Using Shapes of COVID-19 Positive Patient-Specific Trajectories for Mortality Prediction

Alaleh Azhir, MSc[1], Soheila Talebi, MD[2], Louis-Henri Merino, BSc[3], Thomas Lukasiewicz, PhD[1], Edgar Argulian, MD, MPH[2], Jagat Narula, MD, PhD[2], Borislava Mihaylova, MSc, DPhil[1, 4]

[1]University of Oxford, Oxford, UK; [2]Icahn School of Medicine at Mount Sinai, New York, NY, USA; [3]Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland; [4]Queen Mary University of London, London, UK

Abstract

Machine learning can be used to identify relevant trajectory shape features for improved predictive risk modeling, which can help inform decisions for individualized patient management in intensive care during COVID-19 outbreaks. We present explainable random forests to dynamically predict next day mortality risk in COVID -19 positive and negative patients admitted to the Mount Sinai Health System between March 1st and June 8th, 2020 using patient time-series data of vitals, blood and other laboratory measurements from the previous 7 days. Three different models were assessed by using time series with: 1) most recent patient measurements, 2) summary statistics of trajectories (min/max/median/first/last/count), and 3) coefficients of fitted cubic splines to trajectories. AUROC and AUPRC with cross-validation were used to compare models. We found that the second and third models performed statistically significantly better than the first model. Model interpretations are provided at patient-specific level to inform resource allocation and patient care.

Introduction

With the rise of the Delta variant of Covid-19 in June 2021 and Omicron in December 2021, the situation in various countries such as Iran, is to a degree reminiscent of the surge of cases back in mid-April 2020 in New York. The allocation of resources to those most likely to survive is an important and difficult decision that healthcare workers face every day in the pandemic. Covid-19 involves an intricate interplay of various interdependent biological pathways and despite epidemiological and clinical characteristics of patients with COVID-19 in various parts of the world having been reported[1,2], a dynamically updated assessment of patients' disease progression over time is lacking. Monitoring the changes of these measurements over time and dynamically assessing mortality risks for Covid-19 patients could aid in a better allocation of resources.

The increase in automating data extraction from electronic health records has spurred recent efforts to use machine learning to predict individual patient's mortality risk using dynamic time-series data.[3,4] Although scoring systems for ICU populations, such as the Simplified Acute Physiology Score (SAPS)[5], the Acute Physiologic and Chronic Health Evaluation (APACHE)[6], and the Mortality Prediction Model[7] for ICU populations can be useful in predicting mortality risk, they are usually calculated based only on data collected at admission and lack precision.[3,8] Using the information that is captured in patient's measurements during hospital stay, including changes in patient indicators overtime could improve mortality risk scoring.

Previous studies have suggested the use of summary statistics of trajectories (maximum, minimum, median, etc.)[3] or fitting cubic spline and extracting relevant features[4] as inputs for mortality prediction as possible approaches to address this limitation. In this paper, we demonstrate that, compared to including the most recent measurements, each among the use of summary statistics and the automated extraction of cubic spline coefficients by random forest classifiers statistically improves both the area under the receiver operating characteristic (AUROC) curve and the area under the precision- recall curve (AUPRC) of predicting the mortality risk for patients with COVID-19. Such approaches can be used to improve any risk model that contains time series data. Furthermore, we improve on these previous studies by providing a patient-specific model interpretation to facilitate random forest model adoption in a health care setting.

Methods

Data Preprocessing: We retrospectively retrieved de-identified health records of patients tested for Covid-19 admitted to five hospitals within the Mount Sinai Health System (MSHS) between March 1st and June 8th, 2020; approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai. The data included patient demographics, past medical history, daily vitals, and all cardiac, liver panel, blood test, and other labs during their hospital stays (some displayed in Table 1). Relevant patient events and outcomes (ICU admission during patient stay, COVID-19 positive or negative test, and outcome: discharge to home, discharge to hospice care, or death) were recorded. Patients with multiple hospitalizations were removed from the dataset. To remove questionable laboratory measurements, cut off values for ranges as found in the literature were applied for all continuous measurements, and the measurements outside of these ranges were set as missing.

Dataset Generation: To use time for mortality predictions[4], for each patient encounter, we generated (overlapping) observational time period units every day, where each unit has a maximum size of 7 days. For example, a patient hospital stay that lasted 9 days generates the following observational units: (0,1] day, (0,2] days, ..., (0,7] days, (1,8], and (2, 9] days. For each observational unit, the outcome of interest is death within the next day of the end of the unit's time period.

Risk Modeling: We adapted the models presented by Ma et al. to fit our Covid-19 cohorts.[4] Different risk factors in each observational unit can be fed into any classifier to estimate the next day probability of mortality. We used Ranger, a fast implementation of a random forest classifier, due to its superior performance as a nonparametric method.[9,10] Random forests not only avoid overfitting but can also handle large, including dependent, data with numerous features, as included in this dataset.[10] Our main hyperparameters are number of trees and the number of features to be considered for each tree (mtry). We found that 500 trees provided the lowest out-of-bag error rates while not compromising the computational time, and fine-tuned mtry for each of our models using the R Caret package.[11] Three nested random forest models were applied to each observational unit to assess the use of trajectory data in next-day mortality prediction. They are defined below:

Model 1: This model includes static at admission patient information (e.g., age) and last recorded value of each longitudinally measured risk factor.

Model 2: This model appends the summary characteristics of longitudinal risk factor trajectories (minimum, first, last, median, maximum, and count) for each predictor to the information included in Model 1.

Model 3: This model replaces the summary characteristics in Model 2 with the cubic spline coefficients after fitting penalized smoothing splines using Leave-One-Out cross validation.

Since Model 3 fits a cubic spline, to avoid overfitting, we only fitted and compared Model 3's performance to those of Models 1 and 2 in patients whose length of stay exceeded 4 days. As the range for non-missing values across all predictors was bounded away from -10^{10}, we encoded missing values with this number to instruct our tree-based models to treat them differently.

When compared to Ma et al. paper[4], in addition to having a different time frame and window (7 days vs. 24 hours), we included observational units that were shorter than our window size of 7 days, whereas they only include patients that had a stay longer than their window size of 24 hours. Furthermore, in spline fitting for our Model 3, the number of the knots that we chose for smoothing splines was 4 instead of 27, since we had fewer daily measurements.

Model Evaluation: ROC curves are commonly used for model evaluation, however, can be misleading when outcomes (death vs. discharge to home) are highly imbalanced, therefore, we also considered the precision-recall curves.[12,13] We used Monte Carlo cross validation to evaluate and compare the performance of our 3 models, by performing 20 random splits of observational units into training (70%) and validation (30%) sets. For each split, we fit a ranger random forest to the training set, predict the probability of death in the validation set, and evaluate the AUPRC and AUROC. We used the nonparametric Wilcoxon paired signed-rank test to compare AUROC and AUPRC between our three models. Though the baseline value for AUROC is 0.5 (for a random coin toss), the baseline value for AUPRC depends on the fraction of positives in the dataset. For example, in our case, since 3.4% of Covid-19 positive observational units resulted in death, our baseline AUPRC value is 0.034. The R pROC package[14] was used to compute the 95% confidence interval (CI) of the sensitivity at the given specificity points, with 2000 stratified bootstrap replicates. We replicated the above analysis among COVID-19 negative patients.

Model Interpretation: We applied Shapley additive explanations (SHAP) algorithm to our best performing model to obtain explanations of the risk factor features that drive patient-specific predictions.[15,16] The SHAP value for a feature does not specify its direct isolated effect but its compound effect while interacting with other features in the model.

Our code is publicly available at https://github.com/Alaleh1191/Covid-19.

Results

Our study includes 3699 COVID-19 positive and 5488 COVID-19 negative patients, admitted between March 1st and June 8th 2020. Patients' median age is 67 years for COVID-19 positive and 59 years for COVID-19 negative patients. 57% of the COVID-19 positive are male and 45% of the COVID-19 negative are male. The overall median BMI is 27.0 (27.7 for COVID-19 positive, 26.5 for COVID-19 negative). The median length of hospital stay is 5 days. Laboratory markers (such as troponin, WBC count, etc.) were categorized and tracked during the first month of patients' admission. Selected characteristics of the study patients are presented in Table 1.

Table 1: Characteristics of the hospitalized COVID-19 positive and negative patients. Continuous variables are presented as mean (interquartile (IQR) range), and categorical variables are presented as count (% percentage).

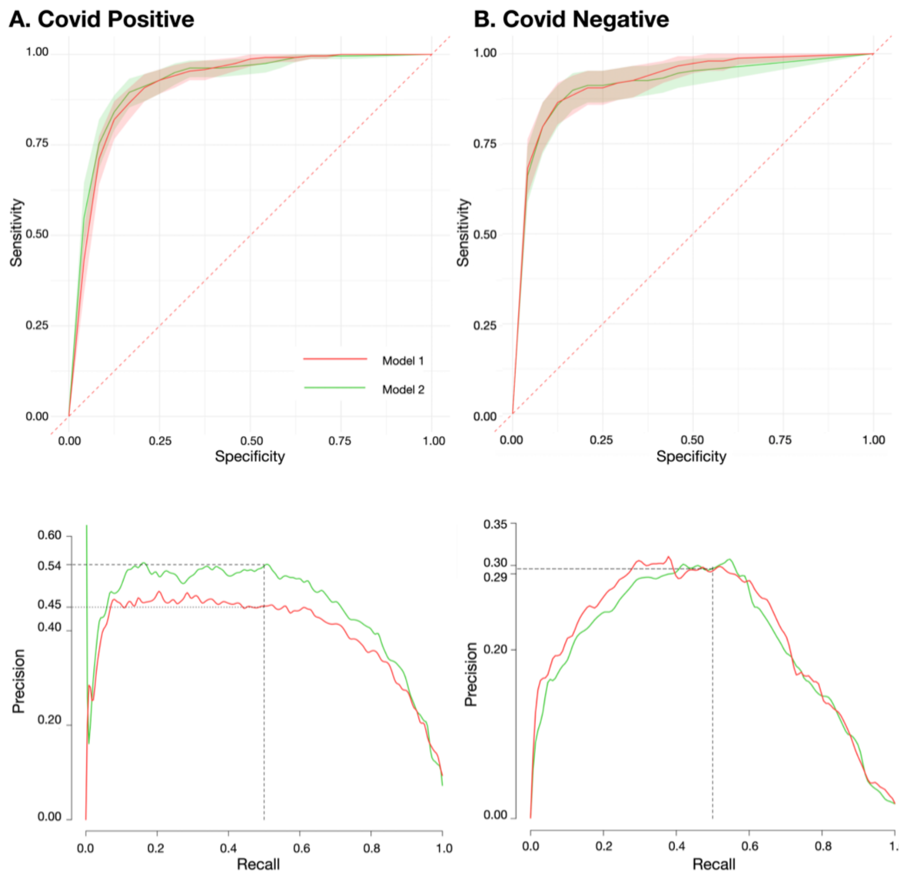
	Covid-19 Positive		**Covid-19 Negative**	
Characteristic	**Survivors**	**Non- survivors**	**Survivors**	**Non-survivors**
Sex, Female	1120 (44 %)	480 (42 %)	2776 (55 %)	241 (52 %)
Missing	0 (0%)	0 (0%)	2 (0.0%)	0 (0%)
Age, years	63 (51 - 73)	75 (65 - 84)	58 (35 - 73)	75 (61 - 85)
BMI	28 (24 - 33)	27 (24 - 33)	27 (23 - 31)	25 (21 - 31)
Underweight	80 (3 %)	56 (5 %)	401 (8 %)	43 (9 %)
Normal	658 (26 %)	299 (26 %)	1516 (30 %)	170 (37 %)
Overweight	790 (31 %)	324 (28 %)	1596 (32 %)	107 (23 %)
Obese	883 (35 %)	368 (32 %)	1413 (28 %)	124 (27 %)
Missing	134 (5.3%)	107 (9.3%)	100 (2.0%)	18 (3.9%)
Smoking Status				
Never	1397 (55 %)	533 (46 %)	2575 (51 %)	207 (45 %)
Previous or Current	635 (25 %)	327 (28 %)	1712 (34 %)	168 (36 %)
Missing	513 (20.2%)	294 (25.5%)	739 (14.7%)	87 (18.8%)
Race and Ethnicity				
African American	711 (28 %)	289 (25 %)	1355 (27 %)	113 (24 %)
American Indian or Alaska Native	2 (0 %)	0 (0 %)	3 (0 %)	0 (0 %)
Asian	121 (5 %)	52 (5 %)	265 (5 %)	21 (5 %)
White	543 (21 %)	314 (27 %)	1614 (32 %)	169 (37 %)
Hispanic	743 (29 %)	295 (26 %)	1076 (21 %)	79 (17 %)
Native Hawaiian or Pacific Islander	3 (0 %)	1 (0 %)	8 (0 %)	0 (0 %)
Other	347 (14 %)	158 (14 %)	549 (11 %)	61 (13 %)
Unknown	75 (3 %)	45 (4 %)	156 (3 %)	19 (4 %)
Hospital Location				
Brooklyn	372 (15 %)	258 (22 %)	475 (9 %)	71 (15 %)
Queens	302 (12 %)	261 (23 %)	430 (9 %)	56 (12 %)
Manhattan – St. Luke's	515 (20 %)	218 (19 %)	879 (17 %)	87 (19 %)
Manhattan – West	363 (14 %)	93 (8 %)	1004 (20 %)	66 (14 %)
Manhattan – East.	993 (39 %)	324 (28 %)	2238 (45 %)	182 (39 %)
Comorbidities				
Asthma	137 (5 %)	52 (5 %)	352 (7 %)	26 (6 %)

COPD	86 (3 %)	58 (5 %)	245 (5 %)	31 (7 %)
Sleep Apnea	48 (2 %)	29 (3 %)	74 (1 %)	13 (3 %)
Diabetes	487 (19 %)	279 (24 %)	884 (18 %)	85 (18 %)
Chronic Kidney Disease	257 (10 %)	154 (13 %)	538 (11 %)	60 (13 %)
Cancer	188 (7 %)	102 (9 %)	502 (10 %)	112 (24 %)
Coronary Artery Disease	254 (10 %)	168 (15 %)	609 (12 %)	68 (15 %)
Atrial Fibrillation	113 (4 %)	106 (9 %)	401 (8 %)	51 (11 %)
Heart Failure	153 (6 %)	110 (10 %)	499 (10 %)	61 (13 %)
Chronic Viral Hepatitis	29 (1 %)	8 (1 %)	66 (1 %)	9 (2 %)
Liver Disease	57 (2 %)	36 (3 %)	145 (3 %)	24 (5 %)
ARDS	13 (1 %)	11 (1 %)	5 (0 %)	0 (0 %)
Acute Kidney Injury	135 (5 %)	105 (9 %)	240 (5 %)	48 (10 %)
Acute venous thromboembolism	26 (1 %)	7 (1 %)	64 (1 %)	4 (1 %)
Cerebral Infarction	31 (1 %)	8 (1 %)	103 (2 %)	13 (3 %)
Intracerebral hemorrhage	5 (0 %)	2 (0 %)	18 (0 %)	5 (1 %)
Acute MI	32 (1 %)	25 (2 %)	79 (2 %)	9 (2 %)
Length of Stay	6.9 (4.0 - 11)	8.8 (5.0 - 15)	4.5 (2.9 - 7.9)	9.2 (5.2 - 17)
ICU (yes: anypoint during admission)	316 (12 %)	521 (45 %)	762 (15 %)	204 (44 %)

In total, there were 33,864 observational units for the COVID-19 positive patients and 32,589 observational units for the COVID-19 negative patients. 3.4% and 1.4% of all COVID-19 positive and negative observational units resulted in death, respectively.

We then divided the study patient population into two groups, those with shorter length of stay (≤ 4 days), and those with longer length of stay (> 4 days). Since most of repeated risk factor measurements (besides vital signs) are only measured daily, to avoid overfitting, we only compared Models 1 and 2 for patients with length of stay up to 4 days. As expected, among these patients due to shorter trajectories, only a slight improvement in performance for Model 2 compared to Model 1 is observed (Figure 1); both models, however, have a high predictive power.

The ROC curves and precision-recall curves (PRC) of the three next-day mortality risk models for COVID-19 positive patients with length of stay above 4 days, including comparisons of the AUROC and AUPRC for the next day mortality averaged over the 20 random data splits are presented (Figure 2). Models 2 and 3 have a better precision than Model 1 for both COVID-19 positive and negative patients (all p-values < 0.01). The AUROC of Model 2 is larger than that of Model 1 (p-values < 0.0001). The AUROC of Model 3, however, did not differ significantly from that of Model 1. Furthermore, comparing Models 2 and 3, Model 2's AUROC is significantly larger than that of Model 3 in both COVID-19 positive and negative patients (both p-values < 0.001). However, Model 2's AUPRC is only statistically significantly higher than that of Model 3 in COVID-19 positive patients. This suggests that Model 2, by using key statistical summaries of trajectories (minimum, maximum, median, first, and last), performs better than fitting cubic splines (Model 3) and conveys additional predictive information that is not captured by the most recent measurements alone (Model 1). A similar analysis confirmed the superiority of Model 2 to Models 1 and 3 in Covid-19 negative patients. All three models, however, performed significantly better than a random forest baseline survival model fitted only on initial patient admission data (95% AUROC CI index: 0.850 - 0.857, 95% AUPRC CI index: 0.380 - 0.400).

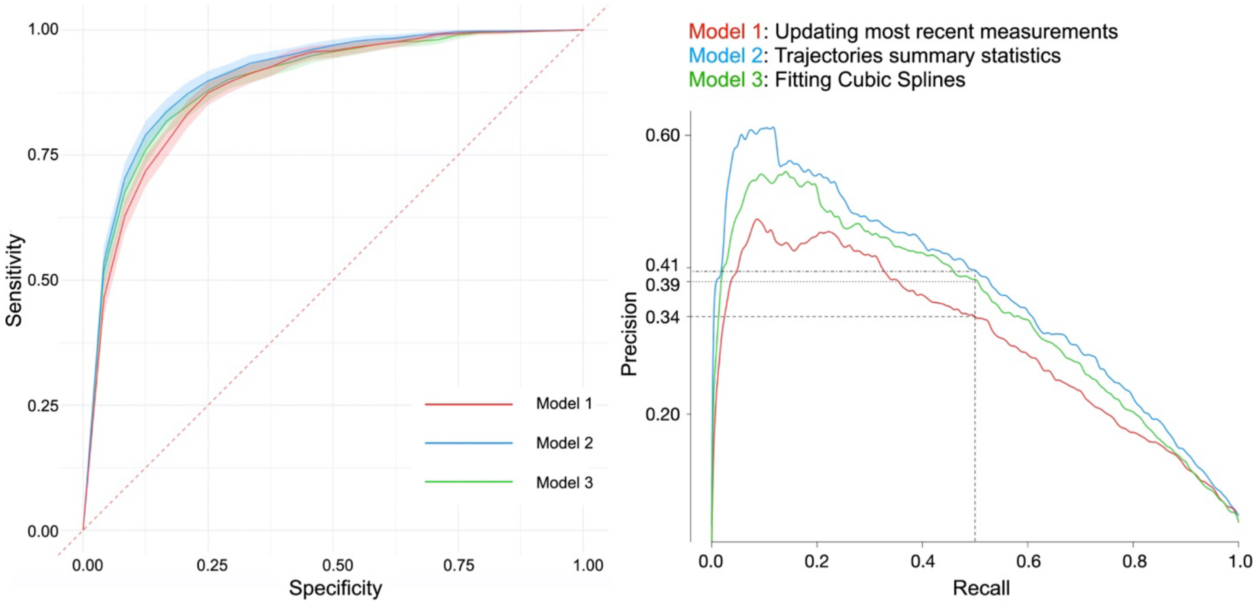


	Model 1: Updating Most Recent Values	Model 2: Trajectories summary statistics
AUROC Covid-19 Positive	0.917 (0.006)	0.927 (0.009); <0.0001
AUPRC Covid-19 Positive	0.415 (0.006)	0.520 (0.012); <0.0001
AUROC Covid-19 Negative	0.934 (0.003)	0.930 (0.008); 0.0063
AUPRC Covid-19 Negative	0.248 (0.015)	0.234 (0.015); 0.0192

Figure 1. Average receiver operating characteristics and precision (%) versus recall (%) plot of Models 1 (most recent measurements only) and 2 (trajectories' summary statistics) of next day mortality risk among Covid-19 positive and negative patients with length of stay ≤ 4 days. AUROC and AUPRC are represented as mean (standard error) for Model 1 column and as mean (standard error); p-value for Model 2 column. p-values are obtained a two-sided Wilcoxon paired signed rank test comparing Model 2 to Model 1 for patients with length of stay ≤ 4 days.

To understand which features are most important in the best performing model (Model 2) among COVID-19 positive patients with length of stay above 4 days, both the SHAP values and Gini Importance values of the top features are summarized (Figure 3). This reveals that lower O2 saturation, lower blood pressure, older age, higher anion gap, and higher white blood cell count are associated with a higher probability of death. However, certain features such as higher troponin level, respiratory rate, and heart rate can be associated with a higher or lower probability of survival depending on patients' other risk factors. The importance values of predictors were also calculated in COVID-19 negative patients, and the top predictors in those patients were vitals, followed by length of stay followed by an intermix of kidney and complete blood count markers.

To illustrate the influence of features of risk factors, Model 2 predictions are presented for 2 COVID-19 patients from our test cohort with the same sex and age who were admitted to the ICU and followed for up to 22 days. Figure 4 displays features that contributed to the mortality risk prediction for these two patients evaluated using Model 2 after 1, 2 and 3 weeks in hospital. The first case is a 66-year-old male with a history of hypertension, systolic heart failure, atrial fibrillation, and end-stage renal disease with BMI 40 presented with dyspnea. On admission, the vital signs were recorded as blood pressure of 150/100 mmHg, heart rate 140bpm, respiratory rate of 38bpm, and O2 saturation of 91%. The initial lab was significant for leukocytosis with lymphopenia, increased coagulation factors, abnormal troponin, and increased liver enzyme (WBC: 32K/uL, neutrophil: 93%, lymphocyte: 0.6%, platelet: 277K/uL, AST: 93U/L, albumin: 2.9g/dl, troponin: 1.19ng/ml, creatinine: 1.84mg/dL, potassium: 5.6mmol/L, calcium: 8.5mg/dL, LDH: 2212U/L, CRP: 419mg/L, ferritin: 393ng/ml, and D-Dimer: 12.9ug/mL). He had a worsening of liver function and increased LDH and ferritin level. After a hospital stay of 21 days, he was discharged to rehab. In this case, the patient's mortality risk decreases over time from 0.06 to 0.01. Predicting death after 21 days in the hospital, glucose median of 68.5, last recorded creatinine of 3.69, and min respiratory rate of 25 drive the mortality prediction towards non-survival, whereas the last recorded systolic blood pressure, O2 saturation, neutrophil, and heart rate are the most important feature pulling the prediction towards survival.



	AUROC	AUPRC	pval ROC	pval PRC
Model 1	0.897 (0.894 - 0.900)	0.424 (0.414 - 0.434)	.	.
Model 2	0.903 (0.899 - 0.906)	0.450 (0.441 - 0.459)	<0.0001	<0.0001
Model 3	0.896 (0.892 - 0.899)	0.432 (0.422 - 0.441)	0.28	0.0023

Figure 2. Average receiver operating characteristics and precision (%) versus recall (%) plots of Models 1 to 3 for next day mortality prediction in Covid-19 positive patients. AUROC and AUPRC are represented as mean (interquartile (IQR) range). p-values are obtained a two-sided Wilcoxon paired signed rank test comparing Model 2 and 3 to Model 1 for patients with length of stay > 4 days.

The second case is a 66-year-old male with BMI 29 and no significant past medical history, who presented with dyspnea. On arrival, the vital signs were recorded as blood pressure 130/80mmHg, heart rate 83bpm, respiratory rate 29bpm, and O2 saturation of 91%. The initial lab revealed a leukocytosis with lymphopenia, kidney injury, mild increase of liver enzyme, and inflammatory factors (WBC: 8.1K/uL, neutrophil: 89.75%, lymphocyte: 1.6%, platelet: 191K/uL, AST: 43U/L, albumin: 2.55g/dl, troponin: 0.013ng/ml, creatinine: 1.84mg/dL, potassium: 5mmol/L, calcium: 8.5mg/dL, LDH: 458U/L, CRP: 298mg/L, ferritin: 1063ng/ml and D-Dimer: 0.85ug/mL). During hospital stay, the lowest oxygen saturation was 80% and mild increase of creatinine, D-Dimer, and liver enzyme were observed. He died in ICU 22 days after admission. In this case, we can see that this patient's predicted mortality risk increases over time from 0 to 0.25. Predicting survival after 21 days in the hospital, maximum lymphocyte, age, and minimum basophil pull the risk down towards survival (blue arrows), whereas last systolic and diastolic blood pressure, PCO2 venous, BUN, and minimum O2 saturation drive mortality prediction towards non-survival (red arrows). Such dynamic risk prediction by visualizing the main features contributing to non-survival at any point in time can aid clinicians determine whether there are actions that can be taken to lower the mortality risk.

Discussion

Although many mortality-risk models have been recently developed for Covid-19 patients, they mostly use static data at admission. In this study, we developed a dynamic next-day mortality risk prediction model incorporating time-updated patient information on the past week's lab measurements. Our results suggest that trajectories' summary statistics significantly improve prediction performance compared to including most recent measurements or fitted cubic spline coefficients. The model's high classification power (AUROC = 0.902, AUPRC = 0.450), suggest that it can be used by clinicians to identify patients at immediate risk, and identify factors contributing to their increased risk of next day mortality for implementing better individualized treatments. The method was also beneficial in improving the next day mortality prediction in COVID-19 negative patients. Lastly, given the evolving COVID-19 landscape, it is important to re-train these models on more recent COVID-19 patient datasets and externally validate them in other hospitals. Due to its small computational cost and public code availability, this method can be easily implemented and continuously updated using new patients' time series data to improve mortality prediction among hospitalized patients.

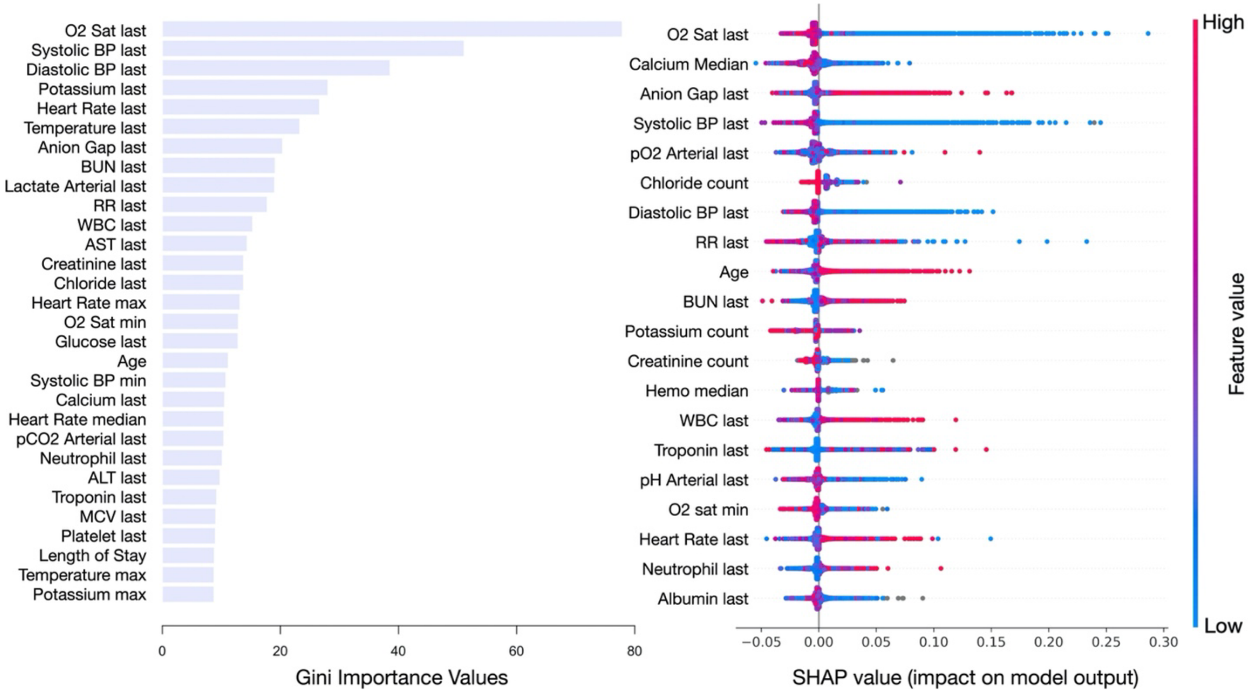


Figure 3. Model 2 (with trajectories' summary statistics): Gini Importance values for top 30 features (left) and SHAP values for the top 20 features (right) for COVID-19 positive patients. The Gini Importance values were calculated by Model 2 random forest model. SHAP values for the top 20 features calculated by sum of SHAP value magnitudes over all Covid-19 positive observations. Higher SHAP values indicates higher probability of death. All patients in the dataset are run through Model 2, and a dot is created for each person on each feature's line.

Whereas our results found that trajectories' summary statistics performs best, Ma et al.'s paper found fitting cubic spline coefficients performs best.[4] The differences between our findings can be due to most of our laboratory factors (besides vitals) being measured only once daily, thus not providing as many datapoints for fitting a spline as are present in their dataset. Furthermore, we also include observational units that were shorter than our maximum observation unit size of 7 days, whereas they only looked at observational units of their maximum size to obtain longer trajectories. These differences in model design could have led to different findings.

Our approach also allows for a clinically interpretable understanding of its top features driving mortality for both an individual patient and the entire dataset. Whereas prior studies reported different outcomes associated with high temperature,[17–19] our study shows that although the non-survivor group has a statistically significantly higher temperature both at admission and during their stay compared to the survivor group, the average daily temperature for the non-survivor group still falls below the fever cutoff (100.4ºF). O2 saturation was identified by Gini and SHAP importance scores to be the most important feature overall in predicting next day mortality, and as suggested by previous studies, is an important indicator for guiding physicians when to require admission to the ICU.[20,21]

Full Blood Count trajectories displayed more frequent leukocytosis, neutrophilia, and lymphopenia in non-survivors. Neutrophilia, as an expression of hyperinflammatory state, may also indicate a superimposed bacterial infection.[22,23] Lymphopenia could occur potentially due to high ACE2 receptor expression on lymphocytes causing increased susceptibility to Covid-19,[24] cytokine storms causing lymphocytes' apoptosis,[25] or injured alveolar epithelial cells inducing the infiltration of lymphocytes.[26,27]

Elevated troponin was observed in non-survivors and was deemed as one of the top predictors of next day mortality by SHAP and Gini. This could be secondary to cytokine storm, myocarditis, pro-thrombotic state, or demand ischemia, which all may contribute to the observed poor prognosis.[28–30]

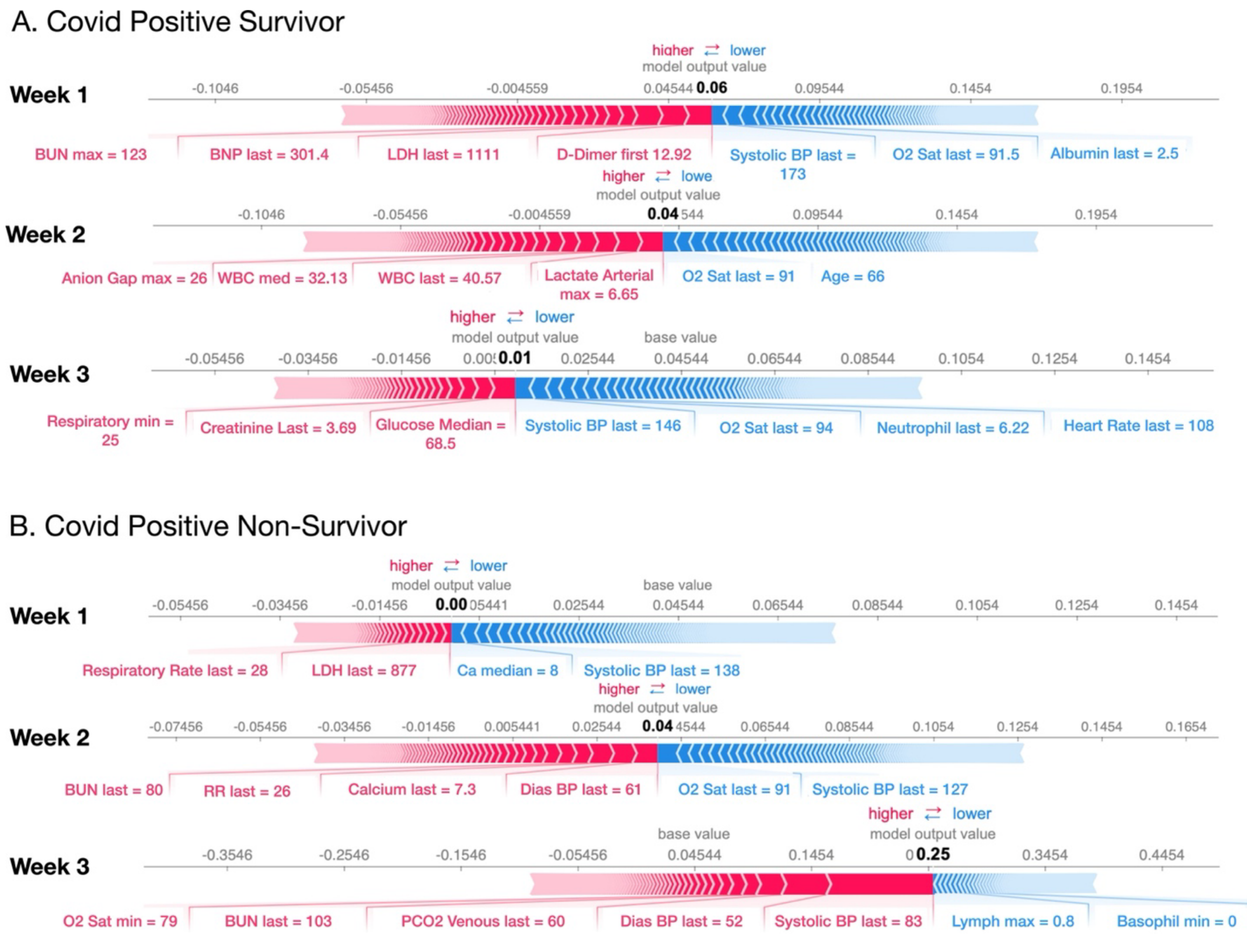


Figure 4. Force plots obtained using SHAP values for 2 COVID-19 positive patients in a test dataset, one survivor and one non-survivor. SHAP values identify risk factor features associated with higher (red color) or lower (blue color) mortality risk.

Despite observing elevated LDH levels in both survivors and non-survivors (as seen in our two example cases), a significantly higher LDH level was observed on average throughout patients' admission in non-survivors. The elevation of LDH has been reported as an independent factor of mortality in patients with severe acute respiratory syndrome.[31,32] Similar to past studies,[33] on average and Gini Importance plot, we also observed higher AST and ALT levels in our non-survivors on admission and throughout their stay, reflecting liver damage. Though the mechanism remains unclear, various hypothesis suggest direct SARS-CoV-2 infection of liver cells, drug-induced toxicity, immune-mediated inflammation, or hypoxia.[33–35]

Our study concurs with the growing evidence of Covid-19-induced hypercoagulable states,[36,37] as we observed increased levels of D-Dimer in non-survivors on average throughout the first two weeks of admission. On the other hand, Fibrinogen despite being high in both survivors and non-survivors, was not significantly different through patients' admission between these 2 groups. Regarding thrombocytopenia, although platelet counts fell within the normal range, non-survivors had significantly lower platelet counts during their stay.

We observed significantly lower levels of calcium in non-survivors, with a drop in average calcium levels in those whose length of stay exceeded 10 days. Furthermore, calcium level had high Gini and Shap importance in next-day mortality prediction. Previous studies have reported that calcium interacts with fusion peptides on viruses such as SARS-CoV, MERS-CoV, and Ebolavirus, promoting their replication.[38–40] Furthermore, it has been shown to be an independent risk factor in Covid-19 hospitalization.[41] A high creatinine and BUN level, and electrolyte abnormality including hyperkalemia, hyperchloremia, acidosis, and hypernatremia were seen more frequently in non-survivors compared to survivors during the first 2 weeks of stay.

Our study has several limitations. Most laboratory factors (besides vitals) were only measured daily, hence not providing a long enough time series for fitting a cubic spline trajectory (Model 3) within a few days. To overcome this limitation, we compared only the use of summary statistics of trajectories to updating the most recent value in patients with hospital length of stay ≤ 4 days and considered comparing fitting a cubic spline to these two methods only in patients with length of stay longer than 4 days. Therefore, our conclusions regarding Model 3 is restricted to patients with longer hospital stays. In addition, we only looked at observational units with a maximum size of 7 days, longer units could be explored in a future study. Furthermore, different hospitals used different protocols in treating COVID-19 positive patients, therefore leading to different frequencies at which labs (i.e., troponin or D-Dimer) were measured. As noted in Table 1, for some measurements, we have a lot of missing values, therefore future work needs to be done to validate the model in further cohorts. Moreover, the Covid-19 negative dataset includes patients who were admitted to the hospital during the severe months of the pandemic in New York, thus are only representative of hospital admissions during the pandemic. Finally, further features of trajectories (curvature, arc length, etc.) may also prove informative for risk assessment and could be assessed in further studies.

Conclusion

In conclusion, we developed an explainable random forest model for dynamic next-day mortality prediction in COVID-19 positive patients using a dataset of 3699 COVID-19 positive and 5488 COVID-19 negative patients in 5 hospitals. The model interpretation showed that risk factors interact and compensate for one another by pulling patients towards survival or non-survival based on patients' other characteristics. By improving the prediction of the next day mortality and identifying features with high-importance values, this new model may help healthcare institutions improve care decisions for COVID-19 positive admitted patients.

References

1. Vaid A, Somani S, Russak AJ, et al. Machine Learning to Predict Mortality and Critical Events in COVID-19 Positive New York City Patients [Internet]. 2020 Apr [cited 2021 Aug 27] p. 2020.04.26.20073411. Available from: https://www.medrxiv.org/content/10.1101/2020.04.26.20073411v1
2. Yan L, Zhang H-T, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell. 2020 May;2(5):283–8.
3. Thorsen-Meyer H-C, Nielsen AB, Nielsen AP, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. The Lancet Digital Health. 2020 Apr 1;2(4):e179–91.
4. Ma J, Lee DKK, Perkins ME, Pisani MA, Pinker E. Using the Shapes of Clinical Data Trajectories to Predict Mortality in ICUs. Critical Care Explorations. 2019 Apr;1(4):e0010.
5. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA. 1993 Dec 22;270(24):2957–63.
6. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med. 1981 Aug;9(8):591–7.
7. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA. 1993 Nov 24;270(20):2478–86.
8. Glance LG, Osler TM, Dick AW. Identifying quality outliers in a large, multiple-institution database by using customized versions of the Simplified Acute Physiology Score II and the Mortality Probability Model II0. Crit Care Med. 2002 Sep;30(9):1995–2002.
9. Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.
10. Goehry B. Random forests for time-dependent process. ESAIM: Probability and Statistics. 2020; 801–826.
11. Kuhn, M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 2008; 28(5), 1–26.
12. Grau J, Grosse I, Keilwagen J. PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. Bioinformatics. 2015 Aug 1;31(15):2595–7.
13. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE. 2015;10(3):e0118432.
14. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC bioinformatics. 2011;12(1):1–8.
15. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in neural information processing systems. 2017. p. 4765–74.
16. Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:180203888. 2018.

17. Tharakan S, Nomoto K, Miyashita S, Ishikawa K. Body temperature correlates with mortality in COVID-19 patients. Critical Care. 2020;24(1):1–3.
18. Zheng Z, Peng F, Xu B, et al. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. Journal of Infection. 2020;
19. Shi L, Wang Y, Wang Y, Duan G, Yang H. Dyspnea rather than fever is a risk factor for predicting mortality in patients with COVID-19. J Infect [Internet]. 2020 May 15 [cited 2020 Aug 8]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7228739/.
20. Xie J, Covassin N, Fan Z, et al. Association Between Hypoxemia and Mortality in Patients With COVID-19. Mayo Clin Proc. 2020 Jun;95(6):1138–47.
21. Siddiqi HK, Mehra MR. COVID-19 illness in native and immunosuppressed states: A clinical–therapeutic staging proposal. J Heart Lung Transplant. 2020 May;39(5):405–7.
22. Terpos E, Ntanasis-Stathopoulos I, Elalamy I, et al. Hematological findings and complications of COVID-19. Am J Hematol. 2020;95(7):834–47.
23. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. Clin Chim Acta. 2020 Jul;506:145–8.
24. Xu H, Zhong L, Deng J, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int J Oral Sci. 2020 24;12(1):8.
25. Mehta P, McAuley DF, Brown M, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet. 2020 28;395(10229):1033–4.
26. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science. 2005 Sep 16;309(5742):1864–8.
27. Ge X-Y, Li J-L, Yang X-L, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature. 2013 Nov 28;503(7477):535–8.
28. Lala A, Johnson KW, Januzzi JL, et al. Prevalence and Impact of Myocardial Injury in Patients Hospitalized With COVID-19 Infection. J Am Coll Cardiol. 2020 04;76(5):533–46.
29. Santoso A, Pranata R, Wibowo A, Al-Farabi MJ, Huang I, Antariksa B. Cardiac injury is associated with mortality and critically ill pneumonia in COVID-19: A meta-analysis. Am J Emerg Med. 2020 Apr 19;
30. Li J-W, Han T-W, Woodward M, et al. The impact of 2019 novel coronavirus on heart injury: A Systematic review and Meta-analysis. Prog Cardiovasc Dis. 2020 Apr 16;
31. Choi KW, Chau TN, Tsang O, et al. Outcomes and prognostic factors in 267 patients with severe acute respiratory syndrome in Hong Kong. Ann Intern Med. 2003 Nov 4;139(9):715–23.
32. Terpstra ML, Aman J, van Nieuw Amerongen GP, Groeneveld ABJ. Plasma biomarkers for acute respiratory distress syndrome: a systematic review and meta-analysis*. Crit Care Med. 2014 Mar;42(3):691–700.
33. Cai Q, Huang D, Yu H, et al. COVID-19: Abnormal liver function tests. J Hepatol. 2020 Apr 13;
34. Boettler T, Newsome PN, Mondelli MU, et al. Care of patients with liver disease during the COVID-19 pandemic: EASL-ESCMID position paper. JHEP Rep. 2020 Jun;2(3):100113.
35. Zhang C, Shi L, Wang F-S. Liver injury in COVID-19: management and challenges. Lancet Gastroenterol Hepatol. 2020;5(5):428–30.
36. Becker RC. COVID-19 update: Covid-19-associated coagulopathy. J Thromb Thrombolysis. 2020 Jul;50(1):54–67.
37. Becker RC. Covid-19 treatment update: follow the scientific evidence. J Thromb Thrombolysis. 2020 Jul;50(1):43–53.
38. Millet JK, Whittaker GR. Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. Virology. 2018;517:3–8.
39. Straus MR, Tang T, Lai AL, et al. Ca2+ Ions Promote Fusion of Middle East Respiratory Syndrome Coronavirus with Host Cells and Increase Infectivity. J Virol. 2020 Jun 16;94(13).
40. Nathan L, Lai AL, Millet JK, et al. Calcium Ions Directly Interact with the Ebola Virus Fusion Peptide To Promote Structure-Function Changes That Enhance Infection. ACS Infect Dis. 2020 14;6(2):250–60.
41. Di Filippo L, Formenti AM, Rovere-Querini P, et al. Hypocalcemia is highly prevalent and predicts hospitalization in patients with COVID-19. Endocrine. 2020;68(3):475–8.