IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 16, 2021, accepted July 13, 2021, date of publication July 26, 2021, date of current version August 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3099996

Improving Recurrent Neural Network Responsiveness to Acute Clinical Events

DAVID R. LEDBETTER, EUGENE LAKSANA, MELISSA ACZON, AND RANDALL WETZEL
The Laura P. and Leland K. Whittier Virtual Pediatric Intensive Care Unit, Children's Hospital Los Angeles, Los Angeles, CA 90027, USA
Corresponding author: David R. Ledbetter (dledbetter@chla.usc.edu)

This work was supported by the L. K. Whittier Foundation.

ABSTRACT Predictive models in acute care settings must immediately recognize precipitous changes in a patient's status when presented with data reflecting such changes. Recurrent neural networks (RNN) have become popular for clinical decision support models but exhibit a delayed response to acute events. New information must propagate through the RNN's cell state before the total impact is reflected in the model's predictions. Input data perseveration is a method to train more responsive RNN-based models. Input data is replicated k times during training and deployment. Each replication propagates through the cell state and output of the RNN, but only the output at the final replication is maintained and broadcast as the prediction for evaluation. De-identified Electronic Medical Records (EMR) of 12, 826 patients admitted to a tertiary care pediatric academic center between 01/2009-02/2019 were analyzed. A baseline Long Short-Term Memory (LSTM) model ($k = 1$), four LSTMs with increasing amounts of input data perseveration ($k = 2$ to $k = 5$), and an LSTM with an attention mechanism were trained to predict ICU-mortality. Performance of models was compared using Area Under the Receiver Operating Characteristic Curve (AUROC) after increasing periods of observation from one to 12 hours. The average variation of the change in predicted mortality immediately following defined acute events measured responsiveness. The AUROC gains due to input perseveration were larger at the earlier times of prediction ($\leq$6 hours), increasing at the first hour from 0.77 with no input data perseveration to 0.83 when $k = 5$. An LSTM with $k = 5$ was $2 - 3$ times more responsive to acute events than a baseline LSTM.

INDEX TERMS Electronic medical records, health and safety, Kalman filters, long short-term memory, machine learning, performance evaluation, predictive models, real time systems, recurrent neural networks, time series analysis.

I. INTRODUCTION

A. THE PROBLEM

Critical care environments require rapid decision making. To be meaningful, predictive models in these settings must immediately recognize precipitous changes in a patient's state when presented with data reflecting such changes [1], [2]. Newly acquired patient data must be integrated quickly with previous data and the model's predictions rapidly updated to inform clinical decisions in a timely fashion.

Long Short-Term Memory (LSTM) models [3] are a type of Recurrent Neural Network (RNN) architecture and have become increasingly popular for modeling tasks in clinical settings [4]–[10]. In particular, they previously have been shown to achieve significantly higher predictive performance than clinical severity of illness (SOI) scores based on assessing a child's risk of mortality in an intensive care unit (ICU) [11]. Critical care requires constant re-evaluations of a patient's status, and the desire for a tool that automatically and continuously evaluates a patient's status in the ICU is underscored by studies investigating the continuous application of static clinical scores [12]–[14]. The RNN is suited for providing automatic and continuous patient monitoring and assessment because of its ability to make ongoing concurrent predictions with new data integrated into the patient's historic context, instead of merely recalculating a new prediction at a single isolated time. The RNN requires computational cycles to incorporate new data into its memory cell [15], which results in a behavior that is recognized but not well described in the literature: when integrating new data, there is a delay between the input and a change in the prediction [16]. This characteristic is frequently observed in stock prediction

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

tasks where the apparent predictive capability of a model is comparable to a model that predicts the last observed stock price [17], [18]. The intersection of deep learning and healthcare, though ever widening, has a paucity of literature on this phenomenon. Commonly reported metrics such as area under the receiver operating characteristic curve, sensitivity, specificity, and mean absolute errors reflect only the accuracy of the final predictions but do not account for this time lag.

In critical care applications, we have noticed two distinct but related observable effects of the time lag. The first occurs during what we refer to as a *pipe-up* period, during which the initial input propagates through the memory cell to overcome the initialized cell state before that input significantly influences the predictions. The second occurs during subsequent times when new information only slightly changes the overall prediction of a model at the time of input, again requiring propagation through the memory cell before fully augmenting the model's prediction. The first phenomenon is a special case of the second. These two effects are further illustrated in Sections II-D and III.

B. PROPOSED SOLUTION

This work aims to enable more timely changes in RNN predictions when a model is presented with data indicating acute clinical events. Because the RNN requires computational cycles to incorporate new data into its memory cell, and because ICU EMR data is usually presented hourly, we reasoned that repeating computational cycles at the time new data becomes available, that is *perseverating* the data presentation to the network, would decrease the response time of the RNN predictions. Instead of giving newly available input data to an RNN model only once, the input is replicated, or *perseverated*, and given to the model multiple times, during both training and deployment, with only the prediction of the final perseveration maintained and made visible to the end-user. This technique of *perseverating* an input k times provides additional computational cycles for new data to be incorporated into the model's internal states before a prediction is broadcast. We hypothesized that the resultant model would react to acute events more quickly than traditionally trained LSTMs, reduce the *pipe-up* time to overcome the initialized cell state, and still maintain overall model performance.

C. RELATED WORKS

Since the seminal paper by R.E. Kalman in 1960 describing what is now called the Kalman Filter [19], engineers have been aware of the tradeoff between integrating historical data and responding to new data. Most training techniques used in modern deep learning affect the balance between relying on historical trends and responding to new information; such techniques include dropout, optimizers, and activation functions [20]–[23]. Generating an appropriate target vector (e.g. using changes in values between consecutive time points instead of predicting raw values) is another training technique that can prevent generation of an auto-correlation model [16].

Attention mechanisms [24] are sometimes used to manage the balance between historical and new data by appending the entire input sequence to the final hidden layer of an RNN. Doing so affords the model another opportunity to learn model weights which expose the moments in time series data most relevant to the predictions. Attention networks were originally developed for post-hoc sequence-to-sequence modeling tasks such as image captioning and neural machine translation which permit access to the entire input sequence when making predictions [24], [25]. An attention mechanism was applied by Zhang, *et al.* to predict risk of future hospitalization using medical records from a fixed observation window [26].

While an attention mechanism allows the model to weight isolated time vectors that are important to the target prediction, *input perseveration* allows all information to propagate through both the input and the memory cell and enables the existing machinery of the RNN (namely *backpropogation*) to determine the magnitude of their contribution.

II. METHODS

A. CLINICAL DATA SOURCES

Data for this work were extracted from de-identified observational clinical data collected in Electronic Medical Records (EMR, Cerner) in the Pediatric Intensive Care Unit (PICU) of Children's Hospital Los Angeles (CHLA) between January 2009 and February 2019. The CHLA Institutional Review Board (IRB) reviewed the study protocol and waived the need for IRB approval. A patient record included static information such as gender, race, and discharge disposition at the end of an ICU *episode*, defined as a contiguous admission in the PICU. A patient may have multiple episodes. The EMR for an episode also contained irregularly, sparsely and asynchronously *charted* measurements of physiologic observations (e.g. heart rate, blood pressure), laboratory results (e.g. creatinine, glucose level), drugs (e.g. epinephrine, furosemide) and interventions (e.g. intubation, oxygen level). Episodes without discharge disposition were excluded, leaving 12,826 episodes (9,250 patients) in the final dataset.

Prior to any of the computational experiments described here, the episodes were randomly partitioned into three sets: a training set for deriving model weights, a validation set for optimizing hyper-parameters, and a holdout test set for measuring performance. To minimize bias in the performance evaluation metrics, partitioning was done at the patient level, i.e. all episodes from a single patient belonged to only one of these sets: 60% in the training set, 20% in the validation set, and 20% in the test set. No other stratification was applied. Table 1 displays basic characteristics of the resulting data partitions.

Preprocessing steps described in previous work [27] converted each episode data to a matrix structure amenable to machine learning model development. A row of values in this matrix represented measurements (recorded or imputed) of

TABLE 1. Basic demographics of data used in this study for each data partition.

	training	validation	test	overall
Number of episodes	7663	2541	2622	12826
Number of patients	5551	1850	1850	9251
Mortality rate	0.035	0.037	0.043	0.037
Fraction male	0.56	0.55	0.58	0.56
Age mean (years)	7.99	8.23	8.09	8.06
std dev	6.45	6.51	6.35	6.44

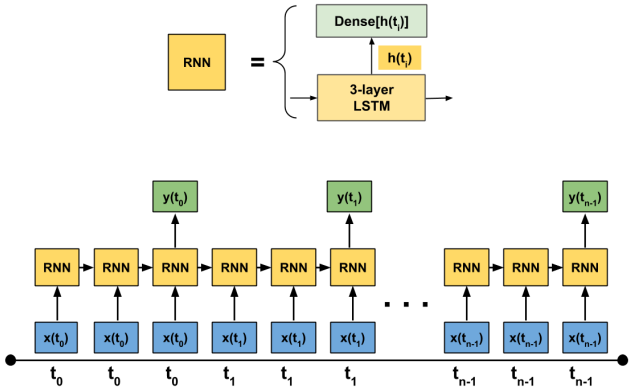

FIGURE 1. The perseverating recurrent neural network (PRNN) is an RNN with repeating inputs. Each input vector, $x(t_i)$, associated with a particular time, t_i, is replicated k times, and only the output from the last replication, $y(t_i)$, is considered as the prediction for that time. The diagram above illustrates the process for $k = 3$.

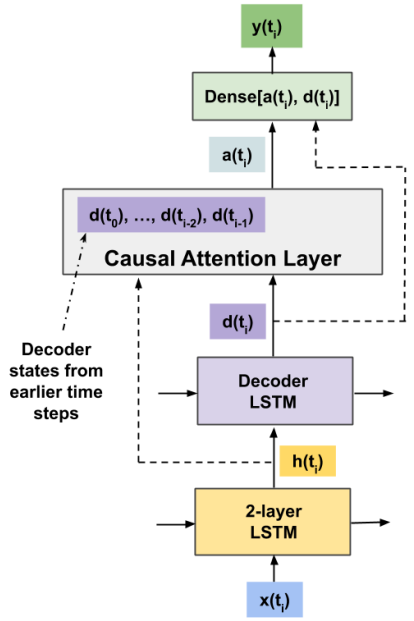

FIGURE 2. Overview of a standard RNN model with a causal attention layer.

different variables at one particular time, while a column contained values of a single variable at different times. A summary of the preprocessing may be found in Appendix A. A complete list of variables used as model inputs can be found in Appendix B. Note that diagnoses, while available in the EMR, were not used as input features.

B. TARGET OUTCOME

ICU mortality was chosen as the target outcome because it is a simple, unambiguous outcome. Importantly, risk of mortality (ROM) is used commonly as a proxy for severity of illness in critical care [13], [28]–[30]. The overall mortality rate of the data was 3.7% (Table 1).

C. RNN MODELS

Many-to-many recurrent neural network models, consisting of stacked Long Short-Term Memory (LSTM) layers followed by a dense layer, were trained to predict ROM in the ICU of each patient episode. All the models output a probability of mortality at each distinct time where an observation or measurement of the patient was made, generating a trajectory of scores that reflect the changing status of a patient during their ICU episode.

The *baseline* RNN model was trained in the traditional manner: when the model acquired new data $x(t_i)$, it generated and immediately broadcast a prediction $y(t_i)$. Other models which share the architecture of the baseline RNN model were trained using input *perseveration*: the same $x(t_i)$ vector was repeatedly given as input to the RNN model k times, where k is a controllable parameter. While all k outputs – corresponding to the k times that $x(t_i)$ was given to the model – were used for optimization during training, only the last one was maintained and broadcast as the prediction at time t_i during performance assessment and deployment. Figure 1 illustrates a *Perseverated Recurrent Neural Network* (PRNN) model with $k = 3$. A python implementation of a perseveration layer written with PyTorch 1.7.1 can be found in Appendix D. Note that the baseline RNN can be considered as a PRNN with $k = 1$. Input perseveration provides the internal cell state memory of the RNN additional computational cycles to incorporate the current state of the patient into the final prediction.

An attention network using the same hyperparameters of the baseline RNN model was also implemented. Most attention networks have access to the entire sequence of inputs – both past and future relative to the current one – when making a prediction at any point in time. However, continuous monitoring of patient status precludes access to future information, available in a retrospective study but not in a real deployment scenario. Therefore, the attention layer between the last hidden layer and output layer used a *causal mask* that exposed only the inputs up to the time when a prediction is being made [31]. This ensured that the network did not use future information to make its predictions while affording it another opportunity to consider the totality of information up to the current time. See Figure 2 for a diagram of the model.

All six models – baseline RNN ($k = 1$), PRNN for $k = 2, 3, 4, 5$, and attention RNN – were implemented and

TABLE 2. Hyperparameters for all permutations of the PRNN.

Hyperparameter	Value
Number of LSTM Layers	3
Hidden Units in LSTM Layers	[128, 256, 128]
Batch Size	32
Initial Learning Rate	1e-5
Loss	binary cross-entropy
Optimizer	rmsprop
Dropout	0.2
Recurrent Dropout	0.2
Regularizer	0.0001
Output Activation	sigmoid

trained using Keras 2.0.7 with the Theano 1.0.2 backend [32], [33]. For each model, weights were optimized on the training and validation sets; performance was computed on the validation set after every epoch (i.e. a full cycle through the training set), and the best performing weights were saved as the final model. Table 2 displays the hyperparameters that were used for all six models. The final models were assessed for performance metrics on the test set.

D. MODEL ASSESSMENTS

Standard metrics such as the Area Under the Receiver Operating Characteristic Curve (AUROC), precision, and recall scores for a binary classification task can capture a model's overall predictive performance. These metrics at a single time point do not measure the lag phenomenon displayed by LSTMs that, although known, is rarely commented on in the literature. The trajectory of ROM predictions should reflect the evolving state of a patient which can have instantaneous changes. Metrics were designed to quantify a model's pipe-up behavior and its ability to capture rapid changes resulting from clinically adverse events. AUROC was used to compare overall predictive performance for the main task – predicting ICU mortality.

1) PIPE-UP BEHAVIOR

A model's prediction at the first time point is a function of its model weights, initial state of its memory cells and the first input. Of these, only the first input varies across different patient episodes; therefore, the distribution of a model's prediction at the first time point of all episodes indicates the model's level of reliance on the first input data. We refer to this period as the *pipe-up* period.

The mean and standard deviation of the distribution of all $\hat{y}_p(t_0)$ predictions, where $\hat{y}_p(t_0)$ represents the first prediction for patient episode p, were computed for all survivors and non-survivors in the test set. These metrics were computed for all six RNN models and used to compare their responsiveness to the first available information about a patient episode.

Additionally, to assess the temporal increase in responsiveness, the AUROCs for the $k = 1$ and $k = 5$ models described in the previous section were computed for increasing periods of observation from one to 12 hours.

2) RESPONSIVENESS TO ACUTE EVENTS

The models were also assessed for their initial responses to clinically adverse events using an average variation metric computed from the predictions during such periods. In consultation with clinicians, the times when any of the following occurred were defined as acute time points or intervals:

1) a patient's heart rate fell to 0;
2) there was a *substantial* decrease in heart rate, mean arterial pressure, Glasgow Coma Score, blood oxygen level (SpO2), arterial blood gas (ABG) pH, and venous blood gas (VBG) pH between two consecutive recordings;
3) there was a *substantial* increase in creatinine levels or inotrope score between two consecutive recordings.

We quantified *substantial* changes as those that were in the top or bottom X percentile ($X \in \{5.0, 1.0, 0.5\}$) of inter-measurement changes. For example, a time point t_i was considered acute in terms of creatinine when the *increase* in creatinine level from t_{i-1} to t_i was in the top X percentile of all creatinine level percent changes between any two consecutive time points in the test dataset. Similarly, acute time points for heart rate were those times when the *decrease* in heart rate was in the bottom $X\%$ of all heart rate changes between any two consecutive time points of all test set episodes. Table 3 shows the specific percent change thresholds indicating *acute* changes in a patient. Thus, a 32% drop in heart rate between two consecutive measurements would be in the top percentile, while a 70% increase in inotrope score would be in the top fifth percentile.

Applying three percentile-based thresholds to define acute patient state changes meant a total of 25 acuity definitions (2 * 3 increases in measured variables, 6 * 3 decreases in measured variables, and zero heart rate) used to assess models' responses when such events occur. It is important to note that these are not proposed as general definitions of clinical acuity. They were designed to capture events that indicate precipitous changes in a patient's state with high specificity.

Each defined acute event, denoted by E, identified a set of *acute time points*, denoted by $S_E(p)$, for each patient episode p. Changes in a model's mortality prediction at these time points were evaluated through an average temporal variation metric given by:

$$V_E(p) = \frac{100}{|S_E(p)|} \sum_{t_i \in S_E(p)} |\hat{y}_p(t_i) - \hat{y}_p(t_{i-1})|, \qquad (1)$$

where $\hat{y}_p(t_i)$ is the prediction for patient episode p at time point t_i, and $|S_E(p)|$ is the number of time points in $S_E(p)$. This metric measures how much the predicted ROM (scaled to [0, 100]) changed, on average, at the defined acute time points of an episode. Model A having a higher $V_E(p)$ than model B means that when both models were presented with data reflecting a precipitous change in a patient's state, model A's prediction underwent a greater change than model B's prediction, indicating that model A had a more pronounced response to the acute event. The average of this metric across

TABLE 3. Thresholds of percent change used to define acute changes in patient state. Negative thresholds correspond to the bottom p% of changes and positive thresholds correspond to the top p% of changes, for $p \in [5, 1, 0.5]$.

	5%	1%	0.5%
Heart Rate	-18.8	-31.6	-37.1
Mean Arterial Pressure	-23.5	-38.1	-44.2
Glascow Coma Score	-50.0	-72.7	-72.7
Pulse Oximetry	-5.6	-20.2	-35.1
ABG pH	-1.5	-2.9	-3.5
VBG pH	-1.4	-2.6	-3.1
Inotrope Score	66.7	150.0	250.0
Creatinine	33.3	60.5	81.8

N_p episodes gives a measure of the model's overall responsiveness to acute event E of those episodes:

$$\bar{V}_E = \frac{1}{N_p} \sum_p V_E(p). \tag{2}$$

3) OVERALL PREDICTIVE PERFORMANCE

Model predictions at ICU admission and at 1, 3, 6, 12, and 24 hours after ICU admission were evaluated for predictive performance via the AUROC. Performance was only measured for episodes lasting at least 24 hours to maintain a consistent cohort across time slices ($n = 2130$, mortality rate $= 0.046$).

III. RESULTS

Figure 3 illustrates trajectories of ROM predictions from two models, the baseline RNN ($k = 1$) and a PRNN with $k = 5$, for two non-surviving episodes. In the first episode (top), the patient had a rapid desaturation event (Sp02 dropping from 100% to 33%) 6 hours after ICU admission. When presented with this defined acute event, the baseline RNN's mortality prediction increased from 0.031 to 0.065, while the PRNN ($k = 5$) model's mortality prediction increased from 0.035 to 0.265. In the second episode (bottom), the patient's Glasgow Coma Score decreased from 7 to 4. When presented with this defined acute event, the baseline RNN's ROM prediction increased from 0.24 to 0.30, while the PRNN ($k = 5$) model's ROM prediction increased from 0.31 to 0.50. In either case, the PRNN model's immediate response to an acute event was more pronounced than that of the baseline RNN model. Further, the trajectory of the baseline RNN's predictions after either acute event appears to lag behind that of the PRNN by about 1-2 hours.

The remainder of this section describes the results from aggregating the assessment metrics described in Section II-D across episodes in the test dataset.

A. PIPE-UP BEHAVIOR

Figure 4 shows the distribution of ROM predictions at the first observation. The mean mortality prediction for survivors decreased with the perseveration parameter k, going from 0.32 ($k = 1$) to 0.10 ($k = 5$). For both survivors and non-survivors, the standard deviation of the ROM predictions

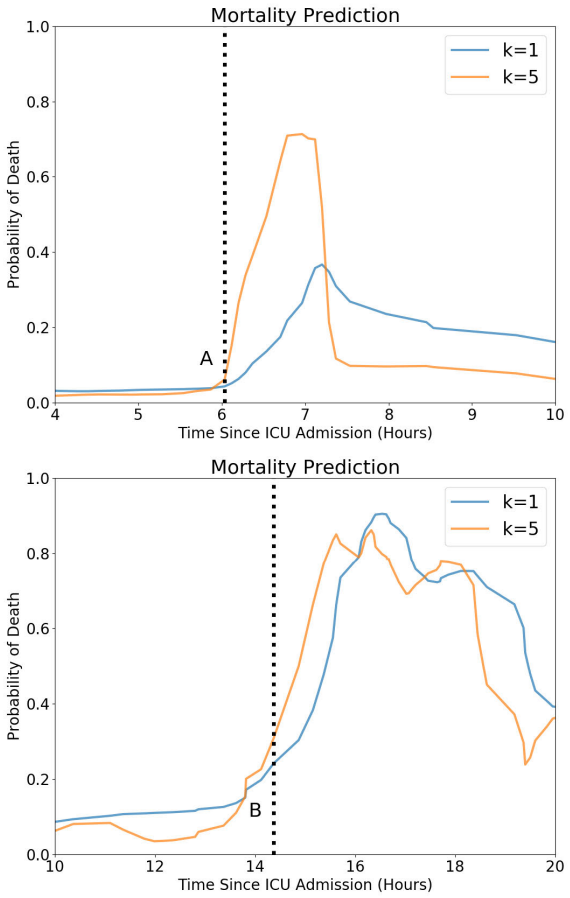

FIGURE 3. Two examples of predictions following acute events in two individual patient episodes. The blue curves correspond to the baseline ($k = 1$) model, while the orange curves correspond to the PRNN model with $k = 5$. The dotted vertical lines denote onset of an acute event for each patient: (A) a patient whose oxygen saturation decreased from 100% to 33%; (B) a patient's Glasgow Coma Score decreasing from 7 to 4, indicating significantly reduced levels of consciousness.

increased with k. The increase in standard deviation was greater in the non-survivors, going from 0.03 ($k = 1$) to 0.13 ($k = 5$).

Figure 5 shows the AUROCs of the $k = 1$ and $k = 5$ models at the first 12 hours of each episode. The $k = 1$ curve is shifted to the right of the $k = 5$ curve by approximately one hour.

B. RESPONSIVENESS TO ACUTE EVENTS

Figure 6 displays the average variation, $\bar{V}_E$, as a function of k and acuity definitions. Two trends are apparent. First is the behavior of $\bar{V}_E$ as a function of percentile change in a given physiologic observation or intervention. For each model, the resulting average variation increased as the change in a physiologic or intervention variable became more severe (i.e. from 95.0 to 99.5 percentile). Generally, the average variation increased monotonically as k increased across all definitions

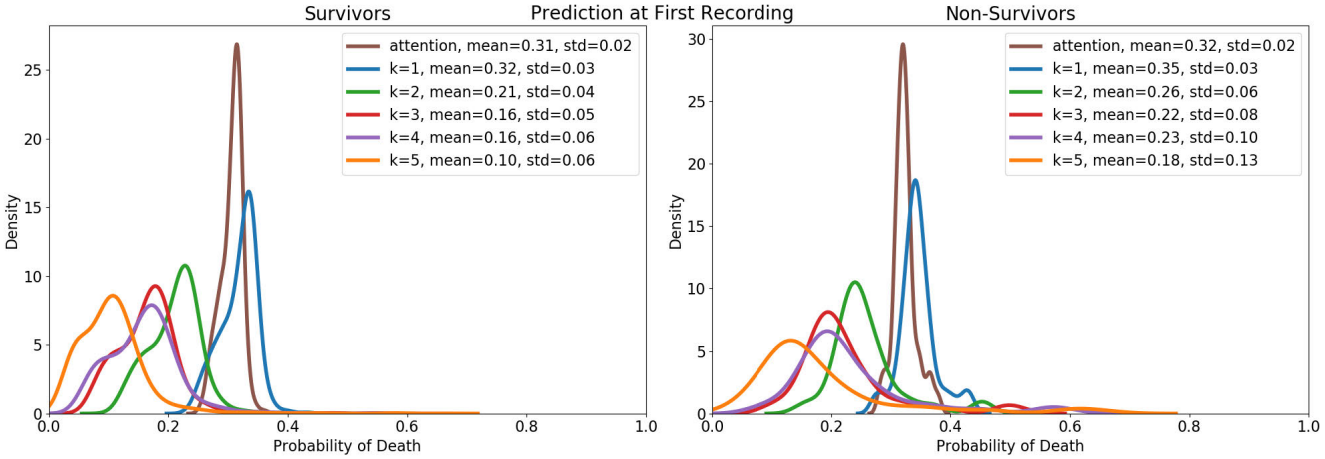

FIGURE 4. Distribution of model predictions at the first available observation for survivors and non-survivors.

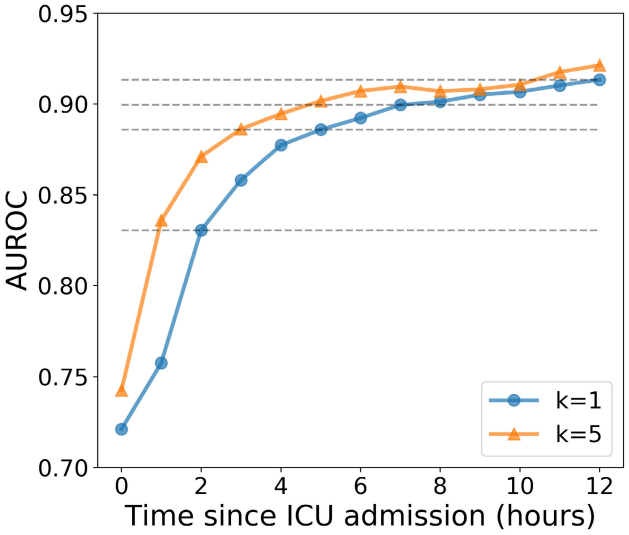

FIGURE 5. Performance for two models ($k = 1$ and $k = 5$) over the first 12 hours of an ICU episode.

of acuity. The attention layer generally demonstrated lower average variation than the $k = 1$ model.

C. OVERALL PREDICTIVE PERFORMANCE

Table 4 summarizes the AUROC of model predictions at different times after ICU admission of all episodes in the test set that lasted at least 24 hours. The AUROC gains due to input perseveration were larger at the earlier times of prediction ($t \leq 6$ hours), increasing at the first hour from 0.77 when k=1 to 0.83 when $k = 5$. By the 12th hour, the AUROC did not significantly change with the perseveration parameter k. For all models, performance increased with greater observation time and reduced lead time [34]. Adding an attention layer had similar performance as the baseline model ($k = 1$) at all evaluation times. The complete ROC curves for each time slice can be found in Appendix C.

TABLE 4. Test set AUROCs for the mortality prediction at admission and after 1, 3, 6, 12, and 24 hours of observation. Only the patient episodes lasting at least 24 hours of ICU time were evaluated.

k	0hr	1hr	3hr	6hr	12hr	24hr
1	0.713	0.751	0.854	0.889	0.911	0.952
2	0.735	0.804	0.871	0.902	0.916	**0.956**
3	0.738	0.819	0.875	0.902	0.916	0.955
4	0.741	0.821	0.876	0.900	0.915	0.954
5	**0.742**	**0.832**	**0.883**	**0.905**	**0.919**	0.955
attention	0.718	0.734	0.853	0.892	0.910	0.949

IV. DISCUSSION

Real time predictions for ICU mortality are proxies for severity of illness [13], [28], [30], [35] and should reflect condition or status in response to a change in observations of the patient. The examples in Figure 3 show that the mortality model trained with input perseveration (PRNN with $k = 5$) responded more pronouncedly and immediately than the traditionally trained model ($k = 1$) when both were presented with data reflecting acute changes. Subsequent to the acute changes, the responses of the traditional RNN appeared to lag behind the PRNN's. Standard metrics such as AUROCs for classification or mean absolute errors for continuous regression at a single time point do not elucidate this predictive lag and other deleterious behaviors that can be detrimental in critical or intensive care settings where rapid recognition and response are crucial.

Input perseveration provides the RNN's internal cells additional computational cycles to incorporate the current state of the patient into the final prediction. The effect of this method was demonstrated by metrics designed to capture a model's initial responses to newly acquired data.

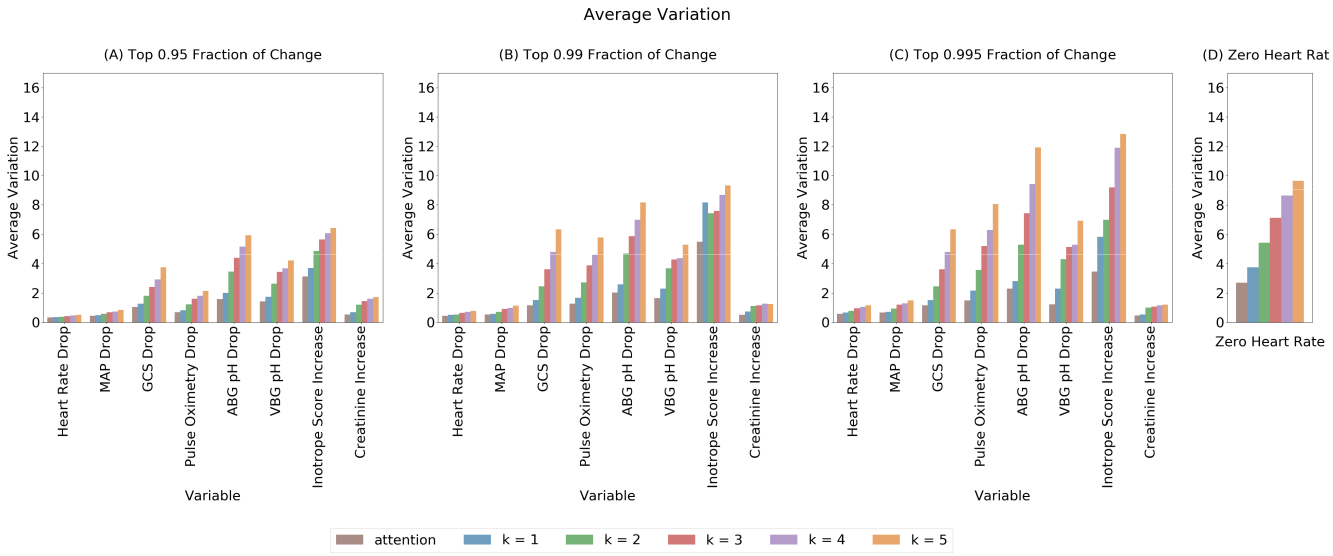

FIGURE 6. Comparison of the average variation metric $\bar{V}_E$ of different models as defined by Equation 2 across all test set episodes, where the events are defined by: A) a 95th percentile change; B) a 99th percentile change; C) a 99.5th percentile change; D) a cardiac arrest.

The variation metric described in Section II-D2 compares the models' initial responses to data indicating precipitous changes in a patient's status. These changes may require quick reaction time from the care team, therefore rapid responsiveness of the predictions is important. The comparison of variation metrics from the different models (Figure 6) shows that the LSTM became more responsive to acute clinical events as the level of perseveration, k, increased. When k increased from 1 to 5, the variation metric corresponding to the defined precipitous events increased by a factor of 2-3 times. This means that the predicted mortality risk after an acute event was 2-3 times more pronounced when perseveration was set to $k = 5$ compared to a baseline LSTM ($k = 1$) or an attention network. The results also show that a given model's responsiveness increased with more acute events (i.e. those belonging to higher percentile changes for a given physiologic or intervention variable). These are consistent with expectations about the variation metric. For example, one would expect the increases in ROM predictions to be greater for large drops in blood gasses than for less severe drops.

Figure 4 compares the distributions of initial predictions from the different models. Since the first prediction is a function of both the initial input (which varies across episodes) and the initial cell state memory (which is fixed), a wider distribution of these predictions across episodes indicates a higher reliance on the initial input. This is important because children admitted to the PICU have different severities of illness [28], [30]. Increasing k resulted in a wider distribution of predictions as measured by the standard deviation. The increase was greater for the non-surviving population ($\sigma = 0.03$ when $k = 1$ to $\sigma = 0.13$ when $k = 5$) relative to the surviving population. This is again consistent with clinical expectations that the non-surviving population would have a broader range of initial SOI scores (especially in the tail) than the surviving population.

Increasing the first prediction's reliance on the initial measurement – as achieved by the PRNN models with higher k – also resulted in higher AUROC at ICU admission (Table 4). As the models observed the patient longer, their AUROCs increased as expected. Importantly, the increase of AUROC due to perseveration was greater at the early hours ($t \leq 3$) when information is most scarce, with the 1-hour AUROC increasing from 0.77 when $k = 1$ to 0.83 when $k = 5$. This means that LSTM-based predictive clinical models relying on scarce data for early detection could benefit from the perseveration approach.

Figure 5 and Table 4 demonstrate that a PRNN model ($k = 5$) more quickly overcame the initialization state of the network during the *pipe-up* period than the baseline model ($k = 1$). A clear translation of approximately 1 hour indicates the ability of the PRNN model to more quickly integrate the patient observations and provide more responsive predictions to a clinical team.

Adding a causal attention layer to the baseline ($k = 1$) model had no apparent performance improvement in any of the metrics. The attention network theoretically can put more weight to the most recent state than to the previous ones. The results indicate that this mechanism did not improve on what the baseline LSTM's gates were already doing, but persistently giving the same input to the model – i.e. perseverating the input – did.

There are limitations to the perseverated input approach. Although LSTMs are theoretically good at understanding temporal trends through their memory cells, there remain practical limitations to how long prior information can be maintained to inform future predictions [36], [37]. The PRNN has the potential to exacerbate these algorithmic deficiencies

TABLE 5. EMR variables (demographics, vitals and labs) in patient episode matrix. The superscript $^+$ indicates the variable is binary, while * indicates the variable is ordinal. Variables without any superscripts are real-valued.

Demographics and Vitals			
Age	Sex_F$^+$	Sex_M$^+$	race_African American$^+$
race_Asian/Indian/Pacific Islander$^+$	race_Caucasian/European Non-Hispanic$^+$	race_Hispanic	race_unknown$^+$
Abdominal Girth	FLACC Pain Face*	Left Pupillary Response Level*	Respiratory Effort Level*
Activity Level*	FLACC Pain Intensity*	Level of Consciousness*	Respiratory Rate
Bladder pressure	FLACC Pain Legs*	Lip Moisture Level*	Right Pupil Size After Light*
Capillary Refill Rate*	Foley Catheter Volume	Mean Arterial Pressure	Right Pupil Size Before Light*
Central Venous Pressure	Gastrostomy Tube Volume	Motor Response Level*	Right Pupillary Response Level*
Cerebral Perfusion Pressure	Glascow Coma Score*	Nasal Flaring Level*	Sedation Scale Level*
Diastolic Blood Pressure	Head Circumference	Near-Infrared Spectroscopy SO2	Skin Turgor_edema$^+$
EtCO2	Heart Rate	Nutrition Level*	Skin Turgor_turgor$^+$
Extremity Temperature Level*	Height	Oxygenation Index	Systolic Blood Pressure
Eye Response Level*	Hemofiltration Fluid Output	PaO2 to FiO2	Temperature
FLACC Pain Activity*	Intracranial Pressure	Patient Mood Level*	Verbal Response Level*
FLACC Pain Consolability*	Left Pupil Size After Light*	Pulse Oximetry	WAT1 Total*
FLACC Pain Cry*	Left Pupil Size Before Light*	Quality of Pain Level*	Weight

Labs			
ABG Base excess	CBG PCO2	GGT	Neutrophils %
ABG FiO2	CBG PO2	Glucose	PT
ABG HCO3	CBG TCO2	Haptoglobin	PTT
ABG O2 sat	CBG pH	Hematocrit	Phosphorus level
ABG PCO2	CSF Bands%	Hemoglobin	Platelet Count
ABG PO2	CSF Glucose	INR	Potassium
ABG TCO2	CSF Lymphs %	Influenza Lab$^+$	Protein Total
ABG pH	CSF Protein	Lactate	RBC Blood
ALT	CSF RBC	Lactate Dehydrogenase Blood	RDW
AST	CSF Segs %	Lactic Acid Blood	Reticulocyte Count
Albumin Level	CSF WBC	Lipase	Schistocytes
Alkaline phosphatase	Calcium Ionized	Lymphocyte %	Sodium
Amylase	Calcium Total	MCH	Spherocytes
Anti-Xa Heparin	Chloride	MCHC	T4 Free
B-type Natriuretic Peptide	Complement C3 Serum	MCV	TSH
BUN	Complement C4 Serum	MVBG Base Excess	Triglycerides
Bands %	Creatinine	MVBG FiO2	VBG Base excess
Basophils %	Culture Blood$^+$	MVBG HCO3	VBG FiO2
Bicarbonate Serum	Culture CSF$^+$	MVBG O2 Sat	VBG HCO3
Bilirubin Conjugated	Culture Fungus Blood$^+$	MVBG PCO2	VBG O2 sat
Bilirubin Total	Culture Respiratory$^+$	MVBG PO2	VBG PCO2
Bilirubin Unconjugated	Culture Urine$^+$	MVBG TCO2	VBG PO2
Blasts %	Culture Wound$^+$	MVBG pH	VBG TCO2
C-Reactive Protein	D-dimer	Macrocytes	VBG pH
CBG Base excess	ESR	Magnesium Level	White Blood Cell Count
CBG FiO2	Eosinophils %	Metamyelocytes %	
CBG HCO3	Ferritin Level	Monocytes %	
CBG O2 sat	Fibrinogen	Myelocytes %	

because of the memory cell's prolonged exposure to the same data. A comprehensive assessment of the impact of a potential reduction in temporal memory was not performed. Nevertheless, the basic data perseveration technique can easily be generalized to any sequential data.

Further, perseveration increases the number of sequences requiring computation. Perseveration is used during both training and inference, and the compute time scales linearly for both, commensurate with the level of perseveration. Training time for $k = 1$ required 6 hours and 8 minutes, while $k = 5$ required 41 hours and 14 minutes. Nevertheless, the time required for inference is on the order of $30ms$ for $k = 1$ (baseline) and $150ms$ for $k = 5$ on an NVIDIA Titan RTX. Since the recording frequency is approximately every 15 minutes, these computational burdens do not hinder deployment.

Finally, this proof-of-concept study used only a single clinical outcome (ICU mortality) and data from a single center. Future work will examine the effect of perseverating the data input on other important clinical tasks such as risk of desaturation, sepsis, and renal failure.

V. CONCLUSION

This work demonstrates that perseverated data input increases the responsiveness of LSTM models, in both timeliness and magnitude, to a variety of acute changes to patient state. After acute events, the magnitude of prediction changes was 2-3 times greater when the input was perseverated. Pipe-up performance showed approximately an hour improvement in the ability to understand a patient. These enable quicker evaluation of patients which is critical in an acute clinical setting.

TABLE 6. EMR variables (drugs and interventions) in baseline patient episode matrix. The superscript + indicates the variable is binary, while * indicates the variable is ordinal. Variables without any superscripts are real-valued.

Drugs			
Acetaminophen/Codeine_inter	Clonazepam_inter	Ipratropium Bromide_inter	Oseltamivir_inter
Acetaminophen/Hydrocodone_inter	Clonidine HCl_inter	Isoniazid_inter	Oxacillin_inter
Acetaminophen_inter	Cyclophosphamide_inter	Isradipine_inter	Oxcarbazepine_inter
Acetazolamide_inter	Desmopressin_inter	Ketamine_cont	Oxycodone_inter
Acyclovir_inter	Dexamethasone_inter	Ketamine_inter	Pantoprazole_inter
Albumin_inter	Dexmedetomidine_cont	Ketorolac_inter	Penicillin G Sodium_inter
Albuterol_inter	Diazepam_inter	Labetalol_inter	Pentobarbital_inter
Allopurinol_inter	Digoxin_inter	Lactobacillus_inter	Phenobarbital_inter
Alteplase_inter	Diphenhydramine HCl_inter	Lansoprazole_inter	Phenytoin_inter
Amikacin_inter	Dobutamine_cont	Levalbuterol_inter	Piperacillin/Tazobactam_inter
Aminophylline_cont	Dopamine_cont	Levetiracetam_inter	Potassium Chloride_inter
Aminophylline_inter	Dornase Alfa_inter	Levocarnitine_inter	Potassium Phosphat e_inter
Amlodipine_inter	Enalapril_inter	Levofloxacin_inter	Prednisolone_inter
Amoxicillin/clavulanic acid_inter	Enoxaparin_inter	Levothyroxine Sodium_inter	Prednisone_inter
Amoxicillin_inter	Epinephrine_cont	Lidocaine_inter	Propofol_cont
Amphotericin B Lipid Complex_inter	Epinephrine_inter	Linezolid_inter	Propofol _inter
Ampicillin/Sulbactam_inter	Epoetin_inter	Lisinopril_inter	Propranolol HCl_inter
Ampicillin_inter	Erythromycin_inter	Lorazepam_inter	Racemic Epi_inter
Aspirin_inter	Factor VII_inter	Magnesium Sulfate_inter	Ranitidine_inter
Atropine_inter	Famotidine_inter	Meropenem_inter	Rifampin_inter
Azathioprine_inter	Fentanyl_cont	Methadone_inter	Risperidone_inter
Azithromycin_inter	Fentanyl_inter	Methylprednisolone_inter	Rocuronium_inter
Baclofen_inter	Ferrous Sulfate_inter	Metoclopramide_inter	Sildenafil_inter
Basiliximab_inter	Filgrastim_inter	Metronidazole_inter	Sodium Bicarbonate_inter
Budesonide_inter	Fluconazole_inter	Micafungin_inter	Sodium Chloride_inter
Bumetanide_inter	Fluticasone_inter	Midazolam HCl_cont	Sodium Phosphate_inter
Calcium Chloride_cont	Fosphenytoin_inter	Midazolam HCl_inter	Spironolactone_inter
Calcium Chloride_inter	Furosemide_cont	Milrinone_cont	Sucralfate_inter
Calcium Gluconate_inter	Furosemide_inter	Montelukast Sodium_inter	Tacrolimus_inter
Carbamazepine_inter	Gabapentin_inter	Morphine_cont	Terbutaline_cont
Cefazolin_inter	Ganciclovir Sodium_inter	Morphine_inter	Tobramycin_inter
Cefepime_inter	Gentamicin_inter	Mycophenolate Mofetl_inter	Topiramate_inter
Cefotaxime_inter	Glycopyrrolate_inter	Naloxone HCL_cont	Trimethoprim/Sulfamethoxazole_inter
Cefoxitin_inter	Heparin_cont	Naloxone HCL_inter	Ursodiol_inter
Ceftazidime_inter	Heparin_inter	Nifedipine_inter	Valganciclovir_inter
Ceftriaxone_inter	Hydrocortisone_inter	Nitrofurantoin_inter	Valproic Acid_inter
Cephalexin_inter	Hydromorphone_cont	Nitroprusside_cont	Vancomycin_inter
Chloral Hydrate_inter	Hydromorphone_inter	Norepinephrine_cont	Vasopressin_cont
Chlorothiazide_inter	Ibuprofen_inter	Nystatin_inter	Vecuronium_inter
Ciprofloxacin HCL_inter	Immune Globulin_inter	Octreotide Acetate_cont	Vitamin K_inter
Cisatracurium_cont	Insulin_cont	Olanzapine_inter	Voriconazole_inter
Clindamycin_inter	Insulin_inter	Ondansetron_inter	

Interventions			
Abdominal X Ray+	Diversional Activity_tv+	NIV Mode*	Range of Motion Assistance Type*
Arterial Line Site+	ECMO Hours	NIV Set Rate	Sedation Intervention Level*
CT Abdomen Pelvis+	EPAP	Nitric Oxide	Sedation Response Level*
CT Brain+	FiO2	Nurse Activity Level Completed+	Tidal Volume Delivered
CT Chest+	Gastrostomy Tube Location+	O2 Flow Rate	Tidal Volume Expiratory
Central Venous Line Site+	HFOV Amplitude	Oxygen Mode Level*	Tidal Volume Inspiratory
Chest Tube Site+	HFOV Frequency	Oxygen Therapy*	Tidal Volume Set
Chest X Ray	Hemofiltration Therapy Mode+	PEEP	Tracheostomy Tube Size
Comfort Response Level*	IPAP	Peak Inspiratory Pressure	Ventilator Rate
Continuous EEG Present+	Inspiratory Time	Peritoneal Dialysis Type+	Ventriculostomy Site+
Diversional Activity_books+	MRI Brain+	Pharmacological Comfort Measures Given	Visitor Mood Level*
Diversional Activity_music+	Mean Airway Pressure	Position Support Given+	Visitor Present+
Diversional Activity_play+	Mechanical Ventilation Mode+	Position Tolerance Level*	Volume Tidal
Diversional Activity_toys+	MultiDisciplinaryTeam Present+	Pressure Support	

APPENDIX A
DATA PREPROCESSING OVERVIEW

Disparate measurements of a common underlying physiologic measure were grouped into a single variable when deemed medically appropriate by an pediatric intensivist. For instance, systolic blood pressures (SBP) measured either invasively via a catheter or non-invasively via a cuff were grouped into a single variable representing SBP. Drugs or interventions which were administered in less than one percent of episodes in the training set were excluded. Measurements incompatible with human life were removed from the dataset, such as heart rates exceeding 400 beats per minute. Means and standard deviations derived from the training set were used to transform physiologic variables and laboratory measurements to have zero mean and unit variance. Therapies applied to the patient were scaled to be between 0 and 1 using clinically defined upper limits (with zero indicating absence of therapy). No age normalization was performed on the data

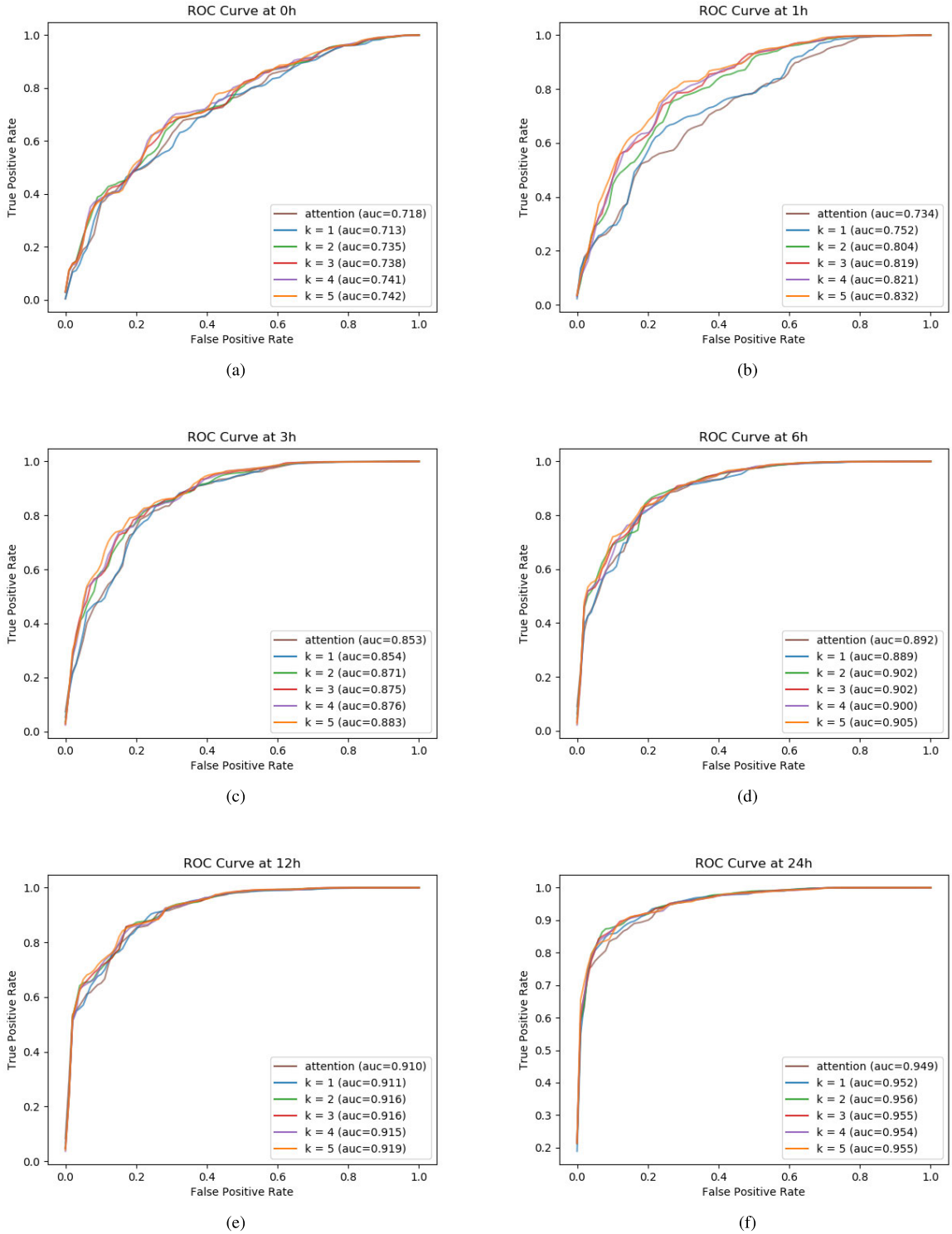

FIGURE 7. **Complete ROC curves for (a) $t = 0$, (b) $t = 1$, (c) $t = 3$, (d) $t = 6$, (e) $t = 12$, and (f) $t = 24$.**

as age was used as an input feature. Diagnoses were not used as model inputs.

EMR measurements were sparsely, asynchronously, and irregularly charted. The median time between measurements was 22 minutes with a minimum interval of one minute. Missing measurements for any drugs or interventions were simply set to zero to indicate absence of treatment. If at any time there was at least one recorded measurement for

```python
import torch

class Perseverate(torch.nn.Module):
    ''' Data augmentation layer that repeats sequences k times
    '''

    def __init__(self, k):
        ''' Initialize the transform layer
            k: integer number of perseverations to augment
        '''
        self.k = k
        super(Perseverate, self).__init__()

    def forward(self, xs, ys, **args):
        ''' repeat input k times
            xs: input x torch.tensor
            ys: output y torch.tensor
        '''
        xs = xs.repeat_interleave(self.k, axis=0)
        ys = ys.repeat_interleave(self.k, axis=0)
        return xs, ys

    def invert_output(self, ys, y0s):
        ''' strip all perseverated elements except the final output
            ys: y truth torch.tensor
            y0s: y pred torch.tensor
        '''
        ys = ys[self.k-1::self.k]
        y0s = y0s[self.k-1::self.k]
        return ys, y0s
```

Listing 1. Code snippet for data perseveration augmentation layer in python.

a physiologic observation or laboratory measurement it was forward filled until another measurement was recorded. This reflects the charting practices of the nurses within the ICU. If there was no recorded physiologic or laboratory value for an entire episode a mean value generated from the training set population was used.

It should be noted that many variables are symbolic representations of symptoms or signs observed within the ICU (such as Glasgow Coma Score or FLACC Pain scores) but were processed in the same manner as direct observations such as heart rate or temperature.

APPENDIX B
EMR VARIABLES

See Table 5 and 6.

APPENDIX C
RECEIVER OPERATING CHARACTERISTIC CURVES FOR EACH TIME SLICE

See Figure 7.

APPENDIX D
PERSEVERATION CODE

See Listing 1.

ACKNOWLEDGMENT

The authors would like to thank Alysia Flynn for help aggregating the data.

REFERENCES

[1] P. S. Chan, A. Khalid, L. S. Longmore, R. A. Berg, M. Kosiborod, and J. A. Spertus, "Hospital-wide code rates and mortality before and after implementation of a rapid response team," *Jama*, vol. 300, no. 21, pp. 2506–2513, 2008.

[2] J. A. Daryl, D. A. Michael, and B. Rinaldo, "Rapid-response teams," *New England J. Med.*, vol. 365, no. 2, pp. 139–146, 2011.

[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] M. C. Winter, T. E. Day, D. R. Ledbetter, M. D. Aczon, C. J. L. Newth, R. C. Wetzel, and P. A. Ross, "Machine learning to predict cardiac death within 1 hour after terminal extubation," *Pediatric Crit. Care Med.*, vol. 22, no. 2, pp. 161–171, Feb. 2021.

[5] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Med.*, vol. 25, no. 1, pp. 24–29, 2019.

[6] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Comput. Biol. Med.*, vol. 102, no. 1, pp. 278–287, Nov. 2018.

[7] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, Jul. 2018.

[8] E. Laksana, M. Aczon, L. Ho, C. Carlin, D. Ledbetter, and R. Wetzel, "The impact of extraneous features on the performance of recurrent neural network models in clinical tasks," *J. Biomed. Informat.*, vol. 102, Feb. 2020, Art. no. 103351.

[9] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, 2017.
[10] C. S. Carlin, L. V. Ho, D. R. Ledbetter, M. D. Aczon, and R. C. Wetzel, "Predicting individual physiologically acceptable states at discharge from a pediatric intensive care unit," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 12, pp. 1600–1607, Dec. 2018.
[11] M. D. Aczon, D. R. Ledbetter, E. Laksana, L. V. Ho, and R. C. Wetzel, "Continuous prediction of mortality in the PICU: A recurrent neural network model in a single-center dataset," *Pediatric Crit. Care Med.*, vol. 22, no. 6, p. 519, 2021.
[12] H. CalebW and S. Peter, "ICU acuity: Real-time models versus daily models," in *Proc. AMIA Annu. Symp. Proc.*, 2009, p. 260.
[13] S. Leteurtre, A. Duhamel, V. Deken, J. Lacroix, and F. Leclerc, "Daily estimation of the severity of organ dysfunctions in critically ill children by using the PELOD-2 score," *Crit. Care*, vol. 19, no. 1, p. 324, Dec. 2015.
[14] O. Badawi, X. Liu, E. Hassan, P. J. Amelung, and S. Swami, "Evaluation of ICU risk models adapted for use as continuous markers of severity of illness throughout the ICU Stay," *Crit. Care Med.*, vol. 46, no. 3, pp. 361–367, Mar. 2018.
[15] C. Olah. (Aug. 27, 2015). Understanding LSTM networks. Colah's Blog. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/
[16] Y. Zhu and Y. Xiong, "Towards data science," *Data Sci. J.*, vol. 14, p. 8, May 2015.
[17] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, pp. 25–37, Aug. 2018.
[18] R. Xiong, E. P. Nichols, and Y. Shen, "Deep learning stock volatility with Google domestic trends," 2015, *arXiv:1512.04916*. [Online]. Available: http://arxiv.org/abs/1512.04916
[19] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, D, J. Basic Eng.*, vol. 82, pp. 35–45, Oct. 1960.
[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[21] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[23] N. Vinod and H. GeoffreyE, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[24] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*. [Online]. Available: http://arxiv.org/abs/1508.04025
[25] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, Feb. 2015, pp. 2048–2057.
[26] J. Zhang, K. Kowsari, J. H. Harrison, J. M. Lobo, and L. E. Barnes, "Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record," *IEEE Access*, vol. 6, pp. 65333–65346, 2018.
[27] H. LongV, L. David, A. Melissa, and W. Randall, "The dependence of machine learning on electronic medical record quality," in *Proc. AMIA Annu. Symp. Proc.*, 2017, p. 883.
[28] M. M. Pollack, K. M. Patel, and U. E. Ruttimann, "PRISM III: An updated pediatric risk of mortality score," *Crit. Care Med.*, vol. 24, no. 5, pp. 743–752, May 1996.
[29] M. M. Pollack, R. Holubkov, T. Funai, J. M. Dean, J. T. Berger, D. L. Wessel, K. Meert, R. A. Berg, and C. J. L. Newth, "The pediatric risk of mortality score: Update 2015," *Pediatric Crit. Care Med.*, vol. 17, no. 1, pp. 2–9, 2016.
[30] A. Slater, f. the PIM Study Group, F. Shann, and G. Pearson, "PIM2: A revised version of the paediatric index of mortality," *Intensive Care Med.*, vol. 29, no. 2, pp. 278–285, Feb. 2003.
[31] V. Ashish, S. Noam, P. Niki, U. Jakob, J. Llion, and G. AidanN, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[32] C. François, "Keras (2.2.2)," *GitHub Repository*, 2018.
[33] T. Theano Development Team *et al.*, "Theano: A Python framework for fast computation of mathematical expressions," 2016, *arXiv:1605.02688*. [Online]. Available: http://arxiv.org/abs/1605.02688
[34] D. E. Leisman, M. O. Harhay, D. J. Lederer, M. Abramson, A. A. Adjei, J. Bakker, and Z. K. Ballas, "Development and reporting of prediction models: Guidance for authors from editors of respiratory, sleep, and critical care journals," *Crit. Care Med.*, vol. 48, no. 5, p. 623, 2020.
[35] P. MurrayM, "Severity of illness confusion," *Pediatric Crit. Care Med.*, vol. 17, no. 6, p. 583, 2016.
[36] Y. Zhang, G. Chen, D. Yu, K. Yao, S. Khudanpur, and J. Glass, "Highway long short-term memory RNNS for distant speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 5755–5759.
[37] Y. Rose, Z. Stephan, A. Anima, and Y. Yisong, "Long-term forecasting using higher order tensor RNNs," 2017, *arXiv:1711.00073*. [Online]. Available: https://arxiv.org/abs/1711.00073

DAVID R. LEDBETTER received the B.Sc. degree in mathematics. He is the Principal Data Scientist at Children's Hospital Los Angeles and the Manager of the Virtual Pediatric Intensive Care Unit (VPICU) Data Science Team. With an extensive experience in signal processing and deep learning, he currently works closely with clinicians to bridge the gap between healthcare professionals and data scientists. He fosters communication and enables collaborations to solve real clinical problems that require significant cross-disciplinary expertise. He has a passion for training aspiring data scientists who are looking to break into healthcare to make a meaningful impact.

EUGENE LAKSANA received the B.Sc. degree in computer science from University of Southern California. He is currently a Data Scientist at Children's Hospital Los Angeles. His research interests include in bridging the gap between clinical experts and data scientists by cultivating meaningful interdisciplinary collaborations.

MELISSA ACZON received the B.Sc. degree in mathematics from Harvey Mudd College and the Ph.D. degree in scientific computing and computational mathematics from Stanford University. She is currently a Principal Data Scientist at Children's Hospital Los Angeles. She works closely with CHLA's ICU physicians to define clinically meaningful problems and develop algorithms to solve them using advanced computational techniques. Prior to joining CHLA, she was a Principal Scientist at Arete Associates.

RANDALL WETZEL received the undergraduate and medical degrees in U.K and the MBA degree, with a focus on information technology, from Hopkins. He is a tenured Professor of pediatrics and anesthesiology at the University of Southern California and has served as an attending Physician in pediatric critical care for over 40 years both at Johns Hopkins Hospital and at CHLA. He then trained in pediatrics at Case Western's Rainbow Babies and Children's Hospital, critical care medicine at Johns Hopkins Hospital, and subsequently anesthesiology in Hopkins. He is board certified in pediatrics, pediatric critical care, anesthesiology, and pediatric anesthesiology. Recently, his research has involved data science applied to critical care medicine. He has served as the Director of critical care medicine and as the Chairman of anesthesiology critical care medicine at CHLA. He has trained hundreds of residents and over 100 critical care fellows. He founded the Virtual Pediatric Intensive Care Unit focused on information technology and data science, which has supported research for over 20 years and received federal, state, and private funding. He also founded Virtual Pediatric Systems, LLC, which collects information of all admissions to over 160 pediatric ICUs in U.S.