Critical Care Explorations

OPEN

Using the Shapes of Clinical Data Trajectories to Predict Mortality in ICUs

Junchao Ma, MSc[1]; Donald K. K. Lee , PhD[2]; Michael E. Perkins, MD[3]; Margaret A. Pisani, MD[4]; Edieal Pinker, PhD[1]

Objectives: 1) To show how to exploit the information contained in the trajectories of time-varying patient clinical data for dynamic predictions of mortality in the ICU; and 2) to demonstrate the additional predictive value that can be achieved by incorporating this trajectory information.
Design: Observational, retrospective study of patient medical records for training and testing of statistical learning models using different sets of predictor variables.
Setting: Medical ICU at the Yale-New Haven Hospital.
Subjects: Electronic health records of 3,763 patients admitted to the medical ICU between January 2013 and January 2015.
Interventions: None.
Measurements and Main Results: Six-hour mortality predictions for ICU patients were generated and updated every 6 hours by applying the random forest classifier to patient time series data from the prior 24 hours. The time series were processed in different ways to create two main models: 1) manual extraction of the summary statistics used in the literature (min/max/median/first/last/number of measurements) and 2) automated extraction of trajectory features using machine learning. Out-of-sample area under the receiver operating characteristics curve and area under the precision-recall curve ("precision" refers to positive predictive value and "recall" to sensitivity) were used to evaluate the predictive performance of the two models. For 6-hour prediction and updating, the second model achieved area under the receiver operating characteristics curve and area under the precision-recall curve of 0.905 (95% CI, 0.900–0.910) and 0.381 (95% CI, 0.368–0.394), respectively, which are statistically significantly higher than those achieved by the first model, with area under the receiver operating characteristics curve and area under the precision-recall curve of 0.896 (95% CI, 0.892–0.900) and 0.905 (95% CI, 0.353–0.379). The superiority of the second model held true for 12-hour prediction/updating as well as for 24-hour prediction/updating.
Conclusions: We show that statistical learning techniques can be used to automatically extract all relevant shape features for use in predictive modeling. The approach requires no additional data and can potentially be used to improve any risk model that uses some form of trajectory information. In this single-center study, the shapes of the clinical data trajectories convey information about ICU mortality risk beyond what is already captured by the summary statistics currently used in the literature.
Key Words: hospital mortality; informatics; machine learning; prognosis; statistical models; time-dependent covariates

The ability to automate data extraction from electronic health record (EHR) systems opens the door for dynamic mortality warning indicators for ICU patients. This has spurred recent efforts to use machine learning to predict individual patient mortality risk (1–8), with an eye toward improving on existing illness severity scores like Acute Physiology and Chronic Health Evaluation (APACHE), Mortality Probability Model, and Simplified Acute Physiologic Score (9–12). These prognostic scores use static information taken at a fixed point in time (e.g., 24 hr after ICU admission) to identify patients at high risk of dying during an ICU visit. However, they ignore potentially valuable information conveyed by the changes in the measurements over time.

Recognizing this limitation, a number of studies use summary statistics of the trajectories of the EHR measurements as inputs for mortality prediction. Some do this by recalculating the scores at regular intervals using the most recent values of the measurements (13), whereas others (1, 14, 15) recalculate the Sequential Organ Failure Assessment (SOFA) score (16) using the worst values in the earlier 24 hours. Sometimes, unstructured text data

[1]School of Management, Yale University, New Haven, CT.
[2]Goizueta Business School, Emory University, Atlanta, GA.
[3]Hartford Hospital, Hartford, CT.
[4]Yale New Haven Hospital Pulmonary and Critical Care Medicine, New Haven, CT.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).

The authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: edieal.pinker@yale.edu

Crit Care Expl 2019; 1:e0010

DOI: 10.1097/CCE.0000000000000010

from patient clinical notes are also used as inputs, with features extracted from the text that have accumulated up to the current point in time (4, 7).

Another approach in the literature is to calculate a score at just one point in time, but using more trajectory inputs such as the minimum, maximum, and mean values in the prior 24 hours (5). Additional features like the first and last values, and also the number of measurements taken during the period, have also been used before (17). Indeed, there is a seemingly unlimited number of additional features that one can use to further describe the shape of the trajectories, such as curvature and arc length. Like text data, the trajectory curve is an unstructured object from which we wish to extract as much information as possible to use for prediction.

This article has two aims. The first is to show how a branch of machine learning called "functional data analysis" (18) can be used to incorporate the entire trajectory of a frequently measured variable into dynamic predictions of near-term mortality (updated regularly during ICU stay). Second, we demonstrate that there is predictive value in incorporating this extra trajectory information: For periodically updated predictions of mortality over a 6-hour window in the ICU, our approach provides statistically significant improvements to both the area under the receiver operating characteristic (AUROC) curve and area under the precision-recall curve (AUPRC) over using trajectory features manually selected from the "same" dataset. In other words, this increase in accuracy comes at no extra cost (aside from negligible computational ones), and the same result is seen for 12- and 24-hour prediction windows as well. The functional data analysis method automatically extracts features from the trajectory. These features capture information about all aspects of the curve and can be fed as inputs into any predictive mortality model. Thus, the method can potentially be used to improve any risk score that uses some form of trajectory information.

MATERIALS AND METHODS

Setting

Our study was a retrospective review of de-identified patient records approved by the Yale University Institutional Review Board using data on 4,557 unique patients admitted to the medical ICU (MICU) of the Yale-New Haven Hospital between January 2013 and January 2015. A total of 5,505 hospitalization episodes and 6,113 MICU visits were recorded. To replicate the trajectory inputs used in the literature (5, 17), which are defined over a trailing window of at least 24 hours, we removed MICU visits that lasted less than 24 hours. This is in line with both APACHE and SOFA, which wait for 24 hours after ICU admission before being calculated. We ended up with 3,763 unique patients who were admitted to the MICU, a total of 4,945 times (across 4,472 episodes of hospitalizations).

For each ICU visit, we generated (overlapping) observational units every 6 hours, where a unit is defined as a 24-hour window during the visit. For example, an ICU visit that lasted 39 hours generates the observational units (0, 24] hours, (6, 30] hours, and (12, 36] hours (**Supplemental Fig. 1**, Supplemental Digital Content 1, http://links.lww.com/CCX/A11). In total, we have 74,067 units of observations, and for each unit, the outcome of interest is death within 6 hours of the end of the interval. The observations are used to build a predictive model that updates mortality forecasts every 6 hours after the initial prediction at hour 24. We feel that the 6-hour prediction window trades off the tension between giving physicians enough advance warning to intervene on patients at imminent risk of dying, but not so much time that too many patients are flagged as at risk of dying within an overly wide time window, thus making it difficult for physicians to prioritize. We also provide similar results for 12- and 24-hour prediction windows.

Data Collection

In addition to patient demographic data, the EHR also provided records for six different types of physiologic measures that were sampled periodically during the ICU stay. In addition, records on 26 types of laboratory values and 18 types of prescribed medications were also included. A summary is provided in **Table 1**. We also made use of a metric that is relatively unique to the Yale-New Haven Hospital, the Rothman Index (RI). This is an EHR-based measure of patient acuity that is continuously updated throughout an episode of hospitalization. The RI score is a composite measure updated regularly from the electronic medical record based on changes in 26 clinical measures including vital signs, nursing assessments, Braden score, cardiac rhythms, and laboratory test results (shown in **Supplemental Table 1**, Supplemental Digital Content 2, http://links.lww.com/CCX/A12). This score is independent of diagnosis, and it was developed to be used for any inpatient (i.e., medical or surgical patients including critical care patients). With a theoretical range from –91 to 100, the majority of patients on a general medical or surgical unit fall within the range from 0 to 100, with lower scores indicating poorer condition (19). The RI has previously been shown to have predictive power in forecasting 24-hour in-ICU mortality (20). Eleven of the clinical measures in Table 1 overlap with the components of the RI. Although the exact calculation of the RI is proprietary, the main idea is to calculate a 1-year mortality risk function for each of the 26 features selected by stepwise logistic regression. These univariate functions were then combined together in an additive manner to produce the RI score.

Data Processing

Time Series Data. Among the physiologic measures in Table 1, those that were sampled often enough to provide a usable time series during an ICU stay (updated hourly) include systolic blood pressure, diastolic blood pressure, heart rate, and the RI. To automatically extract features from each time series that characterizes its evolution, we fitted a continuous trajectory to each series in each observational unit. The basis functions that were used to interpolate the time series data points were cubic B-splines with 23 evenly spaced interior knots, and they were fitted using penalized least squares (21). **Supplemental Figure 2** (Supplemental Digital Content 3, http://links.lww.com/CCX/A13) displays the 27 basis functions that are each weighted and summed together to form the spline fits, and **Supplemental Figure 3** (Supplemental Digital

TABLE 1. Raw Variables

Category	Variable Names
Rothman Index	Rothman Index[a]
Physiologic	Diastolic blood pressure[a,b], systolic blood pressure[a,b], temperature[b], Glasgow Coma Scale score, pulse (heart rate)[a,b], and total respiratory rate[b]
Laboratory values	Sodium[b], potassium[b], chloride[b], creatinine[b], glucose, glucose meter, calcium, magnesium, phosphorus, WBC count[b], hemoglobin[b], hematocrit, international normalized ratio, lactate, bilirubin total, bilirubin direct, alanine transaminase, aspartate transaminase, alkaline phosphatase, albumin, prealbumin, troponin T, fibrinogen level, pH arterial, Po_2 arterial, and Pco_2 arterial
Medication	Vasopressors/inotropes: dobutamine, dopamine, epinephrine, norepinephrine, vasopressin, and phenylephrine
	Antibiotics/antivirals/antifungals: acyclovir, ceftriaxone, ciprofloxacin, doxycycline, ertapenem, fluconazole, gentamicin, moxifloxacin, vancomycin, valacyclovir, ampicillin-sulbactam, and piperacillin-tazobactam
Demographic	Age, gender, race, height, and weight
Chronic disease	Dialysis, chronic obstructive pulmonary disease, and HIV

[a]Time series data.
[b]A component of Rothman Index.

Content 4, http://links.lww.com/CCX/A14) displays an example of one such fit to RI time series data. Because this approach is summarizing the time series using a collection of functions, the fitted coefficients capture information about the shape of the time series. We can then use these coefficients as features of the trajectory in our estimation model. To compare against the use of summary statistics that are manually selected from the trajectory in the literature, we also calculated the minimum, maximum, median, first value, last value, and the number of samples for each time series. We call these variables the "standard trajectory summaries."

Other Physiologic Measures. For the less frequently sampled physiologic measures, we used summary statistics for the measurements within each observational unit (maximum, minimum, and median) to represent their trajectories.

Laboratory Work. For infrequently performed laboratory tests, we used the latest measurement in the trailing 48 hours as the representative value for each observational unit. If a laboratory value was not available in the last 48 hours, it was treated as missing and was handled using the approach described below. The choice of 48 hours was taken from the protocol used in RI to handle laboratory values (19).

Medication Records. We consolidated patient medication records into two categories: a variable for the number of vasopressors and inotropes administered during the period, and an indicator variable that tracked whether a patient was on antibiotics, antifungals, or antivirals during the period. These variables serve as markers of shock and infection, respectively, and relate to disease processes.

Time in ICU. We also created a variable to record how long each patient had already spent in the ICU at the start of the observational unit.

Missing Values. For nontime series variables, missing values were handled in the following way. Since the range for the nonmissing values across all variables was substantially bounded away from –1,000, we encoded a missing value with this number to instruct our tree-based estimation algorithm to treat it differently. For time series data, an observational unit that had fewer than 12 measurements was considered to have a missing functional data point (the time series). Therefore, the coefficients for the 27-spline functions are all encoded as –1,000. **Supplemental Table 2** (Supplemental Digital Content 5, http://links.lww.com/CCX/A15) reports the amount and percentage of missing data for each variable of interest.

Model Construction

The prediction variables that were created above for each observational unit can be fed into any classifier to estimate the probability of death within the next 6 hours in the ICU (or indeed, any number of hours). We use the random forest classifier (22) as our platform for investigating the use of trajectory data because it is a popular nonparametric method that consistently ranks as one of the top performing prediction tools. Employing this to estimate the probability of death, we assess the performance gains resulting from using our proposed trajectory variables in lieu of the standard trajectory summaries defined earlier. We do this by comparing four nested random forest models M1–M4 (**Table 2** for details). In brief, M1 uses the most current values of all predictor variables aside from RI. M2 appends the additional statistics that make up the standard trajectory summaries used in the literature. M3 also adds the standard trajectory summaries for RI. Finally, M4 replaces all summaries with spline coefficients that capture the shape of the entire trajectory.

Model Evaluation

Subsequently (17), we used Monte Carlo cross validation to evaluate the performance of the four models: We performed 20 random splits of our set of unique patients into training (70%) and validation (30%) sets. Having 20 different splits of the data reduces the bias that could arise from any one particular split. For each split, we fit a random forest classifier to the observational units in the training set. The model was then used to predict the probability of death within the next 6 hours for each observational unit in the validation set.

Receiver operating characteristics (ROCs) plots are commonly used to assess the performance of binary classifiers. However, they

TABLE 2. Out of Sample Area Under the Receiver Operating Characteristic Curve and Area Under the Precision-Recall Curve for 6-Hour Mortality Prediction (95% CIs in Parentheses)

Model	Area Under the Receiver Operating Characteristic Curve	Area Under the Precision-Recall Curve
(M1) Snapshot of variables without RI	0.883 (0.877–0.889)	0.342 (0.333–0.351)
(M2) Trajectory summaries without RI	0.887 (0.882–0.892)	0.350 (0.339–0.361)
(M3) Trajectory summaries with RI	0.896 (0.892–0.900)	0.366 (0.353–0.379)
(M4) Full trajectory	0.905 (0.900–0.910)	0.381 (0.368–0.394)

RI = Rothman Index.

can be misleading in situations where the outcome classes are highly imbalanced (23–25). Such is the case here since the number of observational units with a mortality event (1.2%) is much lower than the number of units without. For imbalanced outcomes, simulation studies (23–25) suggest that the precision-recall plot is more informative where "precision" refers to positive predictive value (PPV) and "recall" refers to sensitivity. In light of this, we calculated both the AUROC curve and the AUPRC averaged over the 20 splits for each model, along with 95% CIs. We ran paired-sample *t* tests to compare average AUROC and average AUPRC between different models.

We also applied the Hosmer-Lemeshow test to each of the 20 test sets to evaluate the calibration of the predictive model. Conceptually, if a well-calibrated model assigns (say) *z*% chance of a death event to each of 100 observational units, then about *z* out of the 100 should result in an actual death. The Hosmer-Lemeshow test measures the discrepancy between the expected and observed death rates for the observational units, with the null hypothesis being that the two quantities agree. In performing the test, we followed the guidelines in (26) for analyzing large datasets. All calculations were performed in R (R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org).

In addition to making mortality predictions over a 6-hour window, we also repeated the abovementioned analysis for 12- and 24-hour windows. For 12-hour windows, predictions are made every 12 hours based on the trajectory of the measurements over the previous 24 hours. For 24-hour windows, predictions were made every 24 hours.

RESULTS

The mean age of patients was 63 years (SD, 17 yr), and 53% were male patients with mean weight of 179 lb (SD, 57 lb) and mean height of 5.5 ft. (SD, 0.4 ft). About 6.8% of patients had multiple ICU admissions, and 53% of ICU admissions occurred within 24 hours of being admitted to the hospital. After removing ICU stays shorter than 24 hours, the average length of ICU stay in our final dataset was 3.9 days. Among all observational units, 1.2% were followed by death within 6 hours (1.9% for 12 hr, 3.4% for 24 hr, and 17% were followed by death during the same ICU stay). **Supplemental Table 3** (Supplemental Digital Content 6, http://links.lww.com/CCX/A16) and **Supplemental Table 4** (Supplemental Digital Content 7, http://links.lww.com/CCX/A17) provide the distributions for the number of ICU visits per episode of hospitalization, and also the number of observational units per ICU admission. **Supplemental Table 5** (Supplemental Digital Content 8, http://links.lww.com/CCX/A18) shows how many ICU patients died in each 6-hour time interval and the corresponding hazard rate.

Table 2 compares the AUROC and AUPRC for 6-hour mortality averaged over the 20 splits of the data for models M1–M4 described in **Table 3**. The table encapsulates three findings, which hold for the 12- and 24-hour prediction windows as well (**Supplemental Table 6**, Supplemental Digital Content 9, http://links.lww.com/CCX/A19). **Supplemental Table 7** (Supplemental Digital Content 11, http://links.lww.com/CCX/A21) shows *p* values for paired-sample *t* tests between different models.

First, the AUROC and AUPRC for M4 (0.905 and 0.381, respectively) are both higher than those for M3 (0.896 and 0.366), and the differences are statistically significant ($p = 0.004$ and 0.017, respectively). In other words, the spline representation of the trajectories of time series data conveys additional predictive information that are not already captured by the standard trajectory summaries. Furthermore, using the Gini measure (27), **Figure 1** shows that more than half of the 20 most predictive variables are the spline coefficients for RI and pulse, particularly the ones describing the most recent evolution of the times series (e.g., "pulse_27" is the coefficient for the last spline function in the 24-hour period in Supplemental Figure 2 [Supplemental Digital Content 3, http://links.lww.com/CCX/A13]). This reflects the time decay in the predictive power of the time series data.

Second, the performance of M3 is in turn statistically significantly better than M2 with AUROC of 0.896 versus 0.887 ($p < 0.001$) and AUPRC of 0.366 versus 0.350 ($p = 0.002$). That is, the RI conveys additional predictive information over the other variables used in M2, including 11 of the 26 components used to calculate RI. This is reinforced by Figure 1, which shows that eight of the 20 most predictive variables are related to the trajectory of the RI.

Third, the difference in performance between M1 and M2 is small (0.883 vs 0.887 for AUROC and 0.342 vs 0.35 for AUPRC) and not statistically significant ($p = 0.363$ for AUROC and $p = 0.236$ for AUPRC). In other words, beyond the most recent measurement, the additional trajectory summaries employed in the literature do not add meaningful predictive power.

To make the performance measures more concrete, it is helpful to consider a single point on the precision-recall curve. **Figure 2**

TABLE 3. Variables Used in the Nested Predictive Models

Model	Variables Used
(M1) Snapshot of variables without RI	Most recent values for time series data except RI (diastolic blood pressure, systolic blood pressure, and heart rate)
	All other variables except RI
(M2) Trajectory summaries without RI	Add standard trajectory summaries of time series data (except RI) to model M1
(M3) Trajectory summaries with RI	Add standard trajectory summaries for RI to model M2
(M4) Full trajectory	Replace standard trajectory summaries in model M3 with the spline coefficients that capture complete trajectories

RI = Rothman Index.

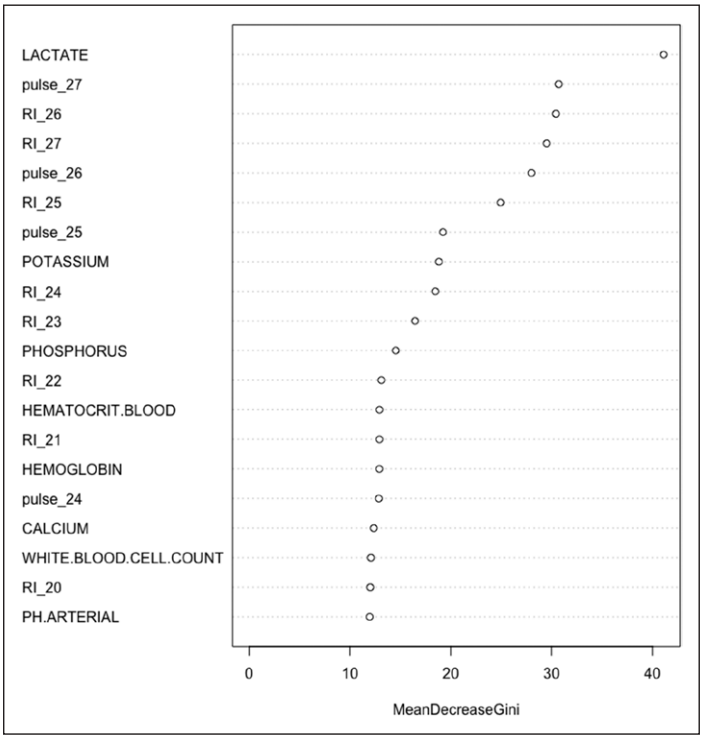

Figure 1. Top twenty most important variables for model M4. Rows are model variables and horizontal axis reports the mean decrease in Gini caused by each variable. RI = Rothman Index.

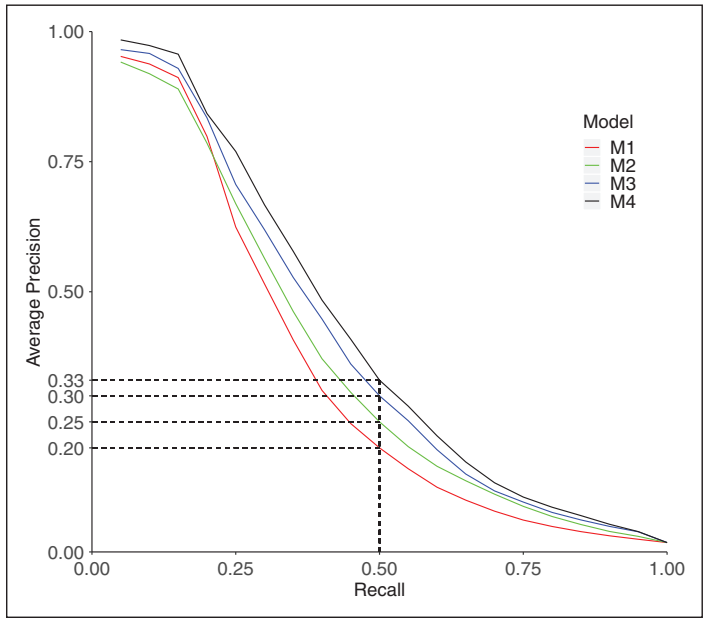

Figure 2. Average precision (%) versus recall (%) plot of M1, M2, M3, and M4 for 6-hr mortality prediction.

and **Figure 3** display the precision-recall curves and ROCs averaged over the 20 splits of the data for models M1–M4. At 50% recall (sensitivity), the average PPV for 6-hour mortality prediction is 33% (compared with 21% for M1, 25% for M2, and 30% for M3). In other words, if the observational units flagged by our algorithm are to include half of those that resulted in death within 6 hours, then one-third of all flagged cases will be correct. This seems like a low accuracy but compared with low rate of mortality in 6-hour windows it is informative. To elaborate, because 1.2% of the observational units were followed by death within 6 hours, this means that we can identify half of all potential deaths by focusing on just the top 1.2% × 50%/33% = 1.8% of observational units with the highest predicted probabilities of death. **Supplemental Figure 4** (Supplemental Digital Content 10, http://links.lww.com/CCX/A20) shows the survival time distributions for the true positive predictions and for the false negatives produced by model M4.

Finally, the results of the Hosmer-Lemeshow test showed that our final model M4 is well calibrated: For the 20 test sets, the *p* values for only three of them were less than 0.20 with the smallest one being 0.094 (**Supplemental Table 8**, Supplemental Digital Content 12, http://links.lww.com/CCX/A22). Thus, the null hypothesis was never rejected.

DISCUSSION

Most modern attempts to predict patient mortality acknowledge the value of incorporating information about changes in patient measurements over time. They use as predictors a number of manually created summaries of the trajectory paths traced out by the measurements. Interestingly, our results suggest that these summary statistics add little predictive power over just using the most recent measurement value.

To incorporate all information about the trajectory into a model, we show how functional data analysis can be employed to automatically extract predictor variables that capture the complete evolution of the trajectory. We demonstrate that using these features improves the accuracy of mortality predictions over 6-,

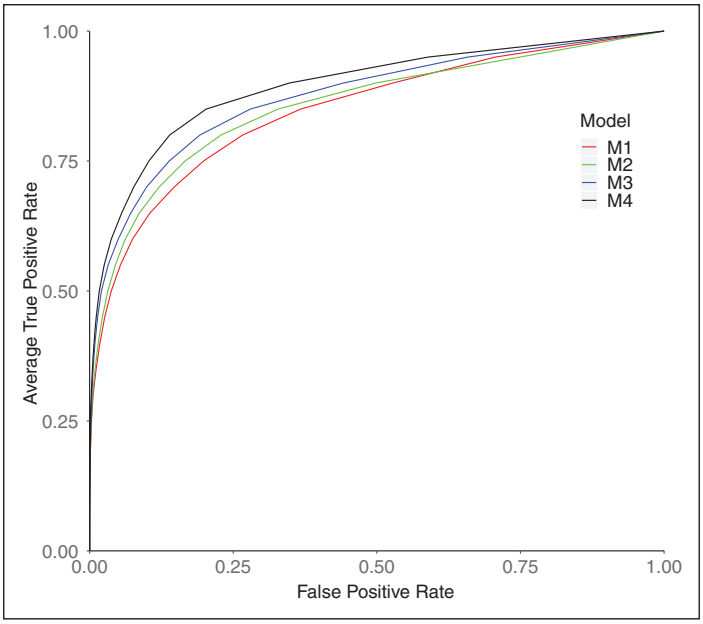

Figure 3. Average receiver operating characteristics of M1, M2, M3, and M4 for 6-hr mortality prediction.

12-, and 24-hour horizons when compared with using the standard trajectory summaries from the literature. Although there is an infinite number of different time horizons one can consider, we suspect that the same qualitative finding will hold for similar-sized time intervals.

Although the magnitude of the improvements is modest, it is important to bear in mind that the proposed trajectory features can be easily generated at negligible computational cost using popular open-source packages like R. Therefore, to not use them is to discard free information. Moreover, they can be substituted for the standard trajectory summaries in any model that uses some form of trajectory information. Thus, our work complements the burgeoning research on predictive mortality modeling in the ICU.

Several limitations of this study are worth noting. Our results are based on comparing the performances of using different sets of variables as input into the random forest algorithm. Should another machine learning algorithm be used, there is no guarantee that the proposed trajectory features will still outperform the use of standard trajectory summaries. However, because random forest is a nonparametric prediction method that consistently ranks as one of the top performing prediction tools, we suspect that our qualitative findings will carry over to other top-of-the-line prediction methods. Evidently, further work is required to confirm this.

To replicate the trajectory inputs used in the literature (5, 17), which require 24–48 hours of measurements, we removed MICU visits that lasted less than 24 hours. Since 17% of deaths in our dataset occurred within the first 24 hours of ICU admission, this restricts the scope of our findings to ICU stays that last at least 24 hours.

We did not have access to data on nursing assessments or code status. This was compensated to some extent by the inclusion of RI, which takes into account nursing assessments and cardiac rhythms. Although our results show use of the RI can be useful for predicting mortality, it is a proprietary product available only to select hospitals that subscribe to the service. Hence, we are also showing that scoring systems like the RI can be improved upon by using trajectory information. Last, our study is based on data from a single medical center, and as noted in Supplemental Table 2 (Supplemental Digital Content 5, http://links.lww.com/CCX/A15) for some variables we have a lot of missing values. Therefore, more work need to be done to validate our approach.

CONCLUSIONS

We have shown that using trajectory information of clinical data can improve the accuracy of mortality predictions. This benefit is apparent for both an approach that uses trajectory summary statistics and an approach that uses an algorithmically generated functional representation of the trajectory. Mortality over short-time horizons is a very rare event even for acutely ill patients, which makes it difficult to predict. Our approach indicates how making fuller use of the available clinical data can help address this difficulty. One would expect that this approach would also benefit predictions of other outcomes for which trends in patient's health performance could be useful indicators such as readmission or response to specific treatments.

Any short-time horizon mortality prediction method has the potential to be the basis for a clinical warning system. Nursing units have acuity-based nurse to patient ratios standards because nurses need to be able to give sufficient attention to patients. A mortality prediction method provides a risk assessment for each patient that can help direct limited nursing resources to the patients most in need. To truly assess if a warning system is useful in this setting it is necessary to determine if the warning would identify high mortality risk cases that: 1) are not already known to the nurses and physicians and 2) could be aided by an intervention. Such an assessment would require a different kind of study.

REFERENCES

1. Badawi O, Liu X, Hassan E, et al: Evaluation of ICU risk models adapted for use as continuous markers of severity of illness throughout the ICU stay. *Crit Care Med* 2018; 46:361–367
2. Johnson AE, Ghassemi MM, Nemati S, et al: Machine learning and decision support in critical care. *Proc IEEE Inst Electr Electron Eng* 2016; 104:444–466
3. Pirracchio R, Petersen ML, Carone M, et al: Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *Lancet Respir Med* 2015; 3:42–52
4. Rajkomar A, Oren E, Chen K, et al: Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018; 18
5. Citi L, Barbieri R: PhysioNet 2012 Challenge: Predicting mortality of ICU patients using a cascaded SVM-GLM paradigm. *Comput Cardiol* 2012; 39:257–260
6. Aczon M, Ledbetter D, Ho L, et al: Dynamic mortality risk predictions in pediatric critical care using recurrent neural networks. arXiv:1701.06675
7. Ghassemi M, Naumann T, Doshi-Velez F, et al: Unfolding physiological state: Mortality modelling in intensive care units. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, August 24-27, 2014, pp 75–84
8. Johnson AE, Kramer AA, Clifford GD: A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy. *Crit Care Med* 2013; 41:1711–1718

9. Knaus WA, Zimmerman JE, Wagner DP, et al: APACHE-Acute Physiology and Chronic Health Evaluation: A physiologically based classification system. *Crit Care Med* 1981; 9:591–597
10. Lemeshow S, Teres D, Klar J, et al: Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993; 270:2478–2486
11. 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; 270:2957–2963
12. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297–1310
13. Sow DM, Sun J, Biem A, et al: Real-time analysis for short-term prognosis in intensive care. *IBM J Res Dev* 2012; 56:1–10
14. Ferreira FL, Bota DP, Bross A, et al: Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 2001; 286:1754–1758
15. Holder AL, Overton E, Lyu P, et al: Serial daily organ failure assessment beyond ICU day 5 does not independently add precision to ICU risk-of-death prediction. *Crit Care Med* 2012; 45:2014–2022
16. Vincent JL, de Mendonça A, Cantraine F, et al: Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med* 1998; 26:1793–1800
17. Johnson AEW, Dunkley N, Mayaud L, et al: Patient specific predictions in the intensive care unit using a Bayesian ensemble. *Comput Cardiol* 2012; 39:249–252
18. Ramsay JO, Silverman BW: Functional Data Analysis. Second Edition. New York, NY, Springer, 2005
19. Rothman MJ, Rothman SI, Beals J 4th: Development and validation of a continuous measure of patient condition using the electronic medical record. *J Biomed Inform* 2013; 46:837–848
20. Finlay GD, Rothman MJ, Smith RA: Measuring the modified early warning score and the Rothman Index: Advantages of utilizing the electronic medical record in an early warning system. *J Hosp Med* 2014; 9:116–119
21. De Boor C: A Practical Guide to Splines. New York, NY, Springer-Verlag, 1978
22. Liaw A, Wiener M: Classification and regression by random forest. *R News* 2002; 2:18–22
23. Leisman DE: Rare events in the ICU: An emerging challenge in classification and prediction. *Crit Care Med* 2018; 46:418–424
24. 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; 10:1–21
25. Ozenne B, Subtil F, Maucort-Boulch D: The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol* 2015; 68:855–859
26. Paul P, Pennell ML, Lemeshow S: Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Stat Med* 2013; 32:67–80
27. Menze BH, Kelm BM, Masuch R, et al: A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 2009; 10:213