1 **Title**: Machine Learning to Predict Mortality and Critical Events in COVID-19 Positive
2 New York City Patients
3
4 **Short Title:** Predicting COVID-19 Outcomes With Machine Learning
5
6 **One-Sentence Summary**:
7 We identify clinical features that robustly predict mortality and critical events in a large
8 cohort of COVID-19 positive patients in New York City.
9

10 **Authors:**
11 Akhil Vaid[1], Sulaiman Somani[1+], Adam J Russak[1,2+], Jessica K De Freitas[1,3], Fayzan F
12 Chaudhry[1,3], Ishan Paranjpe[1], Kipp W Johnson[3], Samuel J Lee[1], Riccardo Miotto[1,3],
13 Shan Zhao[1,4], Noam D Beckmann[3], Nidhi Naik[1], Kodi Arfer[5], Arash Kia[6,7], Prem
14 Timsina[6,7], Anuradha Lala[6,8], Manish Paranjpe[9], Patricia Glowe[1], Eddye Golden[1],
15 Matteo Danieletto[1], Manbir Singh[1], Dara Meyer[3], Paul F O'Reilly[3,10,11], Laura H
16 Huckins[3,10,11], Patricia Kovatch[12], Joseph Finkelstein[6], Robert M Freeman[6,7], Edgar
17 Argulian[13,14], Andrew Kasarskis[3,6,15,16], Bethany Percha[2], Judith A Aberg[2,17], Emilia
18 Bagiella[7,8,], Carol R Horowitz[2,6], Barbara Murphy[2], Eric J Nestler[18,19], Eric E Schadt[3,15],
19 Judy H Cho[20], Carlos Cordon-Cardo[21], Valentin Fuster[8,13,14], Dennis S Charney[22], David
20 L Reich[4], Erwin P Bottinger[1,2], Matthew A Levin[4], Jagat Narula[13,14], Zahi A Fayad[23,24],
21 Allan C Just[5*], Alexander W Charney[3,10,11]*, Girish N Nadkarni[1,2,18]*, Benjamin S
22 Glicksberg[1,3]* on behalf of the Mount Sinai Covid Informatics Center (MSCIC).
23
24 + Equal contribution
25 *Co-correspondence
26
27 Affiliations:
28   1. The Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of
29   Medicine at Mount Sinai
30   2. Department of Medicine, Icahn School of Medicine at Mount Sinai
31   3. Department of Genetics and Genomic Sciences, Icahn School of Medicine at
32   Mount Sinai
33   4. Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of
34   Medicine at Mount Sinai
35   5. Department of Environmental Medicine and Public Health, Icahn School of
36   Medicine at Mount Sinai
37   6. Department of Population Health Science and Policy, Icahn School of Medicine
38   at Mount Sinai
39   7. Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount
40   Sinai

8. The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai
9. Harvard Medical School
10. The Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai
11. The Department of Psychiatry, Icahn School of Medicine at Mount Sinai
12. Mount Sinai Data Warehouse, Icahn School of Medicine at Mount Sinai
13. Mount Sinai Heart, Icahn School of Medicine at Mount Sinai
14. Department of Cardiology, Icahn School of Medicine at Mount Sinai
15. Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai
16. Mount Sinai Data Office Icahn School of Medicine at Mount Sinai
17. Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai
18. Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai
19. Friedman Brain Institute, Icahn School of Medicine at Mount Sinai
20. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai
21. Department of Pathology, Icahn School of Medicine at Mount Sinai
22. Office of the Dean, Icahn School of Medicine at Mount Sinai
23. BioMedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai
24. Department of Radiology, Icahn School of Medicine at Mount Sinai

Figures: 3
Tables: 1
Word Count: 3120

Correspondence:

Allan Just, PhD
allan.just@mssm.edu

Alexander Charney, MD, PhD
alexander.charney@icahn.mssm.edu

Girish Nadkarni, MD
girish.nadkarni@mountsinai.org

81   Benjamin Glicksberg, PhD

82   benjamin.glicksberg@mssm.edu

83

84

Abstract

Coronavirus 2019 (COVID-19), caused by the SARS-CoV-2 virus, has become the deadliest pandemic in modern history, reaching nearly every country worldwide and overwhelming healthcare institutions. As of April 20, there have been more than 2.4 million confirmed cases with over 160,000 deaths. Extreme case surges coupled with challenges in forecasting the clinical course of affected patients have necessitated thoughtful resource allocation and early identification of high-risk patients. However, effective methods for achieving this are lacking. In this paper, we use electronic health records from over 3,055 New York City confirmed COVID-19 positive patients across five hospitals in the Mount Sinai Health System and present a decision tree-based machine learning model for predicting in-hospital mortality and critical events. This model is first trained on patients from a single hospital and then externally validated on patients from four other hospitals. We achieve strong performance, notably predicting mortality at 1 week with an AUC-ROC of 0.84. Finally, we establish model interpretability by calculating SHAP scores to identify decisive features, including age, inflammatory markers (procalcitonin and LDH), and coagulation parameters (PT, PTT, D-Dimer). To our knowledge, this is one of the first models with external validation to both predict outcomes in COVID-19 patients with strong validation performance and identify key contributors in outcome prediction that may assist clinicians in making effective patient management decisions.

Introduction

Despite substantial, organized efforts to prevent disease spread, over 2.4 million people have tested positive for SARS-CoV-2 worldwide, and there have been more than 169,000 deaths to date*(1–3)*. As a result of this pandemic, hospitals are being filled beyond capacity and face extreme challenges with regards to personnel staffing, personal protective equipment availability, and ICU bed allocation. Additionally, patients with COVID-19 demonstrate varying symptomatology, making successful and safe patient triaging difficult. While some infected patients are asymptomatic, others suffer from severe acute respiratory distress syndrome, multiorgan failure, and death*(4).* Identification of key patient characteristics that govern the course of disease across large patient cohorts is lacking but important, particularly given the potential it has to aid physicians and hospitals in predicting disease trajectory, to allocate essential resources effectively, and to improve patient outcomes. With these needs in mind, we report the development of a decision tree-based machine learning model trained on electronic health records from patients with confirmed COVID-19 status at a single center in the Mount Sinai Health System in New York City to predict critical events and mortality; validate this algorithm at four other hospital centers; and perform a saliency analysis using SHAP (SHapley Additive exPlanation) values to identify the most important features used by this model for outcome prediction.

Results

Clinical Data Source and Study Population

We retrieved electronic health records for 3,055 COVID-19-positive inpatient admissions at five hospitals between March 9, 2020 and April 11, 2020 within the Mount Sinai Health System (MSHS). These data included patient demographics, past medical history, and admission vitals and labs (Table 1, Supplementary Table 2). Relevant patient events (intubation, discharge to hospice care, or death) were recorded and subsets were constructed at 3, 5, 7, and 10 day intervals after admission (Figure 1). Of these patients, 17.0% to 31.6% had a critical event (intubation, discharge to hospice care, or death) and 6.0% to 21.5% died over the observed time frames (Supplementary Table 1). In contrast, the control group consisted of patients with all other discharge dispositions and those that were still hospitalized.

Classifier training and performance

Given the large number of patients in the analysis and presence of missing variables in the data, we used XGBoost*(5)*, a boosted decision-tree based machine learning (ML) model, to predict either a critical event or death of a patient within the aforementioned time frames. Patients from the Mount Sinai Hospital (MSH) were split into a training and validation set for the model. To increase model generalizability and help minimize bias, the model's performance was assessed on a test set composed entirely of patients from

147 the other hospitals (OH) in the MSHS. While multiple time limits for event occurrence
148 were assessed, the results and discussion in this letter focus predominantly one week
149 after admission. As a control, both simple and generalized additive logistic regression
150 models were trained to assess performance, given their ubiquity as the preferred model
151 in current COVID research pieces.

152

153 After training, the classifier robustly predicted the presence of a critical event at three,
154 five, seven, and 10 days (Figure 2 and Supplementary Table 4) as measured by area
155 under the receiver operating characteristic curve (AUC-ROC = 0.74 for OH, 0.83 for
156 MSH at 1 week) and area under the precision-recall curve (AUPRC = 0.58 for OH, 0.49
157 for MSH at 1 week). We were able to achieve similar performance at predicting critical
158 events at longer intervals, namely 15 and 20 days (Supplementary Table 5). As a
159 baseline comparison, the logistic regression models (Supplementary Table 6)
160 performed more poorly in prediction by AUC-ROC (0.52 - 0.74) and AUC-PRC (0.19 -
161 0.37) for critical events at three days relative to the XGBoost model (AUC-ROC = 0.77,
162 AUC-PRC = 0.43). With respect to mortality, the model achieved high specificity (0.79 -
163 0.92 for OH) and AUC-ROC (0.79 - 0.92 for OH) with similar AUPRC (0.38 - 0.65 OH)
164 as with critical events. As the event time window increased, the performance of the
165 classifier by the AUPRC value improved, which likely stems from the infrequency of
166 deaths at earlier time points that created a class imbalance for mortality. Comparatively,
167 all logistic regression models underperformed significantly in prediction by AUC-ROC
168 (0.61-0.70) and AUC-PRC (0.06 - 0.18) for mortality.

169

170 *Identifying important features in the model*
171 To identify the most salient features driving model prediction**,** SHAP (SHapley Additive
172 exPlanations) values*(6)* were calculated for the highest-performing model in the cross-
173 validation set during hyperparameter tuning (Figure 3). Among the top features, both
174 high and low levels of lactate dehydrogenase (LDH), procalcitonin, and D-dimer were
175 strong drivers for predicting a critical event at 1 week, while elevated prothrombin time
176 (PT) and partial thromboplastin time (PTT) favored the classifier to predict a critical
177 event. For mortality, both high and low values for age, procalcitonin, and red blood cell
178 distribution width (RDW) were the strongest effectors in guiding mortality prediction by
179 the model within 1 week of admission. Other important variables for increasing the
180 prediction for death included an elevated troponin, LDH, lymphopenia (i.e. low
181 lymphocyte percentage), white blood cell count (WBC), aspartate aminotransferase
182 (AST), and D-dimer. Finally, using SHAP interaction scores, we discovered that
183 covariate interactions between features, relative to each feature's independent
184 importance, contributed less to the model's prediction (Supplementary Figures 1-4).

185

186 **Discussion**

187   We highlight several important findings that have implications in clinical medicine. First,
188   we offer robust prediction algorithms pertaining to the most clinically severe outcomes
189   based solely on admission metrics. This insight provides the likely hospital course for
190   patients up to 10 days into the future. High sensitivity in predicting mortality within three,
191   five, and seven days of admission (0.86 - 1.0) suggests that this model can potentially
192   be used by clinicians in gauging the acute clinical course of an admitted patient. The
193   model's high specificity, particularly for mortality at days three (specificity = 0.92) and
194   five (specificity = 0.86), suggest its role for augmenting clinicians' decision-making when
195   identifying patients at immediate risk of impending clinical decompensation and
196   potentially guide allocation of more intensive care upon admission.
197
198   Additionally, our framework permits a clinically relevant understanding of the model's
199   most salient features defining its decision boundaries. Age was the most important
200   feature for mortality prediction in COVID-19+ patients, with a notable exponential rise of
201   feature contribution as age increased (Figure 3)*(7, 8)*. Elevations in serum LDH,
202   although nonspecific markers of inflammation, are implicated in pulmonary endothelial
203   cell injury and in COVID-19+ patients*(9–11)*. Equivalently, procalcitonin has been
204   implicated as a biomarker of underlying infection and sepsis risk*(12, 13)*. Elevated
205   RDW, which may be an index for enhanced patient frailty and risk of adverse
206   outcomes*(14)*, was also a strong driver of mortality. Other salient features like
207   leukocytosis*(15)*, a natural response to inflammation, in combination with virally-driven
208   lymphopenia*(16, 17)*, have also been associated with COVID-19 burden. Additionally,
209   elevated troponins*(18, 19)*, renal dysfunction from elevated creatinine*(20)*, anemia, vital
210   instability (low oxygen saturation, tachycardia, hypotension), elevated ferritin*(10, 19)*,
211   high lactate, and acidosis were also contributors to driving model prediction towards
212   mortality. With growing evidence of COVID-19-induced hypercoagulable states in these
213   patients*(10, 21, 22)*, it is promising that our model recognized the feature importance of
214   coagulability markers, such as PT, PTT, and D-dimer (Figure 3). Thus, this
215   corroboration of the features learned by XGBoost and highlighted by the SHAP analysis
216   with those findings from pathophysiological principles and more recent correlative
217   studies exploring COVID-19 patients*(2, 3, 18, 23–25)* gives additional credibility to these
218   findings.
219
220   Just as interesting as the features present in the SHAP value analysis were those that
221   were not. For example, because race is both poorly represented ("Unknown") and
222   categorized inadequately in electronic health records, the model did not find race to be
223   important for outcome prediction and instead opted to favor more objective data (vitals,
224   labs). Contrary to our expectation, age was not identified as a significant feature for
225   critical event prediction in these primary analyses. However, SHAP value analyses for
226   critical event prediction at longer time frames (10, 15, and 20 days) revealed an

227 increasing importance on age for outcome prediction (Supplementary Figure 5). This
228 trend suggests the model's decision to capture acute events by relying on more
229 objective measures not confounded by other factors that are cached into age, which
230 may better represent illness severity. However, over time, age may become a better
231 marker for critical event prediction, by offering a more stable container of clinical
232 information, given its invariance to change relative to other features. As such, the
233 classifier becomes optimized for predicting severe events earlier in the course of illness.
234
235 The results of our models should be considered in light of several limitations. First, we
236 base our predictions solely on a patient's admission labs (i.e. within 36 hours); while this
237 restriction encourages the use of this model in patient triage, events during a patient's
238 hospital stay after admission may drive their clinical course away from the prior
239 probability. Furthermore, not all patient labs are drawn at admission, which introduces
240 an element of missingness in our dataset. For example, unlike the general patient
241 population, patients on anticoagulation therapy, who likely have comorbidities
242 increasing their baseline risk, will have coagulation labs (PT, PTT) taken on admission.
243 However, the shift away from predicting death by the model in the absence of PT/PTT
244 (Figure 3) suggests that missingness in coagulation labs is a proxy for this lower
245 baseline risk secondary to not having comorbid conditions that require anticoagulation
246 therapy. Additionally, patients admitted to the hospital later in the crisis were both
247 beneficiaries of improved patient care protocols from experiential learning, but also
248 victims of resource constraints from overburdened hospitals. These effects, while
249 possibly negated by our large sample size, may also induce temporal variation between
250 patient outcomes. Furthermore, inherent limitations exist when using EHRs, especially
251 those integrated from multiple hospitals. In order to facilitate timely dissemination of our
252 results, we chose not to manually chart review patient notes that may have otherwise
253 provided additional potential features, such as symptoms and clinical course, to
254 incorporate in our model. Because all five hospitals operate in a single health system,
255 system-wide protocols in lab order sets and management protocols were an additional
256 source of bias that may lower external validity. Other interhospital effects such as
257 shuttling COVID-19 cases to certain hospitals for balancing systemic patient burden
258 may also imbalance case severity across hospitals and care management between
259 hospitals; certainly, this was the case for MSW, where mortality at 3-days was far lower
260 (1.7%) than other hospital sites. This was ultimately a major reason to restrict model
261 training to a single center and perform testing out of sample in another hospital center.
262 Finally, though XGBoost is superior to other models at handling missing data, a notable
263 drawback is its bias towards continuous features instead of categorical ones, given
264 increased information represented in the form*(28)*. However, collinearities between
265 some categorical features in this dataset may be present with other continuous features,
266 as exhibited by covariance strength between hypertension and systolic BP and

267   creatinine in Supplemental Figure 4, which can then serve as vehicles for capturing
268   these categorical pieces of information.
269
270   In conclusion, the COVID-19 pandemic unequivocally represents an unprecedented
271   public health crisis. Healthcare institutions are facing extreme difficulties in managing
272   resources and personnel. Physicians are treating record numbers of patients and
273   continuously expose themselves to a highly contagious and virulent disease with
274   varying symptomatology. Only few therapeutic options have demonstrated improvement
275   to patient outcomes. As COVID-19 moves outside of the current epicenter in New York
276   City, healthcare institutions will see a larger influx of affected patients and can benefit
277   from immediate insights regarding assessment of disease severity*(29, 30)*. These
278   models successfully predict critical illness and mortality up to 10 days in advance in a
279   diverse patient population from admission information alone and provide important
280   markers for acute care prognosis that can be used by healthcare institutions to improve
281   care decisions at both the physician and hospital level for management of COVID-19
282   positive patients.
283
284   **Materials and Methods**
285
286   This study has been approved by the Institutional Review Board at the Icahn School of
287   Medicine at Mount Sinai (IRB- 20-03271).
288
289   *Clinical Data Source and Study Population*
290   In this study, patient data came from five hospitals within the Mount Sinai Hospital
291   System (MSHS): the Mount Sinai Hospital (MSH) located in East Harlem, Manhattan;
292   Mount Sinai Morningside (MSM) located in Morningside Heights, Manhattan; Mount
293   Sinai West (MSW) located in Midtown and the West Side, Manhattan; Mount Sinai
294   Brooklyn (MSB) located in Midwood, Brooklyn; and Mount Sinai Queens (MSQ) located
295   in Astoria, Queens. The dataset was obtained from different sources and aggregated by
296   the Mount Sinai COVID Informatics Center (MSCIC).
297
298   We included patients who were over 18 years old that had a laboratory-confirmed
299   COVID-19 infection, and were admitted between March 9 and April 11, 2020 to any of
300   the hospitals previously mentioned. A confirmed case of COVID-19 was defined by a
301   positive reverse transcriptase polymerase chain reaction (RT-PCR) assay of a
302   nasopharyngeal swab. We excluded patients who had a positive COVID-19 RT-PCR
303   result more than two days after admission. Additional exclusion criteria are presented in
304   Figure 1. Full patient characteristics by site are provided in Supplementary Table 1.
305
306   *Study Data*

307 Demographics included age, sex, as well as reported race, and ethnicity. Race was
308 collapsed into seven categories based off of the most recent US census race
309 categories: American Indian or Alaskan Native, Asian, Black or African-American,
310 Other, Native Hawaiian or Other Pacific Islander, Unknown, and White. Ethnicity was
311 collapsed into three categories: Hispanic/Latino, Non-Hispanic/Latino, and Unknown.
312 We obtained demographics, diagnosis codes (International Classification of Diseases-
313 9/10-Clinical Modification (ICD-9/10-CM) codes and procedures), as well as vital signs
314 and laboratory measurements during hospitalization. A pre-existing condition was
315 defined as the presence of ICD-9/10-CM codes associated with specific diseases. We
316 chose to include as covariates conditions that have been previously reported to have
317 increased incidence in hospitalized COVID-19 patients, specifically: atrial fibrillation,
318 asthma, cancer, coronary artery disease, chronic kidney disease, chronic obstructive
319 pulmonary disease, diabetes mellitus, heart failure, hypertension, and stroke*(15, 23–25,*
320 *31).*
321
322 We included laboratory measurements and vital signs near the time of admission for
323 prediction. Specifically, because records for laboratory values and vitals may appear
324 with some lag, only the first available value within 36 hours of admission was included,
325 otherwise the value was assigned as missing. Height was absent in a large percentage
326 of the patients (18.2%). Because height is generally invariant in the adult population,
327 and given the resource constraint of the pandemic, it was common for triage nurses to
328 use the height from a previous and recent admission. In an effort to be as cohesive in
329 our data gathering methods as possible, these earlier records were not retrieved for this
330 dataset. However, weight was used as the next approximate proxy for body habitus,
331 with additional information being presented through sex and age for body habitus as
332 well.
333
334 All lab orders from the five hospitals were queried for patients included in this study
335 within the timeframe of interest. Due to discrepancies in how labs were named in
336 different hospitals, a comprehensive review of all lab field names was conducted by a
337 multidisciplinary team of clinical and statistical experts to ensure that there was a direct
338 mapping between all sites. Additionally, many labs represented a single component
339 (e.g. sodium), but were acquired from either an arterial blood gas (ABG), venous blood
340 gas (VBG), and basic metabolic panel (BMP). Based on the utility of these lab values in
341 clinical practice and the similarity between their statistical distributions, labs derived
342 from a VBG or BMP were collapsed into a single category (i.e. 'SODIUM') and those
343 derived from an ABG were moved to a separate category (i.e. 'SODIUM_A'). Finally, the
344 earliest lab, by time of result, in the set of all lab order names that were combined into a
345 single lab category was chosen as the representative lab value for that category.
346 Finally, lab data below the 0.5th and above the 99.5th percentiles were removed to

347   avoid inclusion of any obvious outliers that could represent incorrect documentation and
348   measurement error.

349

350   *Definition of Outcomes*

351   The two primary outcomes were 1) death versus survival or discharge, and 2) critical

352   illness versus survival or discharge, through time horizons of 3, 5, 7, and 10 days.

353   Critical illness is defined as discharge to hospice, intubation ≤ 48 hours prior to ICU

354   admission, or death. To address potential concerns of censoring by limiting exploration

355   of only these time frames, particularly in the case of model enrichment for acute critical

356   events, we also predicted critical events at days 15 and 20.

357

358   *Statistical Analysis*

359   Our primary model was fit with the Extreme Gradient Boosting (XGBoost)
360   implementation of boosted decision trees on continuous and one-hot encoded
361   categorical features. The XGBoost algorithm provides state-of-the-art prediction results
362   through an iterative process of averaging in decision trees (we used 100) fit to the
363   residual error of the prior ensemble. While each tree is too simple to accurately capture
364   complex phenomena, the combination of many trees in the XGBoost model
365   accommodates non-linearity and interactions between predictors. Missing data values
366   are routed through split points based on the direction to minimize loss. XGBoost models
367   were trained and evaluated using 10-fold stratified cross validation. For each fold,
368   hyperparameter tuning was performed by randomized grid searching directed towards
369   maximizing the sensitivity metric over 2,000 discrete grid options. Cross-validation was
370   performed inside each grid option. We present the model hyperparameters for all
371   experiments in Supplementary Table 3. The performance of the models were measured
372   using the area under the receiver operator characteristic curve (AUC-ROC), area under
373   the precision-recall curve (AUC-PRC), F1-score, sensitivity, and specificity. To interpret
374   the significance of input features on the model's prediction, SHAP values across all
375   features on the best-performing model, by AUC-ROC, in the cross-validation set were
376   calculated. Finally, we tested these models built on patient data from MSH on patients
377   from four other hospitals (Figure 1).

378

379   As a baseline, we also fit a logistic regression model and a generalized additive logistic
380   model to compare our XGBoost model. We decided to use a generalized additive model
381   because of its ability to extend generalized linear models by allowing for non-linear
382   functions of features. Four main models were generated: 1) Logistic regression using
383   only Age 2) Generalized Additive Model using only Age 3) Logistic Regression with all
384   available features 4) Generalized Additive Model using all available features. Since

385 these models do not have a built in method of dealing with missing data, we dropped

386 features with over 70% missingness and samples that lacked values for the remaining

387 feature space. The models were trained and evaluated using 10-fold stratified cross

388 validation on patient data from MSH and subsequently evaluated on patient data from

389 the other hospitals. The same metrics were recorded for these models; however, this

390 model was only trained at outcome prediction on Day 3, which was the time frame at

391 which model AUC-ROC was highest for the XGBoost classifier. Performance results for

392 this classifier are represented in Supplementary Table 5.

References

1. G. Onder, G. Rezza, S. Brusaferro, Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy*JAMA* (2020), doi:10.1001/jama.2020.4683.

2. W.-J. Guan, Z.-Y. Ni, Y. Hu, W.-H. Liang, C.-Q. Ou, J.-X. He, L. Liu, H. Shan, C.-L. Lei, D. S. C. Hui, B. Du, L.-J. Li, G. Zeng, K.-Y. Yuen, R.-C. Chen, C.-L. Tang, T. Wang, P.-Y. Chen, J. Xiang, S.-Y. Li, J.-L. Wang, Z.-J. Liang, Y.-X. Peng, L. Wei, Y. Liu, Y.-H. Hu, P. Peng, J.-M. Wang, J.-Y. Liu, Z. Chen, G. Li, Z.-J. Zheng, S.-Q. Qiu, J. Luo, C.-J. Ye, S.-Y. Zhu, N.-S. Zhong, China Medical Treatment Expert Group for Covid-19, Clinical Characteristics of Coronavirus Disease 2019 in China, *N. Engl. J. Med.* (2020), doi:10.1056/NEJMoa2002032.

3. E. Livingston, K. Bucher, Coronavirus Disease 2019 (COVID-19) in Italy*JAMA* **323**, 1335 (2020).

4. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020, *Eurosurveillance* **25**, 2000180 (2020).

5. T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System (2016), doi:10.1145/2939672.2939785.

6. S. M. Lundberg, S.-I. Lee, in *Advances in Neural Information Processing Systems*, (2017), pp. 4765–4774.

7. K. Wang, P.-Y. Zuo, Y. Liu, M. Zhang, X. Zhao, S. Xie, H. Zhang, X. Chen, C. Liu, Clinical and Laboratory Predictors of In-Hospital Mortality in 305 Patients with COVID-19: A Cohort Study in Wuhan, China, (2020), doi:10.2139/ssrn.3546115.

8. Q. Ruan, K. Yang, W. Wang, L. Jiang, J. Song, Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China, *Intensive Care Med.* , 1–3 (2020).

9. M. Drent, N. A. Cobben, R. F. Henderson, E. F. Wouters, M. van Dieijen-Visser, Usefulness of lactate dehydrogenase and its isoenzymes as indicators of lung damage or inflammation, *Eur. Respir. J.* **9**, 1736–1742 (1996).

10. E. Terpos, I. Ntanasis-Stathopoulos, I. Elalamy, E. Kastritis, T. N. Sergentanis, M. Politou, T. Psaltopoulou, G. Gerotziafas, M. A. Dimopoulos, Hematological findings and complications of COVID-19, *Am. J. Hematol.* (2020), doi:10.1002/ajh.25829.

11. B. M. Henry, M. H. S. de Oliveira, S. Benoit, M. Plebani, G. Lippi, Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis, *Clin. Chem. Lab. Med.* (2020), doi:10.1515/cclm-2020-0369.

429 12. A. L. Vijayan, Vanimaya, S. Ravindran, R. Saikant, S. Lakshmi, R. Kartik, M. G,
430 Procalcitonin: a promising diagnostic marker for sepsis and antibiotic therapy, *J.*
431 *Intensive Care Med.* **5**, 51 (2017).

432 13. J. T. Poston, B. K. Patel, A. M. Davis, Management of Critically Ill Adults With
433 COVID-19, *JAMA* (2020), doi:10.1001/jama.2020.4914.

434 14. C. Fava, F. Cattazzo, Z.-D. Hu, G. Lippi, M. Montagnana, The role of red blood cell
435 distribution width (RDW) in cardiovascular risk assessment: useful or hype?, *Ann Transl*
436 *Med* **7**, 581 (2019).

437 15. J. Zhu, P. Ji, J. Pang, Z. Zhong, H. Li, C. He, J. Zhang, C. Zhao, Clinical
438 characteristics of 3,062 COVID-19 patients: a meta-analysis, *J. Med. Virol.* (2020),
439 doi:10.1002/jmv.25884.

440 16. J. Liu, H. Li, M. Luo, J. Liu, L. Wu, X. Lin, R. Li, Z. Wang, H. Zhong, W. Zheng, Y.
441 Zhou, D. Jiang, X. Tan, Z. Zhou, H. Peng, G. Zhang, Lymphopenia acted as an adverse
442 factor for severity in patients with COVID-19: a single-centered, retrospective study*In*
443 *Review* (2020).

444 17. L. Tan, Q. Wang, D. Zhang, J. Ding, Q. Huang, Y.-Q. Tang, Q. Wang, H. Miao,
445 Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study,
446 *Signal Transduct Target Ther* **5**, 33 (2020).

447 18. S. Shi, M. Qin, B. Shen, Y. Cai, T. Liu, F. Yang, W. Gong, X. Liu, J. Liang, Q. Zhao,
448 H. Huang, B. Yang, C. Huang, Association of Cardiac Injury With Mortality in
449 Hospitalized Patients With COVID-19 in Wuhan, China, *JAMA Cardiol* (2020),
450 doi:10.1001/jamacardio.2020.0950.

451 19. A. N. Kochi, A. P. Tagliari, G. B. Forleo, G. M. Fassini, C. Tondo, Cardiac and
452 arrhythmic complications in patients with COVID-19, *J. Cardiovasc. Electrophysiol.*
453 (2020), doi:10.1111/jce.14479.

454 20. Y. Cheng, R. Luo, K. Wang, M. Zhang, Z. Wang, L. Dong, J. Li, Y. Yao, S. Ge, G.
455 Xu, Kidney disease is associated with in-hospital death of patients with COVID-19,
456 *Kidney Int.* (2020), doi:10.1016/j.kint.2020.03.005.

457 21. Y. Zhang, M. Xiao, S. Zhang, P. Xia, W. Cao, W. Jiang, H. Chen, X. Ding, H. Zhao,
458 H. Zhang, C. Wang, J. Zhao, X. Sun, R. Tian, W. Wu, D. Wu, J. Ma, Y. Chen, D. Zhang,
459 J. Xie, X. Yan, X. Zhou, Z. Liu, J. Wang, B. Du, Y. Qin, P. Gao, X. Qin, Y. Xu, W. Zhang,
460 T. Li, F. Zhang, Y. Zhao, Y. Li, S. Zhang, Coagulopathy and Antiphospholipid Antibodies
461 in Patients with Covid-19, *N. Engl. J. Med.* (2020), doi:10.1056/NEJMc2007575.

462 22. F. A. Klok, M. J. H. A. Kruip, N. J. M. van der Meer, M. S. Arbous, D. A. M. P. J.
463 Gommers, K. M. Kant, F. H. J. Kaptein, J. van Paassen, M. A. M. Stals, M. V. Huisman,
464 H. Endeman, Incidence of thrombotic complications in critically ill ICU patients with
465 COVID-19, *Thromb. Res.* (2020), doi:10.1016/j.thromres.2020.04.013.

466   23. Y. Liu, Y. Yang, C. Zhang, F. Huang, F. Wang, J. Yuan, Z. Wang, J. Li, J. Li, C.
467   Feng, Z. Zhang, L. Wang, L. Peng, L. Chen, Y. Qin, D. Zhao, S. Tan, L. Yin, J. Xu, C.
468   Zhou, C. Jiang, L. Liu, Clinical and biochemical indexes from 2019-nCoV infected
469   patients linked to viral loads and lung injury, *Sci. China Life Sci.* **63**, 364–374 (2020).

470   24. L. Wynants, B. Van Calster, M. M. J. Bonten, G. S. Collins, T. P. A. Debray, M. De
471   Vos, M. C. Haller, G. Heinze, K. G. M. Moons, R. D. Riley, E. Schuit, L. J. M. Smits, K. I.
472   E. Snell, E. W. Steyerberg, C. Wallisch, M. van Smeden, Prediction models for
473   diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal,
474   *BMJ* **369** (2020), doi:10.1136/bmj.m1328.

475   25. J. Gong, J. Ou, X. Qiu, Y. Jie, Y. Chen, L. Yuan, J. Cao, M. Tan, W. Xu, F. Zheng,
476   Y. Shi, B. Hu, A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) :
477   A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China, *Clin.*
478   *Infect. Dis.* (2020), doi:10.1093/cid/ciaa443.

479   26. R. Witzig, The medicalization of race: scientific legitimization of a flawed social
480   construct, *Ann. Intern. Med.* **125**, 675–679 (1996).

481   27. L. M. Hunt, N. D. Truesdell, M. J. Kreiner, Genes, race, and culture in clinical care:
482   racial profiling in the management of chronic illness, *Med. Anthropol. Q.* **27**, 253–271
483   (2013).

484   28. C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable
485   importance measures: illustrations, sources and a solution, *BMC Bioinformatics* **8**, 25
486   (2007).

487   29. K. F. Boreskie, P. E. Boreskie, D. Melady, Age is just a number - and so is frailty:
488   Strategies to inform resource allocation during the COVID-19 pandemic, *CJEM* , 1–3
489   (2020).

490   30. A. J. Singer, E. J. Morley, M. C. Henry, Staying Ahead of the Wave, *N. Engl. J. Med.*
491   (2020), doi:10.1056/NEJMc2009409.

492   31. H. M. Alger, J. H. Williams Iv, J. G. Walchok, M. M. Bolles, G. C. Fonarow, C.
493   Rutan, The Role of Data Registries in the Time of COVID-19, *Circ. Cardiovasc. Qual.*
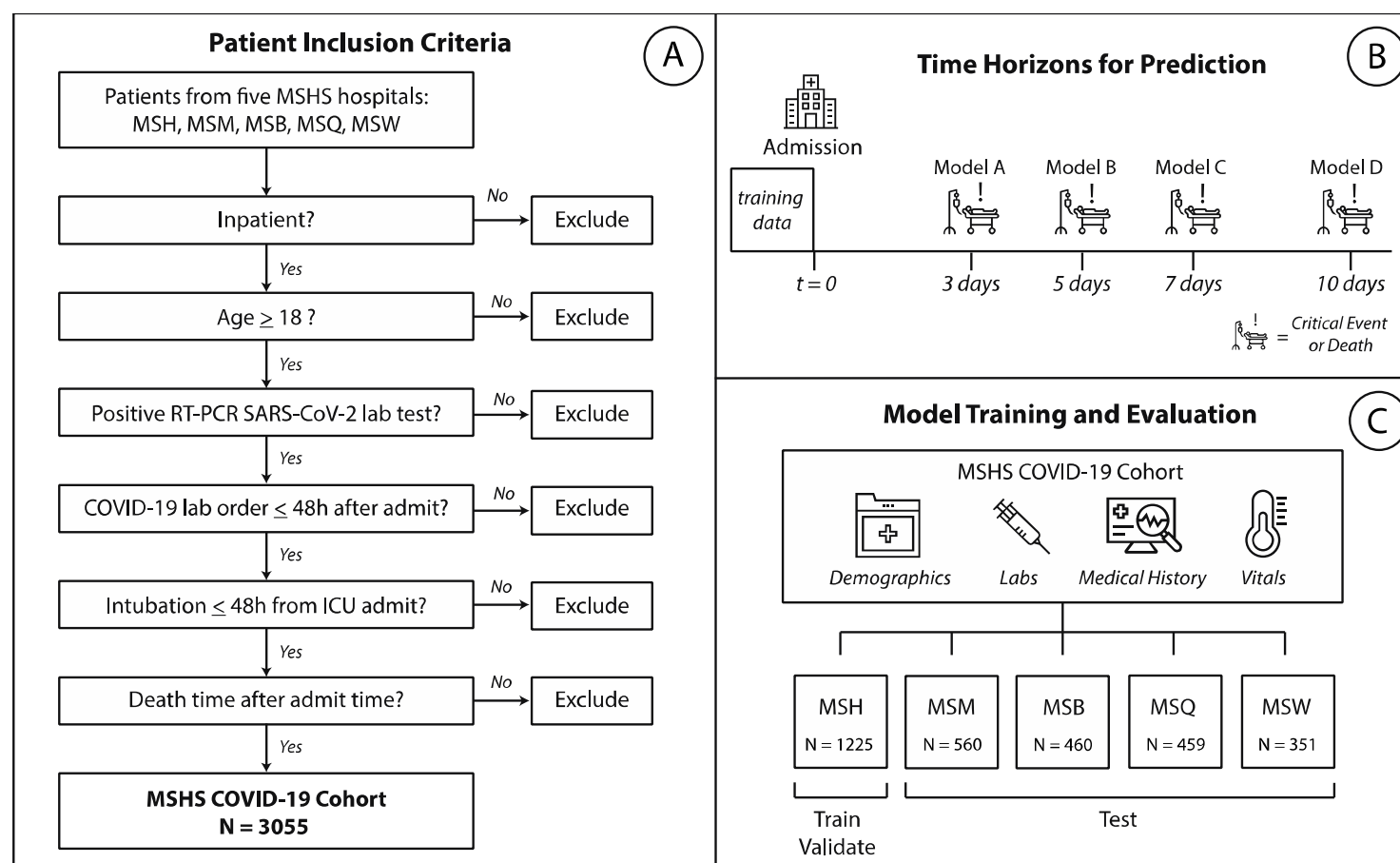494   *Outcomes* (2020), doi:10.1161/CIRCOUTCOMES.120.006766.

495
496
497 **Fig 1. Study Design and Workflow.**
498 **A)** Procedure for patient inclusion in our study. **B)** Strategy and design for experiments. Patient clinical data from Mount
499 Sinai Hospital (MSH) was used to train and validate our machine learning model. We then test these models on patients
500 from four other external hospitals within the Mount Sinai Health System: Mount Sinai Brooklyn (MSB), Mount Sinai
501 Morningside (MSM), Mount Sinai Queens (MSQ), and Mount Sinai West (MSW). **C)** Machine learning experimental
502 design. We train on data from admission to predict mortality and critical illness outcomes.
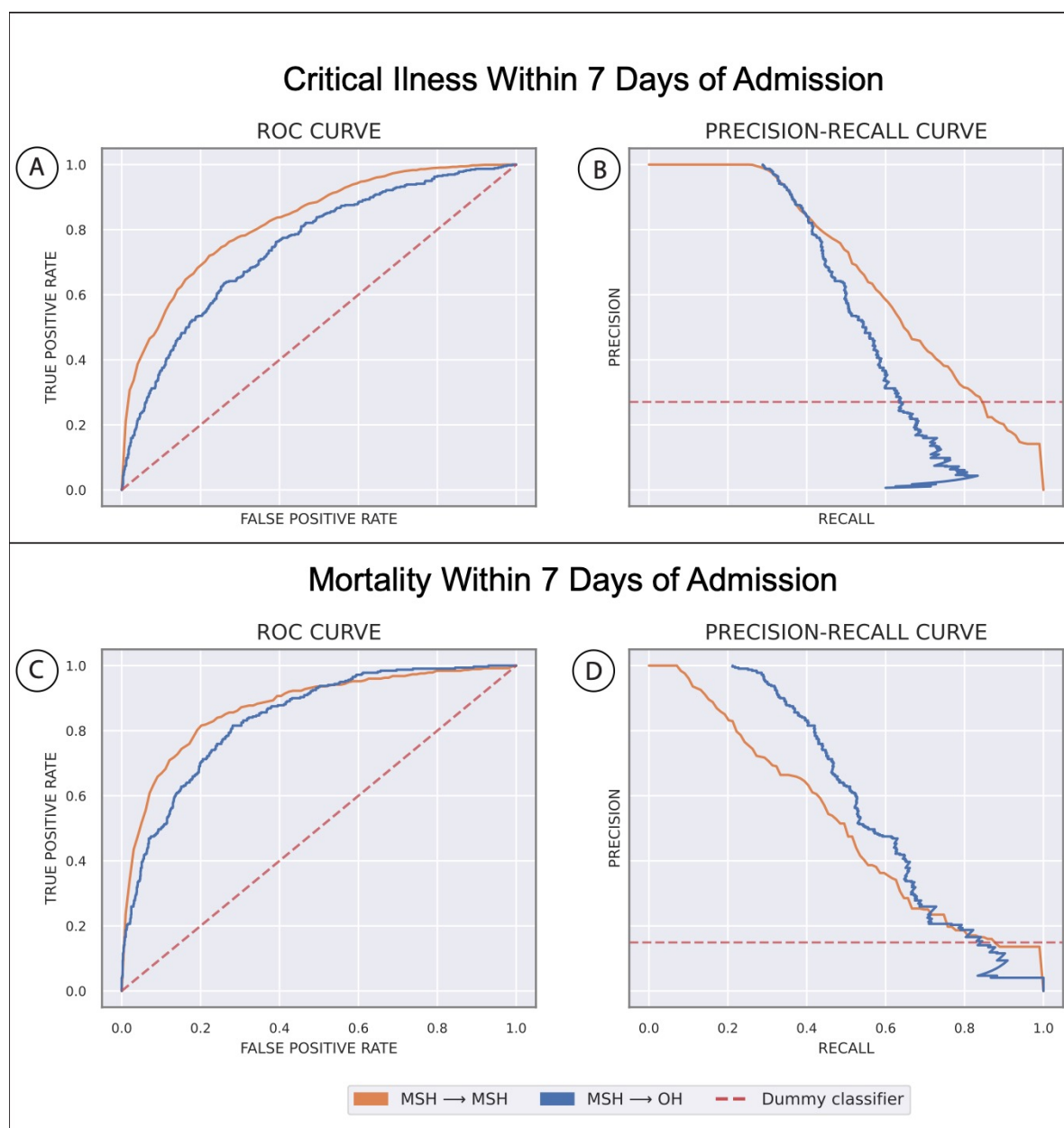
Fig 2. Machine learning experimental results.

In all plots, the orange line reflects training at Mount Sinai Hospital (MSH) and testing
via cross-validation. The blue line reflects testing the model built on patient data from
MSH on external patients from all other hospitals (OH) **A)** Area under the receiver
operator characteristic curve (AUC-ROC) for predicting critical illness at seven days
since admission. **B)** Area under the precision-recall curves (AUPRC) for predicting
critical illness at seven days since admission. **C)** AUC for predicting mortality at seven
days since admission. **D)** AUPRC for predicting critical illness at seven days since
admission.

514
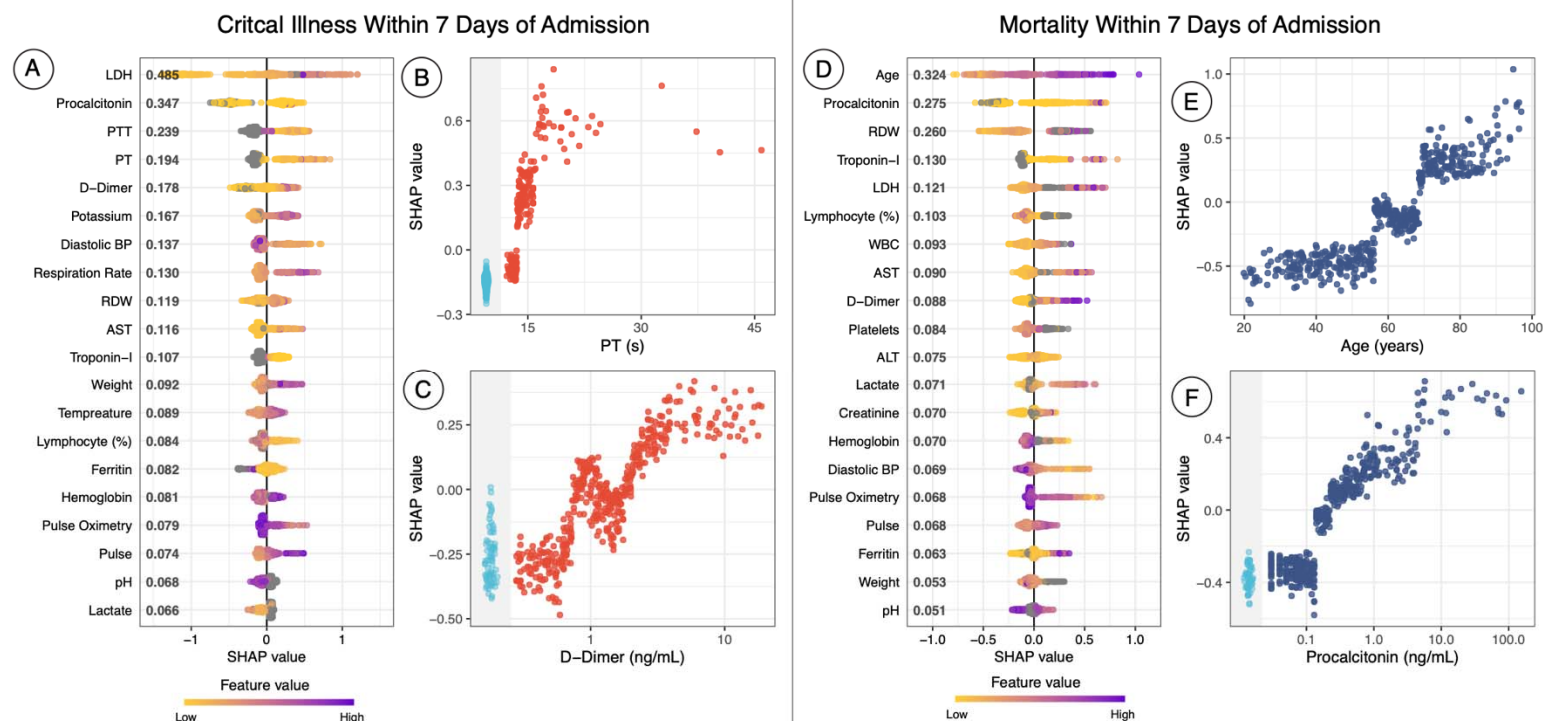
Fig 3. SHAP Summary and Dependency Plots.
SHAP summary plots for critical event **(A)** and mortality **(D)** at 7 days show the SHAP values for the most important features for the respective XGBoost model. Features in the summary plots (y-axis) are organized by their mean absolute SHAP values (x-axis), which represents the importance of that feature in driving the classifier's prediction, for patients. Values of those features for each patient (i.e. a particular LDH value) are colored by their relative value. **(B)** and **(C)** represent dependency plots, which similarly demonstrate how different values of those features can affect the SHAP score and ultimately impact classifier decisions, for prothrombin time (PT) and D-Dimer, respectively, for critical event prediction. **(E)** and **(F)** represent dependency plots for age and procalcitonin levels. Patients with missing values for a feature in the dependency plot are clustered in the shaded area to the left.

526 **Table 1.** Characteristics of Hospitalized Covid-19 Patients at Baseline (n= 3055)
527

Characteristics of Admission	MSH	Other locations
Demographics		
Sex, n (%)		
Male	704 (57.5)	1070 (58.5)
Female	521 (42.5)	760 (41.5)
Race, n (%)		
Other	537 (43.8)	749 (40.9)
Caucasian	293 (23.9)	435 (23.8)
African American	273 (22.3)	503 (27.5)
Unknown	64 (5.2)	60 (3.3)
Asian	44 (3.6)	65 (3.5)
Pacific Islander	14 (1.1)	18 (0.9)
Ethnicity, n (%)		
Non-Hispanic Latino	676 (55.2)	1133 (61.9)
Hispanic/Latino	333 (27.2)	503 (27.5)
Unknown	216 (17.6)	194 (10.6)
Age, Median (IQR)	62.1 (49.4-71.8)	68.33 (56.9-78.8)
Age group, n (%)		
18-30	56 (4.6)	37 (2.0)
31-40	139 (11.6)	93 (5.1)
41-50	133 (10.9)	158 (8.6)
51-60	242 (19.8)	309 (16.9)
61-70	320 (26.1)	431 (23.5)
71-80	202 (16.5)	422 (23.1)
81-90	102 (8.3)	308 (16.8)
At least 90	31 (2.5)	72 (3.9)
Previous Medical History, n (%)		
Hypertension (%)	462 (37.7)	686 (37.5)
Atrial Fibrillation (%)	86 (7.0)	130 (7.1)
Coronary Artery Disease (%)	171 (13.9)	307 (16.8)
Heart failure (%)	110 (8.9)	181 (9.9)
Stroke (%)	98 (8)	118 (6.5)

Chronic Kidney Disease (%)	136 (11.1)	158 (8.6)
Diabetes (%)	313 (25.5)	466 (25.5)
Asthma (%)	115 (9.4)	132 (7.2)
Chronic Obstructive Pulmonary Disease (%)	65 (5.3)	103 (5.6)
Cancer (%)	112 (9.1)	94 (5.1)
Vital Signs at Hospital Admission, Median (IQR)		
Heart Rate (bpm)	89 (78 -100)	89 (78 - 100)
Pulse Oximetry (%)	96 (94 - 98)	96 (94 - 98)
Respiration Rate (breaths / minute)	20 (18 - 20)	18 (18 - 20)
Temperature (F)	98.7 (98.1 - 99.9)	97.9 (98.6 - 99.5)
Systolic Blood Pressure (mmHg)	124 (112 - 138)	127 (112 - 142)
Diastolic Blood Pressure (mmHg)	69 (61 - 78)	72 (65 - 81)
Weight (kg)	80.9 (68.9 - 95.3)	78.9 (68.04 - 91.7)
Admission Laboratory Parameters, Median (IQR)		
Metabolic markers		
Sodium (mEq/L)	137 (135 - 140)	138 (135 -141)
Potassium (mEq/L)	4 (3.6 - 4.5)	4.2 (3.9 - 4.7)
Creatinine (mg/dL)	0.9 (0.7 - 1.4)	1.0 (0.8 -1.6)
Lactate (mg/dL)	1.7 (1.3 - 2.2)	1.4 (1.1 - 2.0)
Hematological markers		
White Blood Cells (10^3/µL)	6.8 (5.2 - 9.8)	7.7 (5.7 - 10.5)
Lymphocyte Percentage	13.1 (7.8 - 20.7)	14.1 (8.9 - 21.4)
Hemoglobin (mEq/L)	12.5 (11.1 - 13.7)	12.8 (11.4 - 14)
Red Blood Cell Distribution Width (%)	14.4 (13.6 -15.6)	12.7 (12.1 - 13.9)
Platelets (#)	212 (162 - 279.3)	194 (148 - 260)
Liver Function		
Alanine Aminotransferase (units/L)	31 (19 - 54)	31 (20 - 53)
Aspartate Aminotransferase (units/L)	41.5 (29 - 67)	47 (31 - 74)
Albumin (g/dL)	3 (2.6 - 3.3)	3 (2.6 - 3.3)
Total Bilirubin (mg/dL)	0.6 (0.4 - 0.8)	0.6 (0.4 - 0.8)
Coagulation markers		
Prothrombin Time (s)	14.3 (13.5 - 15.7)	14.2 (13.5 - 15.5)
Partial Thromboplastin Time (s)	32.5 (29.2 - 37.1)	33.3 (30 - 38.8)

Gases		
PCO$_2$ (mmHg)	37 (41 - 47)	41 (35 - 47)
pH	7.4 (7.4 - 7.4)	7.4 (7.3 - 7.4)
Inflammatory markers		
C Reactive Protein (mg/L)	152.3 (83.1 - 205.6)	124.6 (62.3 - 214.5)
Ferritin (ng/mL)	720 (343.5 - 1813.8)	818 (388 - 1966)
D-Dimer (ng/mL)	1.3 (0.7 - 2.5)	1.5 (0.9 - 3.0)
Creatinine Phosphokinase (units/L)	156 (80 - 505)	237.5 (187.8 - 440.5)
Lactate Dehydrogenase (units/L)	417 (315 - 559)	450 (346.3 - 610.8)
Procalcitonin (ng/mL)	0.2 (0.1 - 0.6)	0.2(0.1 - 0.7)
Cardiac markers		
Troponin I (ng/mL)	0.1 (0.02 - 0.2)	0.2 (0.1- 0.9)

528
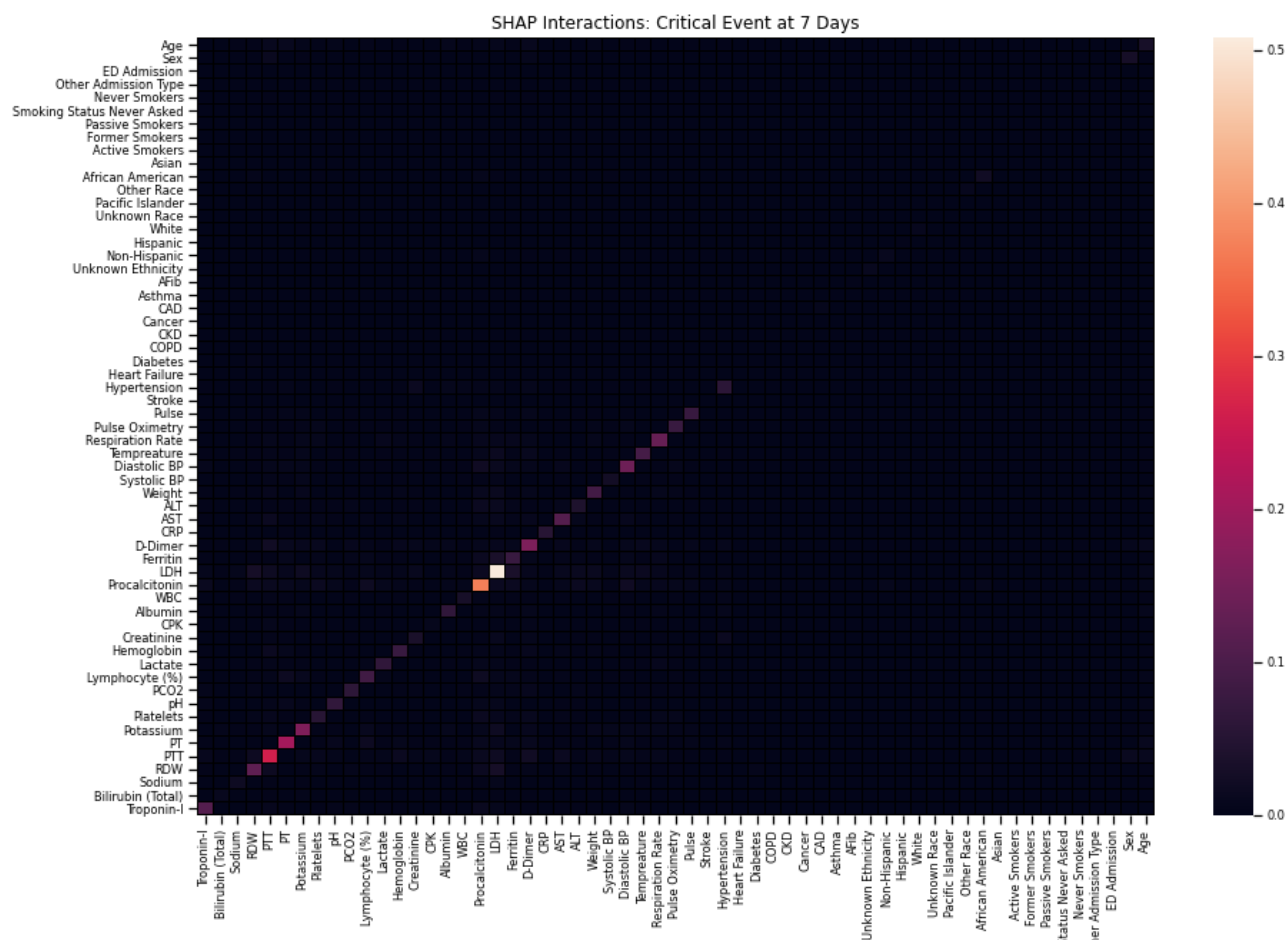
Supplementary Fig 1.SHAP Interactions: Critical Event at 7 Days.

Heatmap demonstrating the composite SHAP interaction scores between features for patients from the best performing k-th validation fold for critical event prediction at 7 days. The intensity along the diagonal represents independent contributions of a feature towards model prediction (i.e. mean absolute SHAP values in Figure 3). Covariant interactions between features is significantly less relative to the intensity of the independent contributions of each feature towards model prediction.

SHAP Interactions: Critical Event at 7 Days (Univariate Effect Removal)
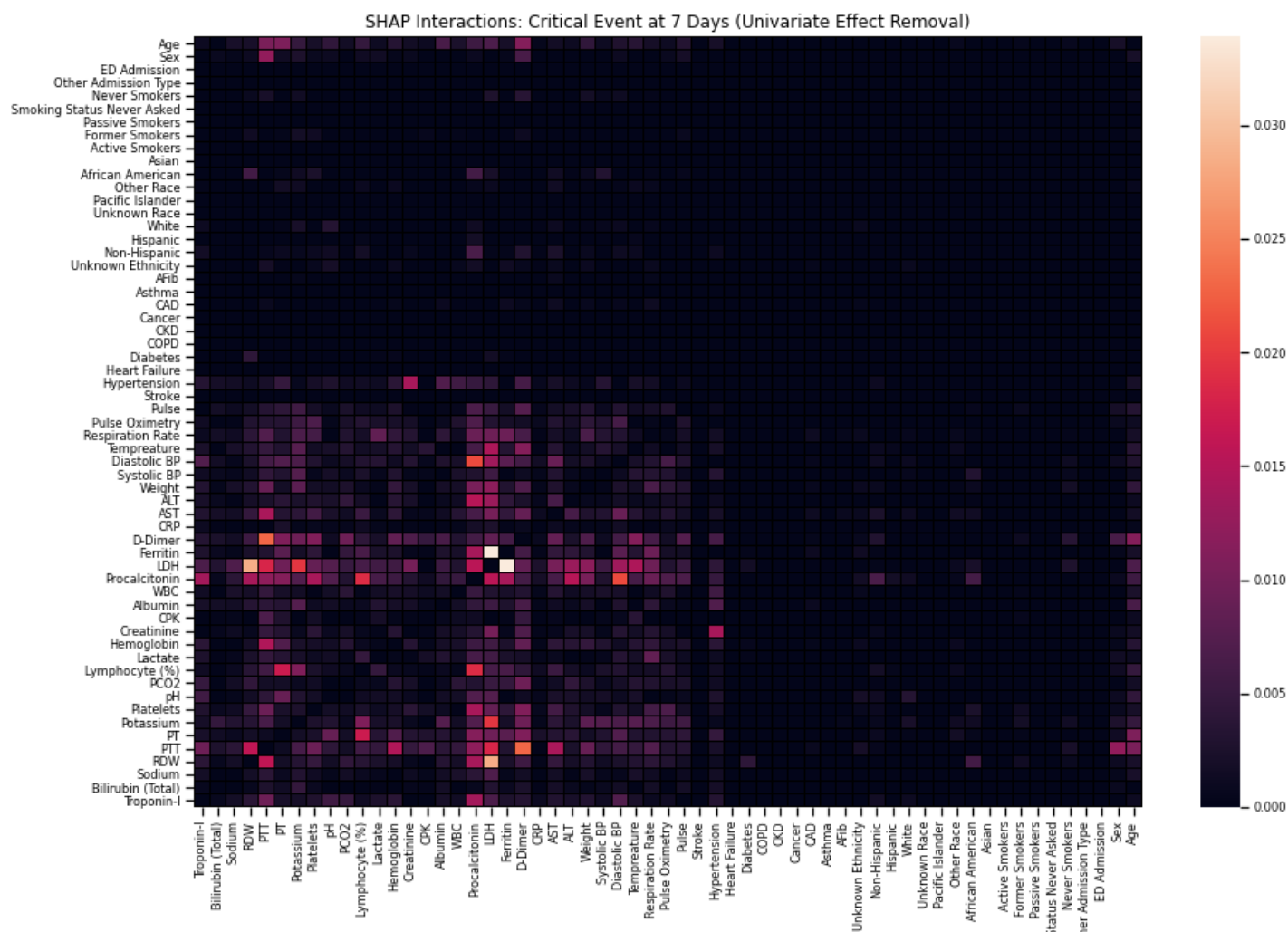
537
538
539 **Supplementary Fig 2. SHAP Interactions: Critical Event at 7 Days (Univariate Effect Removal).**
540 Heatmap demonstrating the composite SHAP interaction scores between features for patients from the best performing k-
541 th validation fold for critical event prediction at 7 days. Univariate feature contributions along the diagonal have been set to
542 0 to better examine the relative strength of covariance between features.
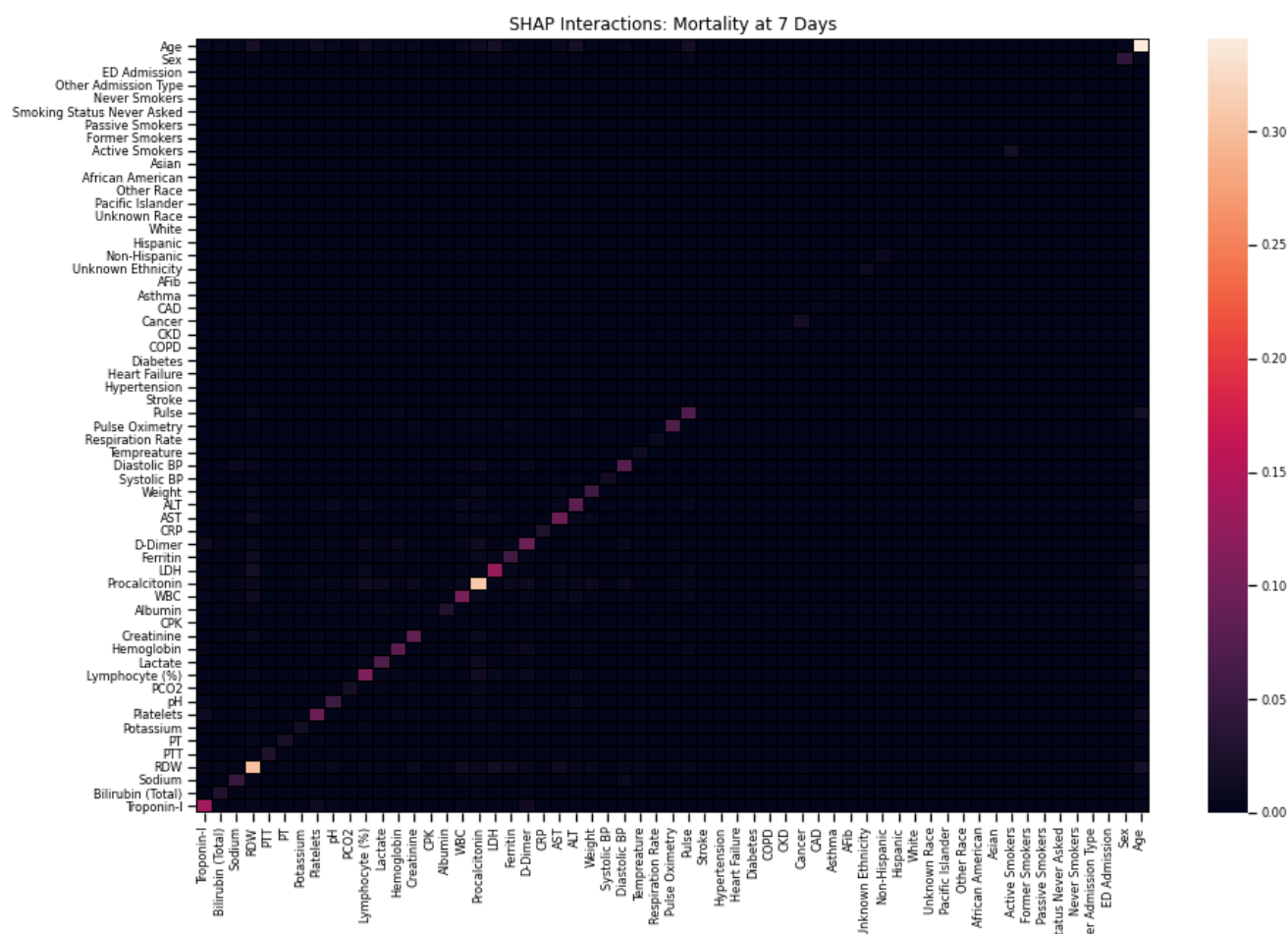
543

544

545 **Supplementary Fig 3. SHAP Interactions: Mortality at 7 Days.**

546 Heatmap demonstrating the composite SHAP interaction scores between features for patients from the best performing k-

547 th validation fold for mortality prediction at 7 days. The intensity along the diagonal represents independent contributions

548 of a feature towards model prediction (i.e. mean absolute SHAP values in Figure 3). Covariant interactions between

549 features is significantly less relative to the intensity of the independent contributions of each feature towards model
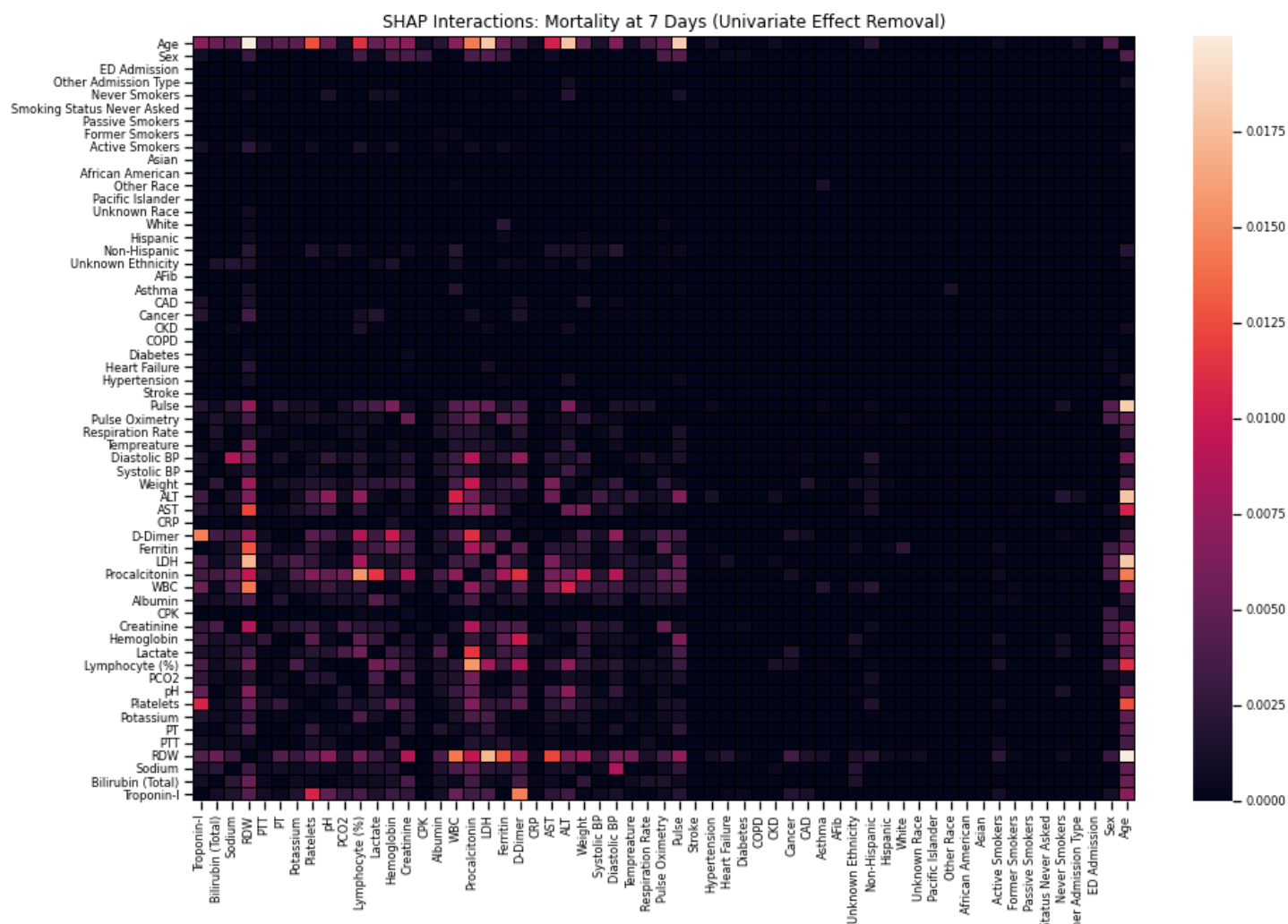
550 prediction.

SHAP Interactions: Mortality at 7 Days (Univariate Effect Removal)

551
552
553 **Supplementary Fig 4. SHAP Interactions: Mortality at 7 Days (Univariate Effect Removal).**
554 Description: Heatmap demonstrating the composite SHAP interaction scores between features for patients from the best
555 performing k-th validation fold for mortality prediction at 7 days. Univariate feature contributions along the diagonal have
556 been set to 0 to better examine the relative strength of covariance between features.
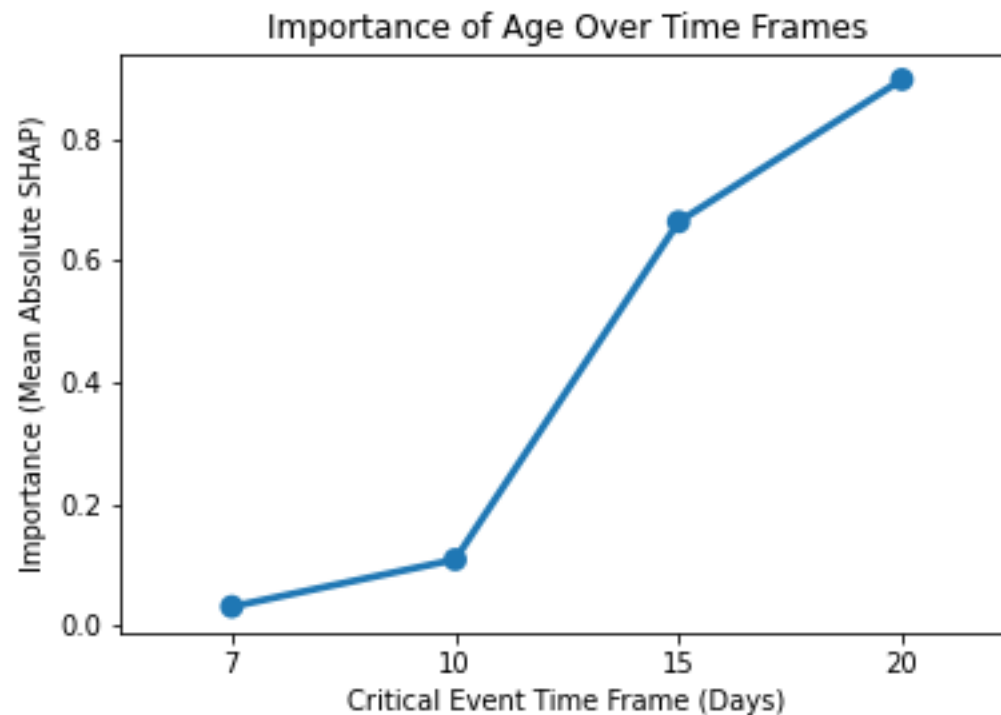
Supplementary Fig 5. Importance of Age Over Time Frames

Mean absolute SHAP value for age (blue) to represent importance in prediction of critical event occurrence over different time frames (7, 10, 15, and 20 days).

561 **Supplementary Table 1. Baseline Patient Characteristics by Hospital.**
562 Characteristics, such demographics, clinical history, vital signs, and laboratory tests, for
563 all patients included in the study and delineated by the hospital site at which the patient
564 was admitted.
565

Characteristics or Admission	MSQ	MSB	MSH	MSW	MSM
Demographics					
Sex, n (%)					
Male	291 (63.4)	207 (58.9)	317 (56.6)	255 (55.4)	704 (57.47)
Female	168 (36.6)	144 (41.0)	243 (43.4)	205 (44.6)	521 (42.53)
Race, n (%)					
Other	33 (7.2)	9 (1.9)	44 (3.6)	16 (4.6)	7 (1.3)
African American	50 (10.9)	191 (41.5)	273 (22.3)	68 (19.4)	194 (34.6)
Caucasian	264 (57.5)	91 (19.8)	537 (43.8)	131 (37.3)	263 (46.9)
Asian	9 (1.9)	0 (0)	14 (1.1)	4 (1.1)	5 (0.9)
Pacific Islander	16 (3.5)	19 (4.1)	64 (5.2)	5 (1.4)	20 (3.6)
Unknown	87 (18.9)	150 (32.6)	293 (23.9)	127 (36.2)	71 (12.7)
Ethnicity, n (%)					
Hispanic/Latino	230 (50.1)	322 (70)	676 (55.2)	251 (71.5)	330 (58.9)
Non-Hispanic Latino	188 (40.9)	17 (3.7)	333 (27.2)	92 (26.2)	206 (36.8)
Unknown	41 (8.9)	121 (26.3)	216 (17.6)	8 (2.3)	24 (4.3)
Age - Median(IQR)	66 (56-77)	65.3 (51.7-76.4)	68.8 (56.0-80.1)	72 (63-82)	62.1 (49.4-71.8)
Age group, n (%)					
18-30	10 (2.2)	16 (4.6)	7 (1.3)	4 (0.9)	56 (4.6)
31-40	16 (3.5)	38 (10.8)	32 (5.7)	7 (1.5)	139 (11.4)
41-50	49 (10.7)	32 (9.1)	60 (10.7)	17 (3.7)	133 (10.9)
51-60	92 (20.0)	62 (17.7)	90 (16.1)	65 (14.1)	242 (19.8)
61-70	110 (23.9)	71 (20.2)	123 (21.9)	127 (27.6)	320 (26.1)
71-80	108 (23.5)	73 (20.8)	120 (21.4)	121 (26.3)	202 (16.5)
81-90	64 (13.9)	51 (14.5)	101 (18.0)	92 (20)	102 (8.3)
At least 90	10 (2.2)	8 (2.3)	27 (4.8)	27 (5.9)	31 (2.5)
Previous Medical History, n (%)					
Hypertension (%)	228 (49.7)	120 (34.2)	210 (37.5)	128 (27.8)	462 (37.7)

Atrial Fibrillation (%)	45 (9.8)	18 (5.1)	40 (7.1)	27 (5.9)	86 (7.0)
Coronary Artery Disease (%)	95 (20.7)	46 (13.1)	85 (15.2)	81 (17.6)	171 (13.9)
Heart failure (%)	51 (11.1)	27 (7.7)	61 (10.9)	42 (9.1)	110 (8.9)
Stroke (%)	43 (9.4)	22 (6.3)	39 (6.9)	14 (3.0)	98 (8)
Chronic Kidney Disease (%)	54 (11.8)	20 (5.7)	58 (10.4)	26 (5.6)	136 (11.1)
Diabetes (%)	154 (33.6)	68 (19.4)	143 (25.5)	101 (21.9)	313 (25.6)
Asthma (%)	42 (9.2)	32 (9.1)	45 (8.0)	13 (2.8)	115 (9.4)
Chronic Obstructive Pulmonary Disease (%)	35 (7.6)	16 (4.6)	31 (5.5)	21 (4.6)	65 (5.1)
Cancer (%)	20 (4.4)	30 (8.5)	33 (5.9)	11 (2.4)	112 (9.1)
Vital Signs at Hospital Admission, Median (IQR)					
Heart Rate (bpm)	89 (78 – 101)	89 (78-100.2)	89 (78-100)	85 (76-96)	89 (79-101.3)
Oxygen Saturation (%)	95 (93-97)	96 (94-98)	96 (94-98)	96 (94.5-98)	96 (94-98)
Respiration Rate (breaths / minute)	18 (18-20)	18 (18-19.3)	20 (18-22)	18 (17- 20)	20 (18-22)
Temperature (F)	98.8 (98.2-99.7)	98 (97.3-98.7)	98.7 (98.1-99.9)	98.7 (98-99.6)	99 (98.2-100)
Systolic Blood Pressure (mmHg)	70 (63-78)	70 (62.75-80)	69 (61-78)	74 (65.5-84)	74 (66.8-82)
Diastolic Blood Pressure (mmHg)	123 (110-138)	130 (112-145)	124 (112-138)	126 (111-137)	128.5 (114-144)
Weight (kg)	76.3 (65.2-90.7)	79.4 (68.1-92.9)	80.9 (69.0-95.3)	78.8 (67.6-90.7)	79.6 (69.1-92.2)
Admission Laboratory Parameters, Median (IQR)					
Metabolic Markers					
Sodium (mEq/L)	138 (136-141)	139 (137-142)	137 (135-140)	137 (134-139)	138 (135-141)
Potassium (mEq/L)	4.3 (3.9-4.7)	4.3 (3.8-4.7)	4 (3.6-4.5)	4.2 (3.9-4.7)	4.2 (3.8-4.7)
Creatinine (mg/dL)	1 (0.8-2.1)	1.2 (0.9-2.1)	0.9 (0.7-1.4)	0.9 (0.7-1.3)	1.0 (0.8-1.7)
Lactate (mg/dL)	-	1.94 (1.1-2.9)	1.7 (1.3-2.2)	1.5 (1.1-2)	1.3 (1-1.8)
Hematological Markers					
White Blood Cells (103/µL)	7.9 (5.9-10.5)	7.5 (5.4-10.5)	6.8 (5.2-9.8)	7.7 (5.6-10.9)	7.5 (5.7-10.5)
Lymphocyte Percentage	12 (7.1-18.3)	14.7 (8.9-22.7)	13.1 (7.8-20.1)	16.1 (9.6-23.1)	13.05 (9-20.5)
Hemoglobin (mEq/L)	12.6 (11.1-13.8)	12.5 (11.2-13.8)	12.5 (11.1-13.7)	13 (11.6-14.2)	13 (11.6-14)

Red Blood Cell Distribution Width (%)	12.8 (12.1-14.2)	13.1 (12.5-14.1)	14.4 (13.6-15.6)	12.5 (12-13.5)	12.6 (12.1-13.7)
Platelets (#)	188 (146-259)	205.5 (153-266.2)	212 (162-279.3)	197 (152-264)	187 (144-252)
Liver Function					
Alanine Aminotransferase (units/L)	34 (21-55)	28 (19-49)	31 (19-54)	30 (19-54)	32 (19-54.8)
Aspartate Aminotransferase (units/L)	50 (34-76.8)	42 (28-74)	41.5 (29-67)	48.5 (32-77)	45 (31-71)
Albumin (g/dL)	2.7 (2.3-3.2)	3.4 (3.1-3.7)	3 (2.6-3.3)	3 (2.6-3.3)	2.9 (2.5-3.1)
Total Bilirubin (mg/dL)	0.6 (0.4-0.7)	0.6 (0.4-0.7)	0.6 (0.4-0.8)	0.5 (0.4-0.8)	0.6 (0.4-0.8)
Coagulation Markers					
Prothrombin Time (s)	14 (13.4-17.2)	14.6 (13.5-16.1)	14.3 (13.5-15.7)	13.9 (13.4-14.9)	14.2 (13.6-15.3)
Partial Thromboplastin Time (s)	35.7 (32.2-61.2)	35.2 (32.3-38.6)	32.5 (29.2-37.2)	32.1 (29.8-37.5)	32.9 (29.4-38.7)
Gases					
PCO2 (mmHg)	-	40 (34.3-46)	41 (37-47)	41.5 (36.3-48)	41 (35-47)
pH	-	7.4 (7.3-7.4)	7.4 (7.4-7.4)	7.4 (7.3-7.4)	7.4 (7.3-7.4)
Inflammatory Markers					
C Reactive Protein (mg/L)	207.2 (159.4-233.4)	147.7 (90.7-243.3)	152.3 (83.1-205.6)	104.9 (50.6-190.4)	125.4 (62.7-201.6)
Ferritin (ng/mL)	968 (473.6-2378.3)	999.5 (522.2-2228.5)	720 (343.5-1813.8)	713.5 (305.8-1466)	763 (360- 1784)
D-Dimer (ng/mL)	1.5 (1.0-2.9)	2.3 (1.1-4.1)	1.3 (0.7-2.5)	1.4 (0.7-2.6)	1.5 (0.9-2.9)
Creatinine Phosphokinase (units/L)	237.5 (187.8-440.5)	-	156 (80-505)	-	-
Procalcitonin (ng/mL)	0.3 (0.1-0.8)	0.4 (0.2-1.4)	0.2 (0.1-0.6)	0.2 (0.1-0.5)	0.2 (0.1-0.7)
Lactate Dehydrogenase (units/L)	468 (359.5-628)	489 (390-637.5)	417 (315-559)	448 (345-651.5)	427 (324-559)
Cardiac Markers					
Troponin I (ng/mL)	0.2 (0.1-0.9)	-	0.1 (0.0-0.2)	-	-
Outcomes					
Critical	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Mortality	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)

566 **Supplementary Table 2. Baseline Patient Feature Analysis.**
567 Missingness, means, standard deviations, interquartile ranges, and histograms for all
568 features in the dataset.
569

Variable	Missing (N)	Present (%)	Mean	Standard Deviation	IQR - 0%	IQR - 25%	IQR - 50%	IQR - 75%	IQR - 100%	Histogram
Demographics										
Age	0	100%	64.5	16.5	18.3	54.4	65.7	76.5	102	
Sex	0	100%	0.581	0.494	0	0	1	1	1	
ED Admission	0	100%	0.973	0.164	0	1	1	1	1	
Other Admission Type	0	100%	0.113	0.316	0	0	0	0	1	
Never Smokers	0	100%	0.521	0.5	0	0	1	1	1	
Smoking Status: Never Asked	0	100%	0.0426	0.202	0	0	0	0	1	
Passive Smokers	0	100%	0.000982	0.0313	0	0	0	0	1	
Former Smokers	0	100%	0.205	0.403	0	0	0	0	1	
Active Smokers	0	100%	0.0383	0.192	0	0	0	0	1	
Asian	0	100%	0.0357	0.186	0	0	0	0	1	
African American	0	100%	0.254	0.435	0	0	0	1	1	
Other Race	0	100%	0.421	0.494	0	0	0	1	1	
Pacific Islander	0	100%	0.0105	0.102	0	0	0	0	1	
Unknown Race	0	100%	0.0406	0.197	0	0	0	0	1	
White	0	100%	0.238	0.426	0	0	0	0	1	
Hispanic	0	100%	0.274	0.446	0	0	0	1	1	
Non-Hispanic	0	100%	0.592	0.492	0	0	1	1	1	
Unknown Ethnicity	0	100%	0.134	0.341	0	0	0	0	1	
Past Medical History										

Atrial Fibrillation	0	100%	0.0707	0.256	0	0	0	0	1	
Asthma	0	100%	0.0809	0.273	0	0	0	0	1	
CAD	0	100%	0.156	0.363	0	0	0	0	1	
Cancer	0	100%	0.0674	0.251	0	0	0	0	1	
CKD	0	100%	0.0962	0.295	0	0	0	0	1	
COPD	0	100%	0.055	0.228	0	0	0	0	1	
Diabetes	0	100%	0.255	0.436	0	0	0	1	1	
Heart Failure	0	100%	0.0953	0.294	0	0	0	0	1	
Hypertension	0	100%	0.376	0.484	0	0	0	1	1	
Stroke	0	100%	0.0707	0.256	0	0	0	0	1	
Vitals										
Pulse	0	100%	89.5	16.7	46	78	89	100	150	
Pulse Oximetry	11	100%	95.2	3.24	76	94	96	98	99	
Respiration Rate	4	100%	20.2	4.2	7	18	19	20	46	
Temperature	6	100%	98.9	1.4	95.2	98	98.7	99.7	103	
Diastolic BP	5	100%	71.7	12.2	39	63	71	80	107	
Systolic BP	6	100%	127	21.3	72	112	126	140	189	
Weight	192	94%	82.5	20.6	40	68	79.6	93	174	
Labs										
ALT	111	96%	47.5	56.1	8	19	31	54	597	
AST	124	96%	64.9	73.4	13	30	44	72	905	
CRP	1901	38%	144	94.8	1.91	63.7	127	215	457	
D-Dimer	763	75%	2.53	3.15	0.28	0.79	1.43	2.75	18.8	
Ferritin	546	82%	1529	2161	36	369	776	1901	17466	
LDH	588	81%	498	261	171	328	436	585	2165	
Procalcitonin	661	78%	1.71	7.64	0.03	0.09	0.21	0.68	154	

WBC	9	100%	8.39	4.31	1.9	5.4	7.4	10.3	40	
Albumin	885	71%	2.97	0.518	1.2	2.6	3	3.3	4.1	
CPK	2768	9%	435	757	17	82.5	173	500	8103	
Creatinine	667	78%	1.73	2.22	0.37	0.74	0.97	1.55	16.1	
Hemoglobin	669	78%	12.5	2.01	6.2	11.3	12.7	13.9	17.1	
Lactate	1703	44%	1.86	1.09	0.7	1.2	1.6	2.2	9.9	
Lymphocyte (%)	871	72%	15.7	9.58	1	8.5	13.6	20.9	54.1	
PCO2	2087	32%	42.4	9.54	23	36	41	47	87	
pH	2085	32%	7.38	0.0833	7.01	7.35	7.39	7.43	7.53	
Platelets	668	78%	221	97.8	27	155	202	268	734	
Potassium	641	79%	4.21	0.684	2.9	3.7	4.1	4.6	7	
PT	2284	25%	15.4	4.25	12.2	13.5	14.3	15.6	48.1	
PTT	2293	25%	36.2	13.6	22	29.6	32.9	38	153	
RDW	688	78%	14	2.12	11.1	12.6	13.6	14.8	25.9	
Sodium	639	79%	138	5.48	123	135	138	141	163	
Bilirubin (Total)	984	68%	0.691	0.445	0.3	0.4	0.6	0.8	4.4	
Troponin-I	2605	15%	0.589	1.78	0.01	0.02	0.06	0.24	15.9	
Outcomes										
Critical Event	0	100%	0.166	0.372	0	0	0	0	1	
Mortality	0	100%	0.0606	0.239	0	0	0	0	1	

570

571 **Supplementary Table 3. Final XGBoost Model Hyperparameters.**

572 Final hyperparameters for the XGBoost classifier at day 7 for critical event and mortality

573 prediction after tuning using a grid-search to optimize for AUC-ROC.

574

Hyperparameter	Critical outcome				Mortality			
	3 Days	*5 Days*	*7 Days*	*10 Days*	*3 Days*	*5 Days*	*7 Days*	*10 Days*
reg_alpha	1.0	0.10	0.01	0.1	1.0	1.0	0.1	0.1
min_child_weight	5.0	3.0	3.0	7.0	5.0	1.0	1.0	5.0
max_depth	9.0	6.0	6.0	6.0	6.0	9.0	6.0	6.0
learning_rate	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.1
gamma	0.4	0.3	0.2	0.1	0.3	0.2	0.3	0.4
colsample_bytree	0.3	0.3	0.4	0.3	0.3	0.3	0.3	0.5

575

576 **Supplementary Table 4. Model Performance Across Each Hospital.**
577 Performance of the XGBoost classifier by hospital site, as measured by accuracy
578 (ACC), area under the receiver operating curve (AUCROC), area under the precision
579 recall curve (AUCPRC), F1-score (F1), sensitivity (SENS), and specificity (SPEC). "OH"
580 refers to all hospitals in the external validation set (i.e. MSW, MSM, MSB, MSQ).
581 Outcomes are structured by "<outcome>_<day>", where outcome is either a critical
582 event (CRITICAL) or mortality (MORTALITY) and the time frame it was predicted over.
583 "OUTCOME PROP" refers to the portion of the dataset with the respective outcome.
584

FACILITY	OUTCOME	PATIENTS	PROP OUTCOME	ACC	AUCROC	AUCPRC	F1	SENS	SPEC
MSH	CRITICAL_3	1225	0.17	0.90	0.89	0.71	0.59	0.84	0.90
	CRITICAL_5	1156	0.22	0.84	0.84	0.67	0.53	0.75	0.85
	CRITICAL_7	1052	0.25	0.81	0.83	0.66	0.53	0.71	0.83
	CRITICAL_10	931	0.28	0.80	0.83	0.67	0.57	0.70	0.82
	MORTALITY_3	1225	0.03	0.97	0.87	0.26	0.03	1.00	0.97
	MORTALITY_5	1156	0.05	0.95	0.84	0.37	0.13	0.56	0.95
	MORTALITY_7	1052	0.07	0.93	0.87	0.49	0.21	0.71	0.94
	MORTALITY_10	931	0.12	0.90	0.84	0.50	0.36	0.65	0.91
MSQ	CRITICAL_3	459	0.20	0.81	0.79	0.52	0.21	0.60	0.82
	CRITICAL_5	440	0.29	0.73	0.79	0.59	0.21	0.67	0.73
	CRITICAL_7	414	0.36	0.69	0.77	0.65	0.31	0.71	0.68
	CRITICAL_10	390	0.41	0.69	0.80	0.73	0.46	0.80	0.67
	MORTALITY_3	459	0.14	0.86	0.77	0.42	0.00	-	0.86
	MORTALITY_5	440	0.23	0.80	0.76	0.56	0.21	0.92	0.79
	MORTALITY_7	414	0.30	0.74	0.77	0.61	0.25	0.86	0.73
	MORTALITY_10	390	0.37	0.68	0.78	0.67	0.29	0.74	0.67
MSB	CRITICAL_3	460	0.17	0.85	0.77	0.45	0.20	0.75	0.85
	CRITICAL_5	426	0.28	0.75	0.80	0.63	0.19	0.93	0.74
	CRITICAL_7	397	0.35	0.68	0.76	0.61	0.17	0.81	0.67
	CRITICAL_10	370	0.41	0.65	0.80	0.71	0.32	0.79	0.63
	MORTALITY_3	460	0.10	0.90	0.79	0.39	0.04	1.00	0.90
	MORTALITY_5	426	0.21	0.80	0.83	0.58	0.11	1.00	0.80
	MORTALITY_7	397	0.28	0.75	0.82	0.66	0.22	1.00	0.74
	MORTALITY_10	370	0.36	0.67	0.79	0.67	0.19	0.88	0.66
MSW	CRITICAL_3	351	0.12	0.89	0.81	0.46	0.31	0.53	0.90
	CRITICAL_5	331	0.15	0.86	0.78	0.45	0.43	0.57	0.89
	CRITICAL_7	307	0.16	0.87	0.81	0.52	0.44	0.64	0.89
	CRITICAL_10	280	0.20	0.83	0.80	0.58	0.49	0.59	0.87
	MORTALITY_3	351	0.02	0.98	0.89	0.09	0.00	-	0.98
	MORTALITY_5	331	0.03	0.97	0.86	0.23	0.00	-	0.97
	MORTALITY_7	307	0.05	0.95	0.88	0.38	0.00	0.00	0.95

	MORTALITY_10	280	0.07	0.95	0.80	0.41	0.40	1.00	0.95
MSM	CRITICAL_3	560	0.16	0.82	0.73	0.34	0.20	0.36	0.85
	CRITICAL_5	525	0.22	0.78	0.73	0.41	0.30	0.50	0.81
	CRITICAL_7	481	0.26	0.76	0.75	0.50	0.34	0.63	0.78
	CRITICAL_10	420	0.32	0.75	0.78	0.64	0.50	0.68	0.76
	MORTALITY_3	560	0.06	0.94	0.89	0.43	0.00	-	0.94
	MORTALITY_5	525	0.10	0.90	0.87	0.48	0.04	1.00	0.90
	MORTALITY_7	481	0.14	0.87	0.84	0.49	0.14	1.00	0.87
	MORTALITY_10	420	0.20	0.84	0.84	0.61	0.34	0.82	0.84
OH	CRITICAL_3	1830	0.17	0.84	0.77	0.43	0.23	0.57	0.85
	CRITICAL_5	1722	0.24	0.78	0.76	0.51	0.26	0.65	0.79
	CRITICAL_7	1599	0.29	0.73	0.74	0.52	0.30	0.60	0.75
	CRITICAL_10	1460	0.34	0.73	0.79	0.66	0.44	0.77	0.73
	MORTALITY_3	1830	0.08	0.92	0.83	0.38	0.03	1.00	0.92
	MORTALITY_5	1722	0.15	0.86	0.82	0.52	0.08	1.00	0.86
	MORTALITY_7	1599	0.20	0.82	0.84	0.58	0.18	0.86	0.82
	MORTALITY_10	1460	0.26	0.79	0.83	0.64	0.40	0.74	0.79

585
586

587

588 **Supplementary Table 5. Model Performance At Extended Time Frames.**

589 Performance of the XGBoost classifier on validation set from MSH and external

590 validation set (MSB, MSW, MSQ, MSM), as measured by accuracy (ACC), area under

591 the receiver operating curve (AUCROC), area under the precision recall curve

592 (AUCPRC), F1-score (F1), sensitivity (SENS), and specificity (SPEC). "OH" refers to all

593 hospitals in the external validation set (i.e. MSW, MSM, MSB, MSQ). Outcomes are

594 structured by "<outcome>_<day>", where outcome is either a critical event (CRITICAL)

595 or mortality (MORTALITY) and the time frame it was predicted over. "OUTCOME

596 PROP" refers to the portion of the dataset with the respective outcome.

597

FACILITY	OUTCOME	PATIENTS	PROP OUTCOME	ACC	AUCROC	AUCPRC	F1	SENS	SPEC
MSH	CRITICAL_15	792	0.27	0.81	0.84	0.68	0.60	0.71	0.84
	CRITICAL_20	719	0.25	0.82	0.85	0.68	0.58	0.70	0.85
	MORTALITY_15	792	0.16	0.87	0.86	0.58	0.58	0.64	0.89
	MORTALITY_20	719	0.18	0.86	0.87	0.64	0.54	0.67	0.87
OH	CRITICAL_15	1320	0.37	0.72	0.75	0.66	0.53	0.68	0.73
	CRITICAL_20	1254	0.36	0.75	0.81	0.71	0.57	0.75	0.75
	MORTALITY_15	1320	0.32	0.74	0.81	0.65	0.40	0.75	0.74
	MORTALITY_20	1254	0.34	0.75	0.84	0.72	0.51	0.77	0.75

598
599
600

601 **Supplementary Table 6. Logistic Regression and Generalized Additive Model**
602 **Performance.**
603 Performance of the four models using only Age and all the features, as measured by
604 accuracy (ACC), area under the receiver operating curve (AUCROC), area under the
605 precision recall curve (AUCPRC), F1-score (F1), sensitivity (SENS), and specificity
606 (SPEC). Models are structured by "<model>_<feature space>", where model is either a
607 logistic regression (LOG) or generalized additive model (GAM) and the features used in
608 the model.

MODEL	OUTCOME	PATIENTS	PROP OUTCOME	ACC	AUCROC	AUCPRC	F1	SENS	SPEC
LOG_AGE	CRITICAL_3	3055	0.17	0.83	0.57	0.22	0	1	0
	MORTALITY_3	3055	0.06	0.92	0.70	0.18	0	1	0
GAM_AGE	CRITICAL_3	3055	0.17	0.83	0.52	0.19	0.01	1	0.003
	MORTALITY_3	3055	0.06	0.98	0.62	0.14	0	1	0
LOG_ALL	CRITICAL_3	1278	0.18	0.81	0.74	0.33	0.31	0.90	0.29
	MORTALITY_3	1278	0.02	0.96	0.67	0.06	0.07	0.98	0.07
GAM_ALL	CRITICAL_3	1278	0.18	0.82	0.74	0.37	0.35	0.92	0.31
	MORTALITY_3	1278	0.02	0.98	0.62	0.14	0	1	0

609