or small changes may cause great
432 impact on outcomes. Machine learning technique has been successfully applied to predict critical
433 events under a fast-paced, data-overloaded setting of ICU, which could provide new insights into
434 evidence-based decision support⁹⁶⁻¹⁰¹. Different machine learning algorithms usually used to
435 capture the real relationships across different data types which may provide advantages for
436 predicting AKI far in advance of onset with high sensitivity and specificity. As previous studies,
437 many machine learning algorithms have been used training many ideal models in predicting AKI,
438 such as random forest, multivariate regression model, boosted ensembles of decision trees, etc.^{26,}
439^{31, 59}. Based on these previous studies, we take a further step by fully utilizing two independent
440 databases and applied eight machine learning algorithms with 52 comprehensive features of
441 different data types for a large-scale model assessments and analyses.

442 Our current study has several strengths. First, we integrated most of the routinely
443 available ICU data, including 52 features under different data types, which would be considered
444 containing the largest number of predictors. Second, we not only used both urine output and SCr
445 as diagnostic criteria to identify the onset of AKI, but also included physiological (especially
446 urine output) and laboratory trends as predictors to improve model performance. Third, we
447 compared 8 modern machine learning algorithms for further analysis to select the best model of
448 predicting task. Fourth, we developed and externally validated an ideal model based on two
449 independent general population ICU databases. Last but not least, we did not only rely on using

450 data at a single time point (admission to hospital), but also combining static and time-varying
451 variables as a time series dataset to achieve accurate prediction model performance.

452 From the results and performances across four time-series models and various machine
453 learning algorithms, we proved the feasibility of deploying artificial intelligence assisted early
454 warning system for AKI prediction. We also showed that AdaBoost, one of the ensemble
455 machine learning algorithms, could be a desired model to predict the onset of AKI. The
456 performances measured by AUC is decent and could be applied in real-time to assist clinicians'
457 decision in the future. While the across-model performances demonstrated that increasing the lag
458 (feature collecting window), or in another word, how many hours we look back, doesn't help
459 improve the performance, justified that model 6-6 with 6 hours feature collecting window is
460 enough for the future AKI inference.

461 For feature importance ranking of AKI prediction, the top 15 features are critical and
462 closely related to AKI development. We justified and validated that the fluid balance and
463 laboratory values are not only the AKI criteria of the KDIGO, but also directly associated with
464 the kidney function (such as OUTPUT_12HR, OUTPUT_6HR, INPUT_12HR, SCr, UN, CA_
465 ION, K_ION, etc.)⁵⁴. Most vital signs, which are used to assess patients' physiological stability,
466 are important for prediction as the AKI prediction should be in real-time. Hours stayed in ICU
467 and weight are also indicated as the most important features which have already been reported
468 that both of them have a great association with AKI¹⁰². The model developed has the potential
469 capability of serving as an "alert" for early warning of AKI, which would make the electronic
470 healthcare systems more artificially intelligence and enable bedside application.

471 Although our study bridged many gaps in AKI forecasting research, this study may still
472 exist some inevitable limitations to be addressed in the future. First, the AKI is defined with the
473 baseline creatinine chosen as first in stay, which could miss the true incidence of AKI by up to
474 30% who already had AKI at that time¹⁰³. However, as we face the limitation of the data, the
475 safest way for now is to exclude that group of patients with unclear information. Same exclusion
476 due to the data limitation also happened in the withdrawal of treatment, which contains patients
477 who did not receive mechanical ventilation and died during ICU stay. Second, we predict the
478 onset of AKI at any stage, without predicting the precise stage. In the future, multi-class
479 classification or even regression might be considered instead of binary classification. Third,
480 comorbidities were included as “static” variables, since they could not be accurately
481 timestamped due to the limited information of the eICU and MIMIC-III databases. Fourth, the
482 predictors may need more effort on improving the recall score – with well-performed AUC and
483 precision, the recall is not so promising due to some positive samples (AKI) are misclassified.
484 Fifth, although we used two non-overlapping general population ICU databases, both are from
485 the United States hospitals. Thus, cross-national and multi-background validations are necessary.
486 Sixth, we only evaluated two feature windows (6 and 12 hours), two gap windows (6 and 12
487 hours) and one prediction window (1 hour) due to the data and computational limitation. In the
488 future, we may put more efforts on various time-series models. Last but not the least, study
489 design in this work relied on retrospective data investigation which may cause missing some
490 important information than a prospective study and could not have any result about the impact
491 analysis between model prediction and patients’ outcomes. Therefore, the model also requires
492 further external validation based on different background populations and may only be used
493 inside the research arena.

494

495 **Conclusion**

496 Recent reviews and comments showed that the prediction of AKI in the ICU is relatively
497 difficult and with several limitations. To answer these limitations and problems, We have
498 developed an artificial intelligence assisted early warning model for predicting the onset of AKI
499 within the first week of ICU stay, which identified by using both oliguric and SCr diagnostic
500 criteria in multi-center eICU database and externally validated in single-center MIMIC-III
501 database. We integrated most routinely available ICU data and demonstrated the model with 6
502 hours feature and 6 hours gap developed by AdaBoost achieved optimal performances measured
503 with AUC, precision, and recall among 8 prevalent machine learning algorithms.

504

505 **Author Contributions:**

506 SH, CF designed the study, conducted the data collection, data analysis, data interpretation, and
507 wrote the manuscript. SH, CF conducted the data analysis, data interpretation and wrote the
508 manuscript. LC and WW conducted the data interpretation and reviewed the manuscript. ZZ and
509 TL designed the study and reviewed the manuscript. KL, LW, and XC conducted the data
510 interpretation and reviewed the manuscript.

511

512 **Data Availability Statement**

513 Datasets are used from the publically available collection of eICU® Collaborative Research
514 Database (eICU, <https://eicu-crd.mit.edu/>) (v1.2) and Medical Information Mart for Intensive
515 Care III (MIMIC-III, <https://mimic.physionet.org/>) (v1.4).

516

517 **Code Availability Statement**

518 Experiments were conducted with two open-source libraries: the machine learning library scikit-
519 learn (<https://scikit-learn.org/stable/>), PyTorch (<https://pytorch.org/>), which provides the public
520 implementations of all listed algorithms. Libraries were called from python 3.7 environment
521 (<https://www.python.org/>).

522

523 **Competing interests**

524 The authors declare no conflict of interests.

525 **References**

- 526 1. Bellomo, R, Kellum, JA, Ronco, C: Acute kidney injury. *Lancet*, 380: 756-766, 2012.
- 527 2. Druml, W, Lenz, K, Laggner, AN: Our paper 20 years later: from acute renal failure to acute
528 kidney injury-the metamorphosis of a syndrome. *Intens Care Med*, 41: 1941-1949, 2015.
- 529 3. Joannidis, M, Metnitz, B, Bauer, P, Schusterschitz, N, Moreno, R, Druml, W, Metnitz, PGH:
530 Acute kidney injury in critically ill patients classified by AKIN versus RIFLE using the
531 SAPS 3 database. *Intens Care Med*, 35: 1692-1702, 2009.
- 532 4. Nisula, S, Kaukonen, KM, Vaara, ST, Korhonen, AM, Poukkanen, M, Karlsson, S, Haapio, M,
533 Inkinen, O, Parviainen, I, Suojaranta-Ylinen, R, Laurila, JJ, Tenhunen, J, Reinikainen, M,
534 Ala-Kokko, T, Ruokonen, E, Kuitunen, A, Pettila, V, Grp, FS: Incidence, risk factors and
535 90-day mortality of patients with acute kidney injury in Finnish intensive care units: the
536 FINNAKI study. *Intens Care Med*, 39: 420-428, 2013.
- 537 5. Susantitaphong, P, Cruz, DN, Cerda, J, Abulfaraj, M, Alqahtani, F, Koulouridis, I, Jaber, BL,
538 Nephrology, AS: World Incidence of AKI: A Meta-Analysis. *Clin J Am Soc Nephro*, 8:
539 1482-1493, 2013.
- 540 6. Sutherland, SM, Chawla, LS, Kane-Gill, SL, Hsu, RK, Kramer, AA, Goldstein, SL, Kellum,
541 JA, Ronco, C, Bagshaw, SM, Group, AC: Utilizing electronic health records to predict
542 acute kidney injury risk and outcomes: workgroup statements from the 15(th) ADQI
543 Consensus Conference. *Can J Kidney Health Dis*, 3: 11, 2016.
- 544 7. Perinel, S, Vincent, F, Lautrette, A, Dellamonica, J, Mariat, C, Zeni, F, Cohen, Y, Tardy, B,
545 Souweine, B, Darmon, M: Transient and Persistent Acute Kidney Injury and the Risk of

- 546 Hospital Mortality in Critically Ill Patients: Results of a Multicenter Cohort Study. *Crit*
547 *Care Med*, 43: e269-275, 2015.
- 548 8. Chertow, GM, Burdick, E, Honour, M, Bonventre, JV, Bates, DW: Acute kidney injury,
549 mortality, length of stay, and costs in hospitalized patients. *J Am Soc Nephrol*, 16: 3365-
550 3370, 2005.
- 551 9. Levy, EM, Viscoli, CM, Horwitz, RI: The effect of acute renal failure on mortality. A cohort
552 analysis. *JAMA*, 275: 1489-1494, 1996.
- 553 10. Bellazzi, R, Zupan, B: Predictive data mining in clinical medicine: current issues and
554 guidelines. *Int J Med Inform*, 77: 81-97, 2008.
- 555 11. Prowle, JR, Liu, YL, Licari, E, Bagshaw, SM, Egi, M, Haase, M, Haase-Fielitz, A, Kellum,
556 JA, Cruz, D, Ronco, C, Tsutsui, K, Uchino, S, Bellomo, R: Oliguria as predictive
557 biomarker of acute kidney injury in critically ill patients. *Crit Care*, 15, 2011.
- 558 12. Ostermann, M, Joannidis, M: Acute kidney injury 2016: diagnosis and diagnostic workup.
559 *Crit Care*, 20, 2016.
- 560 13. Lehner, GF, Forni, LG, Joannidis, M: Oliguria and Biomarkers of Acute Kidney Injury: Star
561 Struck Lovers or Strangers in the Night? *Nephron*, 134: 183-190, 2016.
- 562 14. Waikar, SS, Bonventre, JV: Creatinine Kinetics and the Definition of Acute Kidney Injury.
563 *Journal of the American Society of Nephrology*, 20: 672-679, 2009.
- 564 15. Bove, T, Zangrillo, A, Guarracino, F, Alvaro, G, Persi, B, Maglioni, E, Galdieri, N, Comis, M,
565 Caramelli, F, Pasero, DC, Pala, G, Renzini, M, Conte, M, Paternoster, G, Martinez, B,
566 Pinelli, F, Frontini, M, Zucchetti, MC, Pappalardo, F, Amantea, B, Camata, A, Pisano, A,

- 567 Verdecchia, C, Dal Checco, E, Cariello, C, Faita, L, Baldassarri, R, Scandroglio, AM,
568 Saleh, O, Lembo, R, Calabro, MG, Bellomo, R, Landoni, G: Effect of Fenoldopam on
569 Use of Renal Replacement Therapy Among Patients With Acute Kidney Injury After
570 Cardiac Surgery A Randomized Clinical Trial. *Jama-J Am Med Assoc*, 312: 2244-2253,
571 2014.
- 572 16. Kashani, K, Cheungpasitporn, W, Ronco, C: Biomarkers of acute kidney injury: the pathway
573 from discovery to clinical adoption. *Clin Chem Lab Med*, 55: 1074-1089, 2017.
- 574 17. Ho, JL, Tangri, N, Komenda, P, Kaushal, A, Sood, M, Brar, R, Gill, K, Walker, S,
575 MacDonald, K, Hiebert, BM, Arora, RC, Rigatto, C: Urinary, Plasma, and Serum
576 Biomarkers' Utility for Predicting Acute Kidney Injury Associated With Cardiac Surgery
577 in Adults: A Meta-analysis. *Am J Kidney Dis*, 66: 993-1005, 2015.
- 578 18. Dewitte, A, Joannes-Boyau, O, Sidobre, C, Fleureau, C, Bats, ML, Derache, P, Leuillet, S,
579 Ripoche, J, Combe, C, Ouattara, A: Kinetic eGFR and Novel AKI Biomarkers to Predict
580 Renal Recovery. *Clin J Am Soc Nephro*, 10: 1900-1910, 2015.
- 581 19. Darmon, M, Ostermann, M, Joannidis, M: Predictions are difficult ... especially about AKI.
582 *Intens Care Med*, 43: 932-934, 2017.
- 583 20. (UK), NCGC: Acute Kidney Injury: Prevention, Detection and Management Up to the Point
584 of Renal Replacement Therapy. In: *Acute Kidney Injury: Prevention, Detection and*
585 *Management Up to the Point of Renal Replacement Therapy*. London, 2013.
- 586 21. Mehta, RL, Cerda, J, Burdmann, EA, Tonelli, M, Garcia-Garcia, G, Jha, V, Susantitaphong, P,
587 Rocco, M, Vanholder, R, Sever, MS, Cruz, D, Jaber, B, Lameire, NH, Lombardi, R,

- 588 Lewington, A, Feehally, J, Finkelstein, F, Levin, N, Pannu, N, Thomas, B, Aronoff-
589 Spencer, E, Remuzzi, G: International Society of Nephrology's Oby25 initiative for acute
590 kidney injury (zero preventable deaths by 2025): a human rights case for nephrology.
591 *Lancet*, 385: 2616-2643, 2015.
- 592 22. Huen, SC, Parikh, CR: Predicting acute kidney injury after cardiac surgery: a systematic
593 review. *Ann Thorac Surg*, 93: 337-347, 2012.
- 594 23. Silver, SA, Shah, PM, Chertow, GM, Harel, S, Wald, R, Harel, Z: Risk prediction models for
595 contrast induced nephropathy: systematic review. *BMJ*, 351: h4395, 2015.
- 596 24. Caragata, R, Wyssusek, KH, Kruger, P: Acute kidney injury following liver transplantation: a
597 systematic review of published predictive models. *Anaesth Intensive Care*, 44: 251-261,
598 2016.
- 599 25. Wilson, T, Quan, S, Cheema, K, Zarnke, K, Quinn, R, de Koning, L, Dixon, E, Pannu, N,
600 James, MT: Risk prediction models for acute kidney injury following major noncardiac
601 surgery: systematic review. *Nephrol Dial Transplant*, 31: 231-240, 2016.
- 602 26. Flechet, M, Guiza, F, Schetz, M, Wouters, P, Vanhorebeek, I, Derese, I, Gunst, J, Spriet, I,
603 Casaer, M, Van den Berghe, G, Meyfroidt, G: AKIpredictor, an online prognostic
604 calculator for acute kidney injury in adult critically ill patients: development, validation
605 and comparison to serum neutrophil gelatinase-associated lipocalin. *Intensive care*
606 *medicine*, 43: 764-773, 2017.
- 607 27. Pannu, N, Graham, M, Klarenbach, S, Meyer, S, Kieser, T, Hemmelgarn, B, Ye, F, James, M,
608 Investigators, A, the Alberta Kidney Disease, N: A new model to predict acute kidney

- 609 injury requiring renal replacement therapy after cardiac surgery. *CMAJ*, 188: 1076-1083,
610 2016.
- 611 28. Kellum, JA, Sileanu, FE, Murugan, R, Lucko, N, Shaw, AD, Clermont, G: Classifying AKI
612 by Urine Output versus Serum Creatinine Level. *J Am Soc Nephrol*, 26: 2231-2238, 2015.
- 613 29. Kellum, JA: Diagnostic Criteria for Acute Kidney Injury: Present and Future. *Crit Care Clin*,
614 31: 621-632, 2015.
- 615 30. Harris, SK, Lewington, AJP, Harrison, DA, Rowan, KM: Relationship between patients'
616 outcomes and the changes in serum creatinine and urine output and RIFLE classification
617 in a large critical care cohort database. *Kidney Int*, 88: 369-377, 2015.
- 618 31. Malhotra, R, Kashani, KB, Macedo, E, Kim, J, Bouchard, J, Wynn, S, Li, G, Ohno-Machado,
619 L, Mehta, R: A risk prediction score for acute kidney injury in the intensive care unit.
620 *Nephrol Dial Transplant*, 32: 814-822, 2017.
- 621 32. Koyner, JL, Adhikari, R, Edelson, DP, Churpek, MM: Development of a Multicenter Ward-
622 Based AKI Prediction Model. *Clin J Am Soc Nephrol*, 11: 1935-1943, 2016.
- 623 33. Costa e Silva, VT, de Castro, I, Liano, F, Muriel, A, Rodriguez-Palomares, JR, Yu, L:
624 Sequential evaluation of prognostic models in the early diagnosis of acute kidney injury
625 in the intensive care unit. *Kidney Int*, 75: 982-986, 2009.
- 626 34. Al-Jaghbeer, M, Dealmeida, D, Bilderback, A, Ambrosino, R, Kellum, JA: Clinical Decision
627 Support for In-Hospital AKI. *J Am Soc Nephrol*, 29: 654-660, 2018.
- 628 35. Selby, NM, Casula, A, Lamming, L, Stoves, J, Samarasinghe, Y, Lewington, AJ, Roberts, R,
629 Shah, N, Johnson, M, Jackson, N, Jones, C, Lenguerrand, E, McDonach, E, Fluck, RJ,

- 630 Mohammed, MA, Caskey, FJ: An Organizational-Level Program of Intervention for AKI:
631 A Pragmatic Stepped Wedge Cluster Randomized Trial. *J Am Soc Nephrol*, 30: 505-515,
632 2019.
- 633 36. Tomašev, N, Glorot, X, Rae, JW, Zielinski, M, Askham, H, Saraiva, A, Mottram, A, Meyer, C,
634 Ravuri, S, Protsyuk, I, Connell, A, Hughes, CO, Karthikesalingam, A, Cornebise, J,
635 Montgomery, H, Rees, G, Laing, C, Baker, CR, Peterson, K, Reeves, R, Hassabis, D,
636 King, D, Suleyman, M, Back, T, Nielson, C, Ledsam, JR, Mohamed, S: A clinically
637 applicable approach to continuous prediction of future acute kidney injury. *Nature*, 572:
638 116-119, 2019.
- 639 37. Kheterpal, S, Tremper, KK, Heung, M, Rosenberg, AL, Englesbe, M, Shanks, AM, Campbell,
640 DA, Jr.: Development and validation of an acute kidney injury risk index for patients
641 undergoing general surgery: results from a national data set. *Anesthesiology*, 110: 505-
642 515, 2009.
- 643 38. Drawz, PE, Miller, RT, Sehgal, AR: Predicting hospital-acquired acute kidney injury--a case-
644 controlled study. *Ren Fail*, 30: 848-855, 2008.
- 645 39. Kashani, K, Al-Khafaji, A, Ardiles, T, Artigas, A, Bagshaw, SM, Bell, M, Bihorac, A,
646 Birkhahn, R, Cely, CM, Chawla, LS, Davison, DL, Feldkamp, T, Forni, LG, Gong, MN,
647 Gunnerson, KJ, Haase, M, Hackett, J, Honore, PM, Hoste, EA, Joannes-Boyau, O,
648 Joannidis, M, Kim, P, Koyner, JL, Laskowitz, DT, Lissauer, ME, Marx, G, McCullough,
649 PA, Mullaney, S, Ostermann, M, Rimmele, T, Shapiro, NI, Shaw, AD, Shi, J, Sprague,
650 AM, Vincent, JL, Vinsonneau, C, Wagner, L, Walker, MG, Wilkerson, RG, Zacharowski,
651 K, Kellum, JA: Discovery and validation of cell cycle arrest biomarkers in human acute

- 652 kidney injury. *Crit Care*, 17: R25, 2013.
- 653 40. Chiofalo, C, Chbat, N, Ghosh, E, Eshelman, L, Kashani, K: Automated Continuous Acute
654 Kidney Injury Prediction and Surveillance: A Random Forest Model. *Mayo Clin Proc*, 94:
655 783-792, 2019.
- 656 41. Mandelbaum, T, Lee, J, Scott, DJ, Mark, RG, Malhotra, A, Howell, MD, Talmor, D:
657 Empirical relationships among oliguria, creatinine, mortality, and renal replacement
658 therapy in the critically ill. *Intens Care Med*, 39: 414-419, 2013.
- 659 42. Hodgson, LE, Sarnowski, A, Roderick, PJ, Dimitrov, BD, Venn, RM, Forni, LG: Systematic
660 review of prognostic prediction models for acute kidney injury (AKI) in general hospital
661 populations. *BMJ Open*, 7: e016591, 2017.
- 662 43. Kheterpal, S, Tremper, KK, Englesbe, MJ, O'Reilly, M, Shanks, AM, Fetterman, DM,
663 Rosenberg, AL, Swartz, RD: Predictors of postoperative acute renal failure after
664 noncardiac surgery in patients with previously normal renal function. *Anesthesiology*, 107:
665 892-902, 2007.
- 666 44. Bell, S, Dekker, FW, Vadiveloo, T, Marwick, C, Deshmukh, H, Donnan, PT, Van Diepen, M:
667 Risk of postoperative acute kidney injury in patients undergoing orthopaedic surgery--
668 development and validation of a risk score and effect of acute kidney injury on survival:
669 observational cohort study. *BMJ*, 351: h5639, 2015.
- 670 45. Bedford, M, Stevens, P, Coulton, S, Billings, J, Farr, M, Wheeler, T, Kalli, M, Mottishaw, T,
671 Farmer, C: Development of risk models for the prediction of new or worsening acute
672 kidney injury on or during hospital admission: a cohort and nested study. In:

- 673 *Development of risk models for the prediction of new or worsening acute kidney injury on*
674 *or during hospital admission: a cohort and nested study.* Southampton (UK), 2016.
- 675 46. Forman, DE, Butler, J, Wang, YF, Abraham, WT, O'Connor, CM, Gottlieb, SS, Loh, E,
676 Massie, BM, Rich, MW, Stevenson, LW, Young, JB, Krumholz, HM: Incidence,
677 predictors at admission, and impact of worsening renal function among patients
678 hospitalized with heart failure. *J Am Coll Cardiol*, 43: 61-67, 2004.
- 679 47. Hodgson, LE, Dimitrov, BD, Roderick, PJ, Venn, R, Forni, LG: Predicting AKI in emergency
680 admissions: an external validation study of the acute kidney injury prediction score (APS).
681 *Bmj Open*, 7, 2017.
- 682 48. Goldberger, AL, Amaral, LA, Glass, L, Hausdorff, JM, Ivanov, PC, Mark, RG, Mietus, JE,
683 Moody, GB, Peng, CK, Stanley, HE: PhysioBank, PhysioToolkit, and PhysioNet:
684 components of a new research resource for complex physiologic signals. *Circulation*, 101:
685 E215-220, 2000.
- 686 49. Pollard, TJ, Johnson, AEW, Raffa, JD, Celi, LA, Mark, RG, Badawi, O: The eICU
687 Collaborative Research Database, a freely available multi-center database for critical care
688 research. *Sci Data*, 5: 180178, 2018.
- 689 50. Johnson, AE, Pollard, TJ, Shen, L, Lehman, LW, Feng, M, Ghassemi, M, Moody, B,
690 Szolovits, P, Celi, LA, Mark, RG: MIMIC-III, a freely accessible critical care database.
691 *Sci Data*, 3: 160035, 2016.
- 692 51. Johnson, AE, Stone, DJ, Celi, LA, Pollard, TJ: The MIMIC Code Repository: enabling
693 reproducibility in critical care research. *J Am Med Inform Assoc*, 25: 32-39, 2018.

- 694 52. Raita, Y, Goto, T, Faridi, MK, Brown, DFM, Camargo, CA, Jr., Hasegawa, K: Emergency
695 department triage prediction of clinical outcomes using machine learning models. *Crit*
696 *Care*, 23: 64, 2019.
- 697 53. Serpa Neto, A, Deliberato, RO, Johnson, AEW, Bos, LD, Amorim, P, Pereira, SM, Cazati,
698 DC, Cordioli, RL, Correa, TD, Pollard, TJ, Schettino, GPP, Timenetsky, KT, Celi, LA,
699 Pelosi, P, Gama de Abreu, M, Schultz, MJ, Investigators, PN: Mechanical power of
700 ventilation is associated with mortality in critically ill patients: an analysis of patients in
701 two observational cohorts. *Intensive Care Med*, 44: 1914-1922, 2018.
- 702 54. Zhang, Z, Ho, KM, Hong, Y: Machine learning for the prediction of volume responsiveness
703 in patients with oliguric acute kidney injury in critical care. *Crit Care*, 23: 112, 2019.
- 704 55. Thomas, ME, Blaine, C, Dawnay, A, Devonald, MA, Ftouh, S, Laing, C, Latchem, S,
705 Lewington, A, Milford, DV, Ostermann, M: The definition of acute kidney injury and its
706 use in practice. *Kidney Int*, 87: 62-73, 2015.
- 707 56. Thomas, ME, Blaine, C, Dawnay, A, Devonald, MAJ, Ftouh, S, Laing, C, Latchem, S,
708 Lewington, A, Milford, DV, Ostermann, M: The definition of acute kidney injury and its
709 use in practice. *Kidney Int*, 87: 62-73, 2015.
- 710 57. Wang, YN, Cheng, H, Yue, T, Chen, YP: Derivation and validation of a prediction score for
711 acute kidney injury in patients hospitalized with acute heart failure in a Chinese cohort.
712 *Nephrology (Carlton)*, 18: 489-496, 2013.
- 713 58. Forman, DE, Butler, J, Wang, Y, Abraham, WT, O'Connor, CM, Gottlieb, SS, Loh, E, Massie,
714 BM, Rich, MW, Stevenson, LW, Young, JB, Krumholz, HM: Incidence, predictors at

- 715 admission, and impact of worsening renal function among patients hospitalized with heart
716 failure. *J Am Coll Cardiol*, 43: 61-67, 2004.
- 717 59. Mohamadlou, H, Lynn-Palevsky, A, Barton, C, Chettipally, U, Shieh, L, Calvert, J, Saber, NR,
718 Das, R: Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using
719 Electronic Health Record Data. *Can J Kidney Health Dis*, 5: 2054358118776326, 2018.
- 720 60. Hintze, JL, Nelson, RDJTAS: Violin plots: a box plot-density trace synergism. 52: 181-184,
721 1998.
- 722 61. Huang, Z, Zhan, X, Xiang, S, Johnson, TS, Helm, B, Yu, CY, Zhang, J, Salama, P, Rizkalla,
723 M, Han, Z, Huang, K: SALMON: Survival Analysis Learning With Multi-Omics Neural
724 Networks on Breast Cancer. *Front Genet*, 10: 166, 2019.
- 725 62. Breiman, L: *Classification and regression trees*, Routledge, 2017.
- 726 63. Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M,
727 Prettenhofer, P, Weiss, R, Dubourg, VJ: *Scikit-learn: Machine learning in Python*.
728 12: 2825-2830, 2011.
- 729 64. Revelle, WR: *psych: Procedures for personality and psychological research*. 2017.
- 730 65. Holland, M, Wolfe, D: *Nonparametric statistical methods*. John Wiley & Sons, New York,
731 1973 pp 185–194.
- 732 66. Suresh, H, Hunt, N, Johnson, A, Celi, LA, Szolovits, P, Ghassemi, M: Clinical intervention
733 prediction and understanding with deep neural networks. *Machine Learning for
734 Healthcare Conference*. 2017 pp 322-337.

- 735 67. Chanley, VA, Rudolph, TJ, Rahn, WM: The origins and consequences of public trust in
736 government: a time series analysis. *Public Opin Q*, 64: 239-256, 2000.
- 737 68. Chiang, AC: Fundamental methods of mathematical economics. *Auckland (New Zealand)*
738 *McGraw-Hill*, 1984.
- 739 69. Twisk, J, de Vente, WJJoce: Attrition in longitudinal studies: how to deal with missing data.
740 55: 329-337, 2002.
- 741 70. Shah, JS, Rai, SN, DeFilippis, AP, Hill, BG, Bhatnagar, A, Brock, GN: Distribution based
742 nearest neighbor imputation for truncated high dimensional data with applications to pre-
743 clinical and clinical metabolomics studies. *Bmc Bioinformatics*, 18, 2017.
- 744 71. Kim, KY, Kim, BJ, Yi, GS: Reuse of imputed data in microarray analysis increases
745 imputation efficiency. *Bmc Bioinformatics*, 5, 2004.
- 746 72. Komorowski, M, Celi, L, Badawi, O, Gordon, AC, Faisal, AA: The Artificial Intelligence
747 Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*, 24:
748 1716+, 2018.
- 749 73. Liaw, A, Wiener, MJRn: Classification and regression by randomForest. 2: 18-22, 2002.
- 750 74. Freund, Y, Schapire, R, Abe, NJJ-JSFAI: A short introduction to boosting. 14: 1612, 1999.
- 751 75. Friedman, JHJAos: Greedy function approximation: a gradient boosting machine. 1189-1232,
752 2001.
- 753 76. Hochreiter, S, Schmidhuber, J: Long short-term memory. *Neural Comput*, 9: 1735-1780,
754 1997.

- 755 77. Bradley, APJPr: The use of the area under the ROC curve in the evaluation of machine
756 learning algorithms. *30*: 1145-1159, 1997.
- 757 78. Davis, J, Goadrich, M: The relationship between Precision-Recall and ROC curves.
758 *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006 pp
759 233-240.
- 760 79. Goutte, C, Gaussier, E: A Probabilistic Interpretation of Precision, Recall and F-Score, with
761 Implication for Evaluation. Berlin, Heidelberg, Springer Berlin Heidelberg, 2005 pp 345-
762 359.
- 763 80. Altman, DG, Bland, JM: Diagnostic tests 2: Predictive values. *BMJ*, *309*: 102, 1994.
- 764 81. Altman, DG, Bland, JM: Diagnostic tests. 1: Sensitivity and specificity. *BMJ*, *308*: 1552,
765 1994.
- 766 82. Vincent, JL: Critical care - where have we been and where are we going? *Crit Care*, *17*, 2013.
- 767 83. Abul-Husn, NS, Kenny, EE: Personalized Medicine and the Power of Electronic Health
768 Records. *Cell*, *177*: 58-69, 2019.
- 769 84. Jamoom, EW, Yang, N, Hing, E: Adoption of Certified Electronic Health Record Systems
770 and Electronic Information Sharing in Physician Offices: United States, 2013 and 2014.
771 *NCHS Data Brief*: 1-8, 2016.
- 772 85. Hoste, E, Bagshaw, S, Bellomo, R, Cely, C, Colman, R, Cruz, D, Edipidis, K, Forni, L,
773 Gomersall, C, Govil, D, Honore, PM, Joannes-Boyau, O, Joannidis, M, Korhonen, AM,
774 Lavrentieva, A, Mehta, R, Palevsky, P, Roessler, E, Ronco, C, Uchino, S, Vazquez, J,
775 Vidal Andrade, E, Webb, S, Kellum, J: Epidemiology of acute kidney injury in critically

- 776 ill patients: the multinational AKI-EPI study. *Intens Care Med*, 41: 1411-1423, 2015.
- 777 86. Schetz, M, Gunst, J, Van den Berghe, G: The impact of using estimated GFR versus
778 creatinine clearance on the evaluation of recovery from acute kidney injury in the ICU.
779 *Intens Care Med*, 40: 1709-1717, 2014.
- 780 87. Geri, G, Guillemet, L, Dumas, F, Charpentier, J, Antona, M, Lemiale, V, Bouguin, W,
781 Lamhaut, L, Mira, JP, Vinsonneau, C, Cariou, A: Acute kidney injury after out-of-hospital
782 cardiac arrest: risk factors and prognosis in a large cohort. *Intens Care Med*, 41: 1273-
783 1280, 2015.
- 784 88. O'Connor, ME, Kirwan, CJ, Pearse, RM, Prowle, JR: Incidence and associations of acute
785 kidney injury after major abdominal surgery. *Intens Care Med*, 42: 521-530, 2016.
- 786 89. Smith, LE, Smith, DK, Blume, JD, Siew, ED, Billings, FT: Latent variable modeling
787 improves AKI risk factor identification and AKI prediction compared to traditional
788 methods. *Bmc Nephrol*, 18, 2017.
- 789 90. Rind, DM, Safran, C, Phillips, RS, Wang, Q, Calkins, DR, Delbanco, TL, Bleich, HL, Slack,
790 WV: Effect of computer-based alerts on the treatment and outcomes of hospitalized
791 patients. *Arch Intern Med*, 154: 1511-1517, 1994.
- 792 91. McCoy, AB, Waitman, LR, Gadd, CS, Danciu, I, Smith, JP, Lewis, JB, Schildcrout, JS,
793 Peterson, JF: A computerized provider order entry intervention for medication safety
794 during acute kidney injury: a quality improvement report. *Am J Kidney Dis*, 56: 832-841,
795 2010.
- 796 92. Cho, A, Lee, JE, Yoon, JY, Jang, HR, Huh, W, Kim, YG, Kim, DJ, Oh, HY: Effect of an

- 797 electronic alert on risk of contrast-induced acute kidney injury in hospitalized patients
798 undergoing computed tomography. *Am J Kidney Dis*, 60: 74-81, 2012.
- 799 93. Hooper, MH, Weavind, L, Wheeler, AP, Martin, JB, Gowda, SS, Semler, MW, Hayes, RM,
800 Albert, DW, Deane, NB, Nian, H, Mathe, JL, Nadas, A, Sztipanovits, J, Miller, A,
801 Bernard, GR, Rice, TW: Randomized trial of automated, electronic monitoring to
802 facilitate early detection of sepsis in the intensive care unit*. *Crit Care Med*, 40: 2096-
803 2101, 2012.
- 804 94. van der Sijs, H, Aarts, J, Vulto, A, Berg, M: Overriding of drug safety alerts in computerized
805 physician order entry. *J Am Med Inform Assoc*, 13: 138-147, 2006.
- 806 95. Rosenberg, SN, Sullivan, M, Juster, IA, Jacques, J: Overrides of medication alerts in
807 ambulatory care. *Arch Intern Med*, 169: 1337; author reply 1338, 2009.
- 808 96. Henry, KE, Hager, DN, Pronovost, PJ, Saria, S: A targeted real-time early warning score
809 (TREWScore) for septic shock. *Sci Transl Med*, 7, 2015.
- 810 97. ElMoaqet, H, Tilbury, DM, Ramachandran, SK: Multi-Step Ahead Predictions for Critical
811 Levels in Physiological Time Series. *Ieee T Cybernetics*, 46: 1704-1714, 2016.
- 812 98. Saria, S, Rajani, AK, Gould, J, Koller, D, Penn, AA: Integration of Early Physiological
813 Responses Predicts Later Illness Severity in Preterm Infants. *Sci Transl Med*, 2, 2010.
- 814 99. Neto, AS, Deliberato, RO, Johnson, AEW, Bos, LD, Amorim, P, Pereira, SM, Cazati, DC,
815 Cordioli, RL, Correa, TD, Pollard, TJ, Schettino, GPP, Timenetsky, KT, Celi, LA, Pelosi,
816 P, de Abreu, MG, Schultz, MJ, Investigators, PN: Mechanical power of ventilation is
817 associated with mortality in critically ill patients: an analysis of patients in two

- 818 observational cohorts. *Intens Care Med*, 44: 1914-1922, 2018.
- 819 100. Ghassemi, M, Pimentel, MA, Naumann, T, Brennan, T, Clifton, DA, Szolovits, P, Feng, M:
820 A Multivariate Timeseries Modeling Approach to Severity of Illness Assessment and
821 Forecasting in ICU with Sparse, Heterogeneous Clinical Data. *Proc Conf AAAI Artif*
822 *Intell*, 2015: 446-453, 2015.
- 823 101. Ghassemi, M, Wu, M, Hughes, MC, Szolovits, P, Doshi-Velez, F: Predicting intervention
824 onset in the ICU with switching state space models. *AMIA Jt Summits Transl Sci Proc*,
825 2017: 82-91, 2017.
- 826 102. Danziger, J, Chen, KP, Lee, J, Feng, M, Mark, RG, Celi, LA, Mukamal, KJ: Obesity, Acute
827 Kidney Injury, and Mortality in Critical Illness. *Crit Care Med*, 44: 328-334, 2016.
- 828 103. Siew, ED, Matheny, ME, Ikizler, TA, Lewis, JB, Miller, RA, Waitman, LR, Go, AS, Parikh,
829 CR, Peterson, JF: Commonly used surrogates for baseline renal function affect the
830 classification and prognosis of acute kidney injury. *Kidney Int*, 77: 536-542, 2010.
- 831

832 **Figure Legends**

833 Figure 1. Model settings and prediction for a single admission. Timestamp is at one-hour interval.
834 Feature collecting window are the discretized time-series features (6 hours or 12 hours) as inputs
835 of machine learning algorithms for predicting the label of AKI after certain gaps (6 hours or 12
836 hours). We experimented on four different model configurations: model 6-6, model 6-12, model
837 12-6, and model 12-12.

838 Figure 2. The study overview of the artificial intelligence-assisted AKI early warning system,
839 including (1) Data extraction, (2) Data pre-processing and missing data imputation, (3) dataset
840 splitting, (4) supervised machine learning method to predict AKI, (5) external validation, and (6)
841 performance evaluation and important feature identification.

842 Figure 3. Study cohort selection workflow of eICU (A) and MIMIC-III database (B) based on the
843 designed inclusion and exclusion criteria.

844 Figure 4. Boxplot of all five-folds recall results across four different configurations of the time-
845 series models (model 6-6, model 6-12, model 12-6, model 12-12) and eight different machine
846 learning algorithms on eICU testing set (A) and external validation database MIMIC-III (B).
847 Abbreviations: DT: Decision Tree; GBM: Gradient Boosting Machine; LR: Logistic Regression;
848 NN: Neural Networks; NN (L2): Neural Networks with L2 regularization; RF: Random Forest;
849 LSTM: Long short-term memory networks.

850 Figure 5. Model performances (AUC) violin plot of four different configurations of the models
851 (model 6-6, model 6-12, model 12-6, model 12-12) across eight different algorithms with five-
852 fold cross-validation. Black solid horizontal lines in each violin plot are the interquartile range
853 lines. Red points and lines indicate the mean performances (AUC) over four models.

854 Figure 6. Identified top important features in eICU database. Values and rankings are based on
855 time-series model 6-6 and the averaged Gini importance from AdaBoost algorithm. Features are
856 colored group-wisely according to the categories listed in Table 1. The higher value indicates the
857 higher significance of the feature.

858

859 **Table Legends**

860 Table 1. Variables and categorized groups. ¹Body Mass Index: Weight (kg) / Height (meter) *
861 Height (meter). BMI then converted to classes according to following criteria: BMI < 18.5: class
862 = 0; $18.5 \leq \text{BMI} < 23$: class = 1; $23 \leq \text{BMI} < 25$: class = 2; $25 \leq \text{BMI} < 40$: class = 3; BMI ≥ 40 :
863 class = 4. ²P_F_ratio: PO₂/FiO₂. ³SOFA: Sequential Organ Failure Assessment.

864 Table 2. Description of the datasets. Ethnicity of Black includes African American, Cape
865 Verdean, Haitian, and African. Ethnicity of Hispanic includes Latino (central American), Latino
866 (Cuban), Latino (Puerto Rican), Latino (Honduran), Latino (Guatemalan), Latino (Mexican),
867 Latino (Dominican), Latino (Salvadoran), Latino (Colombian), and Portuguese. Ethnicity of
868 Asian includes Vietnamese, Thai, Asian Indian, middle Eastern, Korean, Chinese, Filipino,
869 Cambodian, Japanese, Asian other. White includes Eastern European, Brazilian, Russian, and
870 other European. Ethnicity of Other includes Unknown, not specified, multi race, patient declined
871 to answer, and unable to obtain. s.d.: Standard deviation in short. IQR: interquartile range in
872 short.

873 Table 3. Comparison of model performance (mean values across five-fold validation results) in
874 eICU internal validation. Abbreviations: AUC: area under the ROC curve. L2 stands for Ridge
875 regularization.

876 Table 4. Comparison of model performance (mean values across five-fold validation results) in
877 MIMIC-III cohort for external validation. Abbreviations: AUC: area under the ROC curve. L2
878 stands for Ridge regularization.

879 Table 5. Secondary outcomes of eICU and MIMIC-III databases. s.d.: standard deviation in short.

880 IQR: interquartile range in short. *P-value is derived according to alternative hypothesis: true

881 Spearman ρ is not equal to 0 when comparing the variable with the onset of AKI.

882 Table 6. Pairwise paired t-test comparison among the five-fold cross-validation AUC results over

883 four different time-series models.

884 Table S1. eICU and MIMIC-III databases exclusion criteria.

885 Table S2. Most commonly missing features according to all timestamps in eICU database. Total

886 missing count was the count of total missing occurrence of the associated feature. Missing ratio

887 was the ratio of missing observations to the entire observations (427,527 timestamps in total). All

888 abbreviations are referred to eICU and MIMIC-III databases. Features with high missing ratio

889 were dropped.

890 Table S3. Comparison of other performance metrics (mean values across five-fold validation

891 results) on eICU internal validation. Abbreviations: NPV: negative predictive value. L2 stands

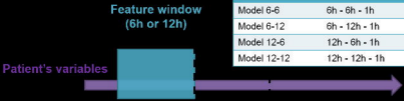
892 for Ridge regularization.

893 Table S4. Comparison of other performance metrics (mean values across five-fold validation

894 results) on MIMIC-III cohort for external validation. Abbreviations: NPV: negative predictive

895 value. L2 stands for Ridge regularization.

Model	Feature window - gap window - prediction window
Model 6-6	6h - 6h - 1h
Model 6-12	6h - 12h - 1h
Model 12-6	12h - 6h - 1h
Model 12-12	12h - 12h - 1h





All patients in the eICU (v1.2) and MIMIC-III (v1.4) databases are included.



6 demographic information and static variables.



31 time-varying variables of vital signs, laboratory values, and fluid balance.



13 variables of additional respiratory and hemodynamic support, primary diagnosis, and comorbidities.



2 other variables.

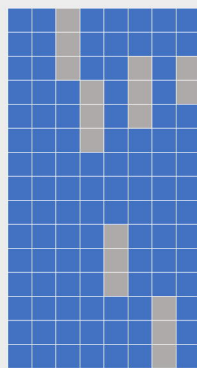


For one admission (with one unique ID)

All variables



Hours

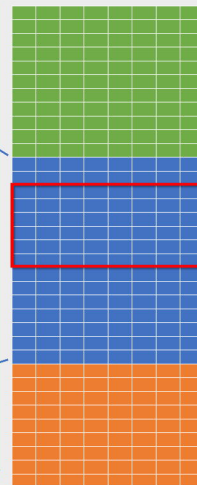


Data after expansion

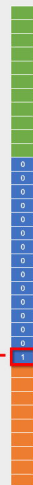
Missing data to be imputed

Unique admissions

All variables



Label: AKI



Prediction window



• Validation dataset: 20% of eICU

• Testing dataset: 20% of eICU

Models

Logistic Regression
Decision Tree (DT)
Random Forest (RF)
AdaBoost
Gradient Boosting Machine (GBM)
Neural Networks
Long short-term memory (LSTM) networks

External validation dataset: 100% of MIMIC-III

Model	Feature window - gap window - prediction window
-------	---

Model 6-6	6h - 6h - 1h
-----------	--------------

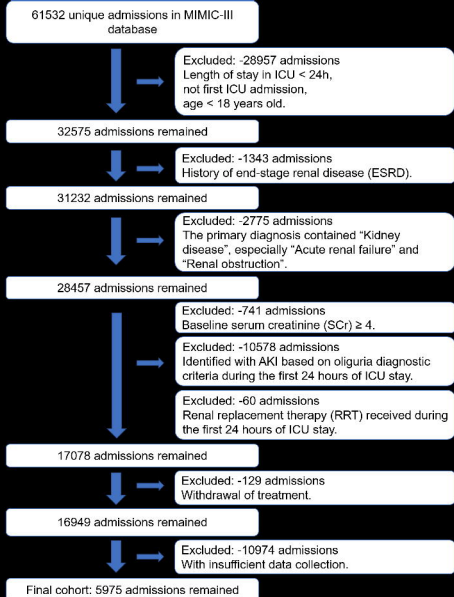
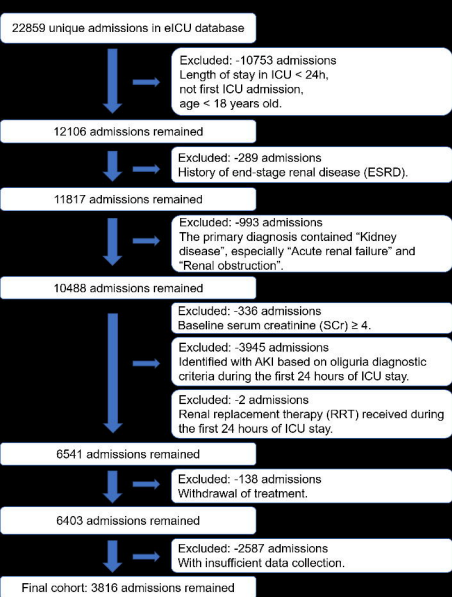
Model 6-12	6h - 12h - 1h
------------	---------------

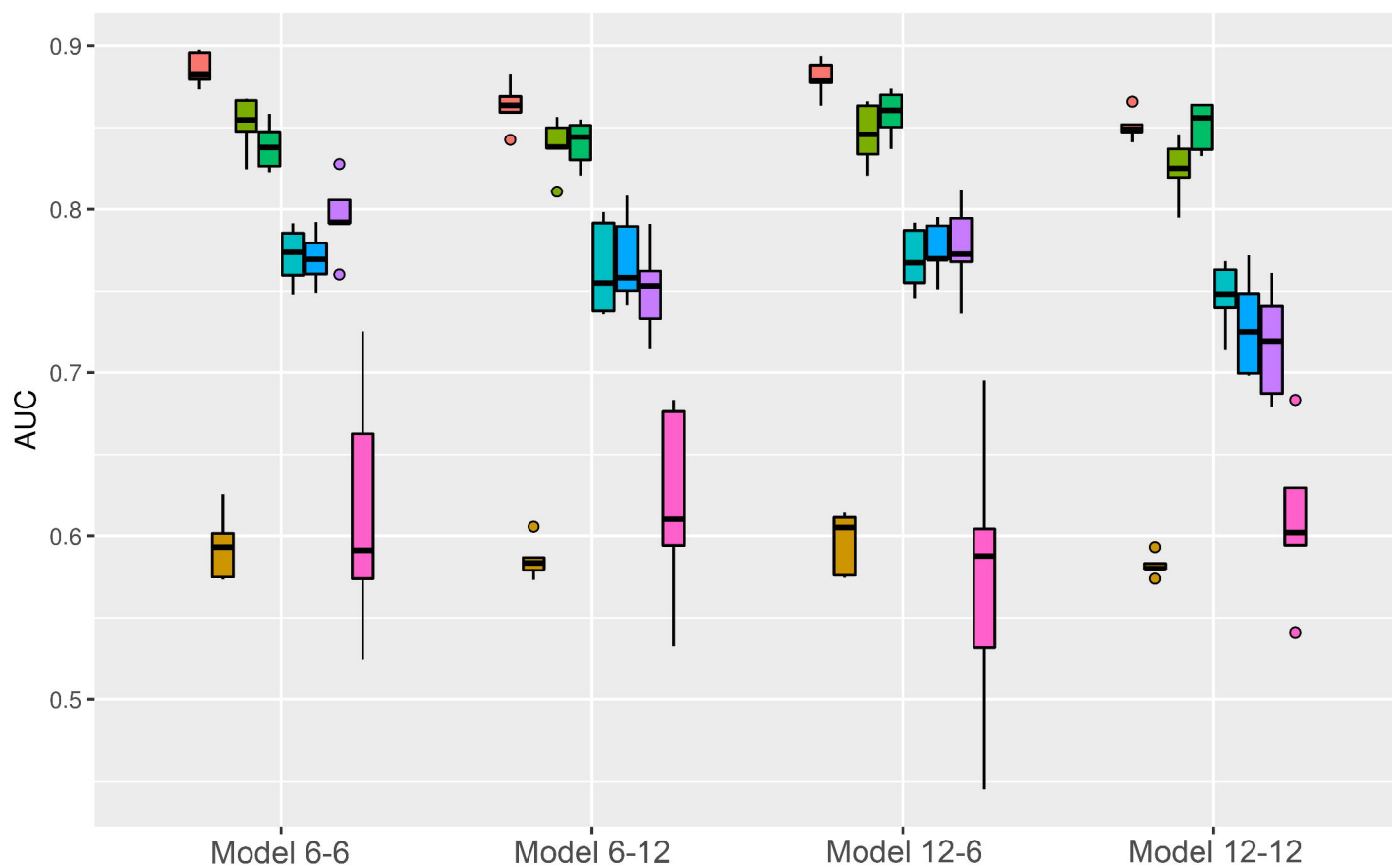
Model 12-6	12h - 6h - 1h
------------	---------------

Model 12-12	12h - 12h - 1h
-------------	----------------



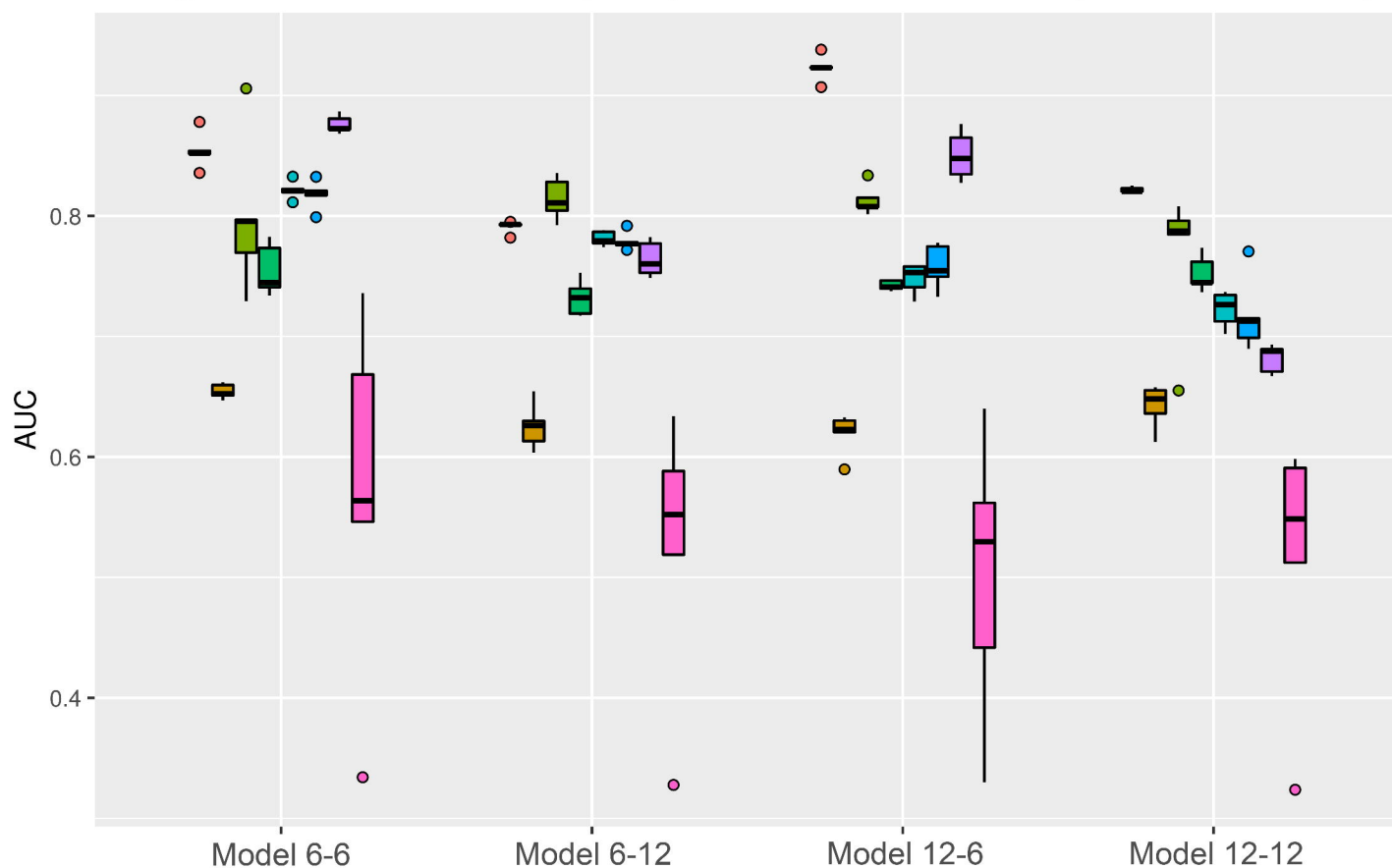
- AUC
- Precision
- Recall
- Gini importance
-



A Algorithm Performances (AUC) with Different Models (EICU Test Set)

Algorithms

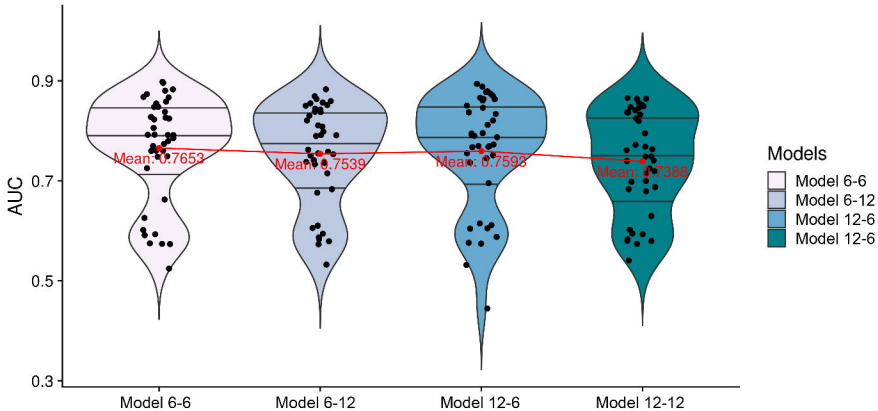
- AdaBoost
- DT
- GBM
- LR (Baseline)
- NN
- NN (L2)
- RF
- LSTM

B Algorithm Performances (AUC) with Different Models (MIMIC Database)

Algorithms

- AdaBoost
- DT
- GBM
- LR (Baseline)
- NN
- NN (L2)
- RF
- LSTM

Model Performances (AUC) Across Different Algorithms



Feature Importance Ranked by AdaBoost Algorithm

