***Improving Prediction Models’ Propriety in Intensive-Care Unit, by Enforcing an Advance Notice Period***

*Tomer Hermelin1, Pierre Singer2 & Nadav Rappoport1*

*1Department of Software and Information Systems Engineering, Faculty of Engineering, Ben-Gurion University of the Negev, ISRAEL*

*2Beilinson, ISRAEL*

*hermelit@post.bgu.ac.il*

[*psinger@clalit.org.il*](mailto:psinger@clalit.org.il)

*nadavrap@bgu.ac.il*

*Key phrases:*

Forecasting (D005544); Intensive Care Units (D007362); Hospital Mortality (D017052); Sepsis (D018805); Deep Learning (D000077321)

*Word Count: 3,923*

Table of Contents

[Abstract 3](#_Toc68961960)

[Objective 3](#_Toc68961961)

[Materials and Methods 3](#_Toc68961962)

[Results 3](#_Toc68961963)

[Conclusions 3](#_Toc68961964)

[Introduction 4](#_Toc68961965)

[Background 4](#_Toc68961966)

[Objective 6](#_Toc68961967)

[Materials and methods 6](#_Toc68961968)

[Formal problem definition 7](#_Toc68961969)

[Evaluation 8](#_Toc68961970)

[Clinical data 9](#_Toc68961971)

[Cohort development and feature selection 9](#_Toc68961972)

[Data preprocessing and feature engineering 10](#_Toc68961973)

[Model development 10](#_Toc68961974)

[Results 12](#_Toc68961975)

[Discussion 12](#_Toc68961976)

[Conclusions 13](#_Toc68961977)

[References 14](#_Toc68961978)

# Abstract

## Objective

Intensive-Care-Units (ICUs) are time-critical, and sufficient reaction time is crucial. There are existing papers and systems for alerting life-threatening events in the ICU, though these models suffer from “immediate” events bias. In this study, we present a new approach for outcome prediction in ICU admissions, which takes into consideration the advance notice of a predicted outcome. We showcase the approach over mortality and sepsis-3 prediction. Further, we examine whether models need to be trained for a specific notice period, or whether the approach could be incorporated at the evaluation level.

## Materials and Methods

We’ve created a set of Neural Network models that implement and evaluate the suggested approach using the MIMIC-III data. We do so by training and evaluating models with and without adding a constrain representing an “Alert Interval” between the prediction time and the prediction window.

## Results

We show that enforcing a notice period can significantly affect performance, but not for all outcomes prediction. Additionally, we see that the “Alert Interval” could be defined post-model training, with no significant performance loss, within the bounds of the trained lookahead.

## Conclusions

When evaluating the applicability of predictive models in the ICU, incorporating an advance-notice constrain to the model for some scenarios can be crucial, and in some cases, can change the results significantly. Doing so could be done over existing, already trained models. The concept of adding Alert-Interval could be applied to other clinical scenarios, where having advance notice is essential.

# Introduction

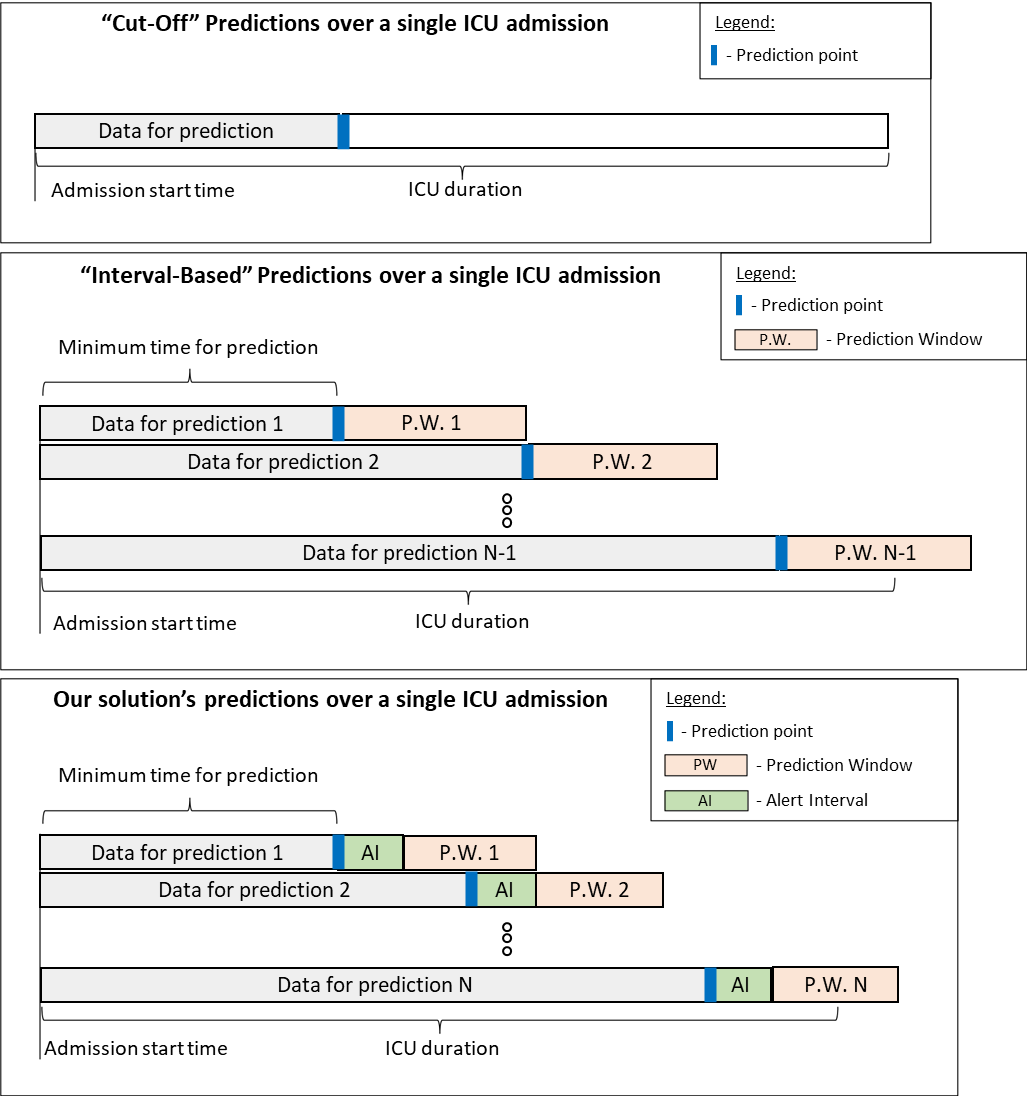
An intensive care unit (ICU) is a special department of a hospital or health care facility that provides intensive treatment care. Patients admitted to the ICU usually have severe or even life-threatening illnesses and injuries, and therefore are at high risk of mortality. The admitted patients are provided with constant care and close supervision. The goal of the ICU admission is to nurse the patients to a vigorous and stable condition, so they can be released from the ICU and continue to receive the care needed in a step-down unit or at home. However, not all admissions end up successfully. Statistics show that around 11.5% of patients admitted to the ICU die during admission.[1] Close monitoring and the adoption of Electronic Medical Records (EMR) has made patient data in the ICUs frequently sampled and abundant for leveraging Data-Science solutions. At these ICUs, where response time is critical, leveraging this data to provide risk alerts for patients’ future events (like death, sepsis onset, cardiac arrest, organ failure, etc.) can improve the care given in the ICU and reduce the death rate.

## Background

Predicting mortality, sepsis onset (or other types of events) within ICU admissions is not a new subject for research. A good literature review that covers this topic for mortality outcome prediction is ‘Development and validation of early warning score system: A systematic literature review”[2] and for sepsis onset prediction is “Prediction of sepsis patients using machine learning approach: A meta-analysis”.[3] Prediction models in this clinical scenario can be categorized into two main groups based on their approach: Cut-Off and Intervallic.

A Cut-Off model is a model which uses information from the first X hours to predict the outcome (for example, died or discharged) of the ICU admission (or in some cases, outcome after some period, like patient’s status after 24 hours from admission or 30 days from release). In this type of model, there is a single prediction per ICU admission. Common values for X are 24 hours and 48 hours.[3–13] A well-known clinical score that matches this profile is APACHE-II (Acute Physiology And Chronic Health Evaluation II),[14] which is applied within 24 hours of admission and assigns a death risk score according to several measurements. In contrast, an Intervallic model is a model which provides multiple predictions during the ICU admission. Each prediction refers to a prediction-window (a slice of time from the patient’s admission), where the prediction is based on the patient’s data up to the prediction-window, and it predicts whether the patient will have an event within the prediction-window’s time. Common setting for a prediction window is 1, 6, and 24 hours.[15–23] Some models use multiple Prediction-Window sizes, evaluating the forecasting ability of different “horizons”.[24]

**Figure 1:**

****

***Figure 1: Graphical illustration of two common prediction types following with the new proposed Notice approach.*** *In Cut-Off, each admission is assigned a single prediction at a pre-defined time, and the prediction is typically about the admission outcome. In Intervallic, each admission is assigned multiple predictions, depending on the length of admission and the prediction’s window-size. In the Notice approach, there are also multiple predictions per admission. However, the prediction window is distanced from the prediction time.*

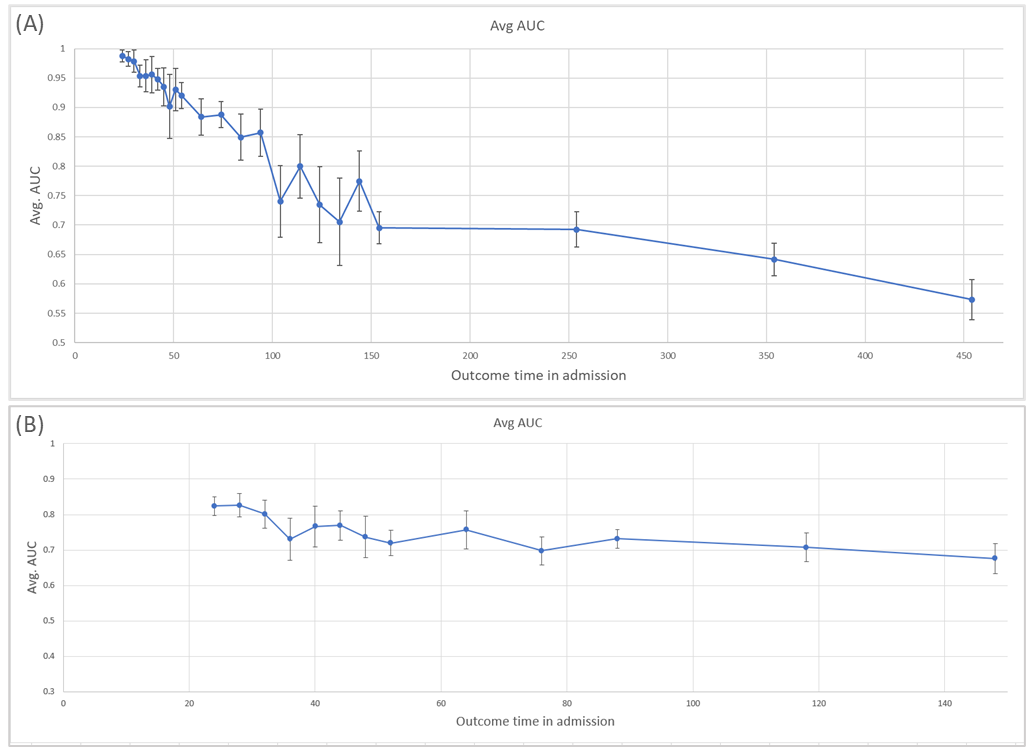
A recent paper that addressed and implemented both of the described model types is “Outcomes prediction in longitudinal data: Study designs evaluation, use case in ICU acquired sepsis”.[25] There are additional variations of these two types of models, which are less common/applicative, that we don’t implement for comparison in this model, like “Rolling Cut-Off” models, which is a hybrid of the two methods. In “Rolling Cut-Off” there is a sliding prediction point, similar to intervallic models, but the prediction is with regards to the rest of the admission.[26,27]

Each type of model has drawbacks. Cut-Off models are not scoped in time, making it hard to focus the efforts when read alerts are needed most. For example, if the Cut-Off time is after 24 hours and the admission duration is 168 hours, the prediction doesn’t tell us “when” in the remaining 144 hours of the admission, the risk of the event is the highest. Therefore, it may not be useful as a real-time alert system. Additionally, it leverages only the available data up to the Cut-Off point, regardless of the patient’s admission duration and when a prediction is needed/asked for. Looking at the above example, when 100 hours into admission, a prediction is required, it does not leverage the data between hour 24 and hour 100, including the most updated data.

In contrast, the Intervallic approach does provide a scoped prediction (for a specific prediction-window) and leverages the data up to the required prediction point. However, by definition, such models have a prediction window that is immediately following the time of prediction. This introduces a disadvantage that characterizes twofold: (i) Applicative-wise: this does not ensure a minimum advance notice period for intervention. For example, in the case where the model predicts a patient’s status every 6 hours. A patient with an event 31 hours after admission will get a negative prediction for the prediction-window of 24-30 hours within admission. For prediction-window of 30-36 hours, if a prediction is correct, the patient will get a positive prediction that gives only 1-hour alert before event time. (ii) Performance evaluation-wise: it can be easier for prediction models to predict events that will occur close to the prediction time over events occurring farther from prediction time. Clinical events are often gradual and progressive events. Predictions that occur adjacent to the predicted event can rely on signals that indicate that event in a “straight-forward” manner. Therefore, it can be considered as a type of data leaking. These disadvantages are actually relevant to the Cut-Off approach as well, however, with less significance.

One way to demonstrate the effect of the distance of the event from the prediction point on the performance’s evaluation is by breaking the ROC AUC performance of a “Cut-Off’s” model by the time of the event. As shown in Figure 2 for our showcased outcomes, when predicting mortality over data from the first 24 hours of admission, ROC AUC drops as admissions get longer. However, for sepsis prediction, the drop is mild.

**Figure 2:**

******

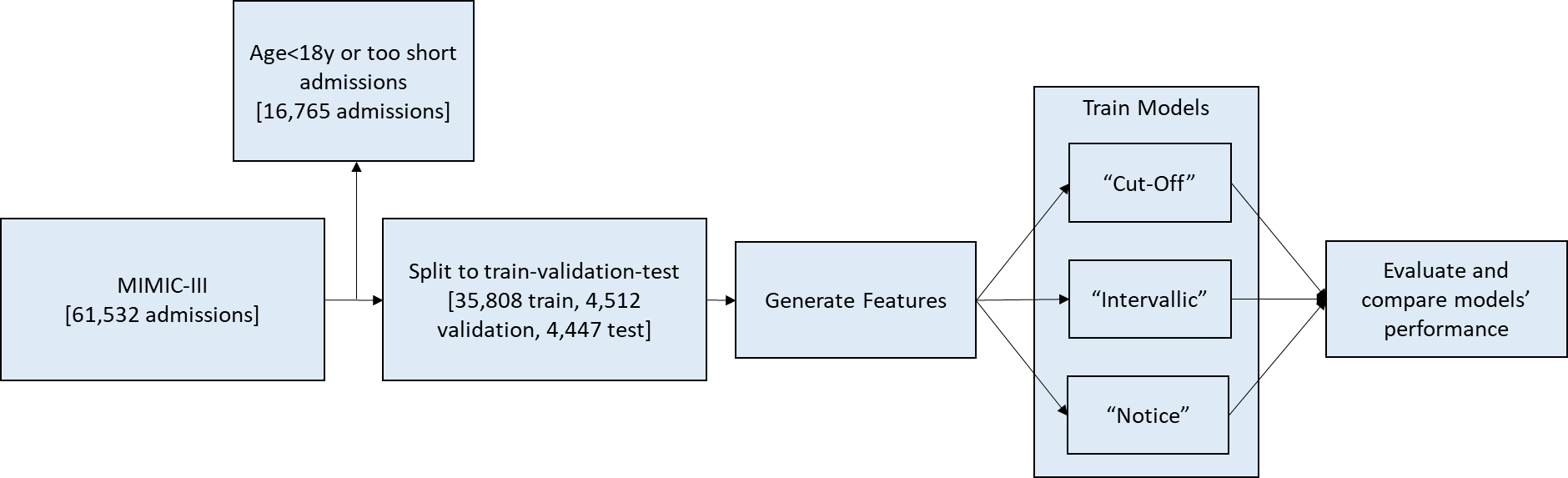
***Figure 2: Prediction performance of a Cut-Off model as measured by ROC AUC per admission length bin, using 10-fold cross-validation.******(A) Mortality prediction and (B) Sepsis onset prediction.*** *Performance drops for more prolonged admissions in mortality, while sepsis onset prediction is more stable. The bins are increasing in width to avoid bins with low number of samples. The ticks mark the bin’s start time, where the last bin has no upper limit. The error bars are the confidence intervals for each averaged data point over the x axis.*

# Objective

We are introducing and implementing a new approach for outcome prediction, and we demonstrate it for mortality and sepsis onset prediction during ICU admissions. The new approach is designed to take into account a minimal advance notice for alerting, while maintaining the prediction scoped like the Intervallic approach. We do so by adding an alert-Interval between the prediction time and the prediction window. We call this new type of models “Notice models”.

# Materials and methods

An overview of the study is presented in Figure 3. The new approach we introduce will be referred to as the Notice model, and its predictions are depicted in Figure 1 (in contrast to the “Cut-Off” and Intervallic models, also illustrated in Figure 1)



***Figure 3: Study overview****. The numbers in the flow vary a bit between configurations and the predicted outcome, as we filter admissions that are shorter than “Start-Time” + ”Alert Interval” (defined in “Formal problem definition” below) to prevent from predicting events that are within the model’s input signals. The numbers in the above flow chart are for the show-cased Notice configuration of 6h Alert-Interval, predicting mortality, detailed in table 1.*

## Formal problem definition

Formally, we want to be able to generate a predictor , where given:

* – Single ICU admission data, limited to events in time window , where , in hours
* – Label. Indicator whether the predicted outcome occurred for an ICU admission within time window , where , in hours
* – Start-time. Defines the time from admission, of the first prediction point, in hours
* – Prediction Step. Defines the time interval between each two prediction points, in hours
* – Prediction Window Size. Defines the length of each “Prediction Window”, in hours
* – Alert Interval. Sets the minimal notice in advance time for the prediction, in hours

Then for a given ICU admission, our target function is:

* =
* ICU admission has not concluded until time
* is the lookback (or observation window), chosen for the model. It is addressed as a hyperparameter to tune.

We refer to the set of values as a **configuration** when examining different models in the paper.

## Evaluation

We evaluated our models using Area Under the Receiver Operating Characteristics (AUC), conventionally used to evaluate such risk-prediction models. The drawback of evaluating AUC in Intervallic/notice models is that longer admissions are counted more times than shorter admissions, as these admissions appear in more prediction-windows. However, there is sense in evaluating the model in a way that gives each admission the same weight. For that we used Weighted AUC (WAUC), where every sample is weighted inversed proportionally to the number of samples (prediction windows) of that admission. The weights of all the predictions that belong to the same admission sums up to 1.

**AUC: Evaluate all predicted time-windows evenly:**

measure for predictions done at time for the outcomes at time

**WAUC: Evaluate all predictions, ICU normalized:**

## Clinical data and cohort

In this study we’ve used the MIMIC-III (Medical Information Mart for Intensive Care III) dataset from the Beth Israel Deaconess Medical Center (BIDMC), Boston, Massachusetts. The MIMIC-III database contains, clinical data, from 53,423 adult ICU stays from 38,597 adult patients.[29]

We’ve removed all admissions of patients that are under the age 18, or admissions shorter than the time of the first prediction-window. As a result, we excluded 8,656 admissions, resulting in remaining 44,767. For sepsis-3 outcome prediction, we’ve also filtered 21,208 admissions of patients’ that had first sepsis occurred prior to their first admission prediction window (per described definition in the “Acquiring labels” section), resulting in 23,559 admissions. For both outcomes, the split to train-validation-test was done in a patient level to avoid information leakage across ICU stays of the same patient such that all time windows of the same patients were not split between train, validation and test set. The train set consisted of 80% of the patients, validation 10% of the patients and test was 10% of the patients.

## Acquiring labels

Patient mortality and its time is logged in the MIMIC-III dataset’s ‘Admissions’ table. However, a parallel label does not exist for sepsis-3 events. Therefore, we followed the approach described at Goldstein et al. for identifying ICU acquired sepsis.[25] Sepsis-3 case was defined by a patient having a culture sampling and antibiotics administration within 24 hours from each other while also having sequential organ failure assessment (SOFA)[30] score of 2 or above within 24 hours. Sepsis time was defined as the first time of culture sampling or antibiotics administration. To avoid interpreting ongoing cases as several instances, we’ve kept only the first case of sepsis of each patient and added a constrain that there was no additional antibiotics administration in the 24 hours preceding the antibiotics administration of the diagnosed case.

## Data preprocessing

The dataset is constructed so that each prediction-point and its associated data and prediction-window is an independent entry. The prediction-point is defined as the time during the admission when a prediction is taking place. The prediction-window is defined as the time span for which the prediction gives the prediction about. The prediction-window’s label (i.e., case or control) is determined by whether the predicted outcome’s time is within that time window.

The data used for prediction include admission’s data from the Chart-Event and Lab-Event tables (containing laboratory test results, vital signs, diagnosis, provided procedures, provided meds, etc.), acquired prior to the prediction-point and during the defined lookback window by the configuration. In Notice models, we drop out the data accumulated between the prediction-point and the prediction-window (the Alert-Interval timeframe) (Figure 1). For each entry, we’ve used the following features:

* Demographic features including age, sex, admission type, and an indicator for whether the patient had a recorded previous ICU admission in the database.
* For each numerical type of measurement taken from the data (like heart rate, oxygen saturation, etc.), we’ve computed over its values within the data that is available to the prediction-point:
  + Count of measures
  + Minimal value
  + Maximal value
  + Average value
  + Variance of the values
  + First value
  + Last value
  + Warning/Flag count for abnormal measurements.
* For categorical measures (like oxygen delivery device), we’ve only taken the count/existence-indicator of measure.

The dataset is sparse and contains different scales as naturally, most admissions did not have values for most types of measures, and different measures have different value ranges. But missing data is missing not at random,[31] as missing data is often a result of caregivers’ decisions. For example, the caregiver decides to order specific laboratory tests, but not others, as the latter are not relevant for the patient in his current situation. As a result, the missing values have an inherent bias, compared to the general population, as they are a result of medical concerns. When handling missing values and normalizing, we wanted to make sure the information of not having a value for a certain feature doesn’t get lost (the missing values get assigned a designated, separable value from all non-missing values). We’ve chosen the following approach for normalizing each feature:

where is a set of all original values of the feature in the train set, and is set to 0.1. In this way, missing values are replaced with 0, and non-missing values are scaled such that the range of the training set is [0.1,1.1]. In this way, missing values are represented with a unique value 0, which does not appear in non-missing results. For non-missing results, we get a min-max scaling translated by a constant .

## Model development

We’ve trained a fully connected Deep Neural Networks model for every configuration examined. Each model type has its own hyperparameters tuned, using the validation set and the ROC AUC metric (rather than the WAUC). Then the models were evaluated on the test set for both AUC and WAUC.

For simplicity, we focus here on a small set of configurations. However, other configurations can be easily adapted, changing the alert-interval, prediction-window size, etc. For each predicted outcome, we’ve implemented the three models, Cut-Off, Intervallic, and Notice. We’ve created the Intervallic and Notice configurations as a pair so that the prediction horizon for the models is the same. However, an Alert-Interval is added for the Notice model, making the Notice configuration a more challenging task, as the events close to the prediction point are excluded.

For each configuration, we’ve tuned hyper-parameters independently to try and maximize its potential. The hyper-parameters include conventional Deep-Learning hyper-parameters along with a parameter of the lookback size the model used to compute the features for each prediction entry.

The show-cased configurations and models are detailed in Table 1.

**Table 1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Configuration** | **Configuration Specs** | | | | **Hyperparameters** |
| **ST** | **PS** | **PWS** | **AI** |
| Mortality  Cut-Off | 24 | Inf | Inf | 0 | Layer dims: 1200, 600, 250  weight decay: 0.0005  Lookback: 24 hours |
| Mortality Intervallic | 24 | 18 | 18 | 0 | Layer dims: 1000, 500, 200  weight decay: 0.00075  Lookback: 24 hours |
| Mortality Notice | 24 | 12 | 12 | 6 | Layer dims: 900, 400, 200  weight decay: 0.0004  Lookback: 24 hours |
| Sepsis  Cut-Off | 24 | Inf | Inf | 0 | Layer dims: 800, 400, 150  weight decay: 0.00075  Lookback: 24 hours |
| Sepsis Intervallic | 24 | 18 | 18 | 0 | Layer dims: 900, 400, 150  weight decay: 0.0004  Lookback: 24 hours |
| Sepsis  Notice | 24 | 12 | 12 | 6 | Layer dims: 600, 300, 100  weight decay: 0.0005  Lookback: 36 hours |

*Configuration parameters ST-Start time; PS-Prediction Step; PWS-Prediction Window Size; AI-Alert Interval. Dropout was set to 0.25 for all configurations.*

# Results

Table 2 details the results of each model configuration that was trained and evaluated with 10 different random model initializations. Results are presented as mean AUC and mean WAUC with 95% confidence interval. In Table 3, we detail the results of the Intervallic configurations models evaluated on their paired counterpart Notice configuration test sets.

**Table 2. Performances of three model architectures for both outcomes.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Outcome** | **Configuration** | **Mean AUC** | **AUC CI** | **Mean WAUC** | **WAUC CI** |
| Mortality | Cut-Off | 0.869 | 0.0026 | 0.869 | 0.0026 |
| Mortality | Intervallic | 0.891 | 0.0015 | 0.933 | 0.002 |
| Mortality | Notice | 0.866 | 0.0017 | 0.91 | 0.0021 |
| Sepsis | Cut-Off | 0.783 | 0.0046 | 0.783 | 0.0046 |
| Sepsis | Intervallic | 0.76 | 0.0034 | 0.739 | 0.0068 |
| Sepsis | Notice | 0.76 | 0.0065 | 0.722 | 0.0093 |

**Table 3. Evaluation of the Notice model on the intervallic test-set for both outcomes.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Outcome** | **Configuration** | **Test Set** | **AUC** | **AUC CI** | **WAUC** | **WAUC CI** |
| Mortality | Intervallic | Notice | 0.864 | 0.0016 | 0.906 | 0.003 |
| Sepsis | Intervallic | Notice | 0.758 | 0.0048 | 0.727 | 0.0087 |

When examining the results in Table 2, we can observe that the WAUC scores are higher than their AUC counterparts for mortality prediction while lesser for sepsis onset prediction. When comparing performance across different configurations, the ranking could change, depending on the used metric, AUC or WAUC. For mortality prediction, the Cut-Off model outscores the Notice model with respect to AUC, but the Notice outperforms with a higher WAUC.

For mortality prediction, the Intervallic model outperforms the Notice model, when evaluating each model with their configuration’s test set. We’ve expected this behavior, as the prediction-horizons for both tasks are the same, but the Notice model is not evaluated using close events, which are ‘easier’ to predict, as illustrated in Figure 2A. When evaluating the Intervallic models on its paired Notices test set, which incorporates alert interval (the size of the injected alert-interval to the Intervallic test set is equivalent to the parallel Notice model), the Intervallic’s model results dropped close to the Notice ones.

For sepsis onset prediction, the Intervallic and Notice models performed quite similarly according to the AUC metric, while there is a slight gap in favor of the Intervallic in the WAUC. Once again, when evaluating the Intervallic models on an adjusted test set, the Intervallic’s model results dropped close to the Notice ones.

# Discussion

We suggest a more adequate approach for evaluating alert systems in the clinical setting, incorporating a constrain for ahead notice into the model’s evaluation. This type of evaluation may be more adequate for such alert systems, as alerting for an event which is going to happen within a short time period, may not be helpful for the staff, as they may know about it already, or they may not be able to do anything to change it. The concept of creating an alert system with a prediction model is not limited to mortality or sepsis onset prediction but can be used for other clinical and even non-clinical settings.

Mortality prediction demonstrated a scenario where “immediate” signals give a strong indication for the upcoming outcome and when enforcing an advance notice, the results changed significantly. In sepsis onset prediction, the “immediate” signals didn’t affect the prediction much. We see this sits well with the observation illustrated in Figure 2. The slope of the mortality prediction is steep and the slope for sepsis is relatively stable, with almost similar performance for faraway events as close ones.

Comparing variant solutions and different architectures could result in having the Intervallic and Notice approaches rank models differently (due to different performance gaps between Intervallic and Notice in each solution), changing the selected “best” model, depending on which approach you take. In our future work, we plan to examine comparisons where this is the case. Additionally, we argue there is still work in incorporating this concept into an applicable system in ICUs. Having a confident short-notice prediction is also valuable and should be considered when planning a holistic solution. It was shown that alert systems integrated into the ICU have much lower AUC than expected.[32] We scope out from this paper the topic of generating a production alert-system from the models.

The fact that the Intervallic performs similar to the Notice model on Notice test set (rather than having the Notice outperform the Intervallic on the Notice), shows there is no gain from “focusing” on this specific subset of events in the prediction-window. This means that the alert-interval could be defined independently from the model development process, configurable in size after the model is trained. Although one can argue that for mortality prediction there was a statistically significant gain, we believe this gain is not sufficient and that the fact the models were tuned independently could also contribute to differences in performance.

While the Cut-Off model doesn’t have Alert-Interval, it generally predicts on further events than the Notice model. On the other hand, the bound that the Notice model provides on the predicted event is much tighter and more informative than the Cut-Off model. These are aspects way against each other, thus it’s hard to rank the tasks’ difficultness.

Naturally, different scenarios require different alert-intervals and different configurations in general. Therefore, we’ve kept the formal problem definition in general form. When shortening the alert-interval of a configuration, the results catch up to the Intervallic results, until the tasks unite when alert-interval is 0. We think there are other scenarios that would benefit from incorporating Notice models, with different configurations. Seizure prediction might benefit from shorter alert intervals, while discharge-readiness might require longer alert intervals. Moreover, the problem definition defined above could be furtherly generalized by transitioning to be:

This enables the problem’s fixed sizes parameters to be dynamically defined per prediction. It could serve applications like having smaller prediction windows or prediction steps at the start of an admission and expanding them as the admission duration increases or tuning the lookback according to the time in admission the prediction takes place. In this work we do not focus on these generalizations, however we believe they can be useful in future research. The fact that the alert-interval could be defined and applied after the model’s training plays well here.

# Conclusion

There are currently two main types of approaches for predicting outcomes in ICU admissions, Cut-Off and Intervallic. The Intervallic is the more applicative one. In our new Notice approach we suggest further improves the applicability of the Intervallic approach, in scenarios that benefit a heads-up on the predicted event of at least a pre-defined time. This is done by adding an Alert-Interval constrain over the model’s data. Empirical experiments show that adding this constrain could affect model performance significantly in some outcome predictions, resulting in better model evaluation (and better model selection, when comparing several models). Adding the alert-interval could be done at inference time alone (and not necessarily during training). This allows the alert-interval to be configured post training and to be applied on already existing, trained models. The concept of adding Alert-Interval could be applied to other clinical scenarios, where having advance notice is important. We also saw that there are scenarios where there is a significant difference between measuring this task with WAUC rather than with AUC.

# References

1 MIMIC-III, a freely accessible critical care database | Scientific Data. https://www.nature.com/articles/sdata201635 (accessed 14 Jul 2020).

2 Fu L-H, Schwartz J, Moy A, *et al.* Development and validation of early warning score system: A systematic literature review. *J Biomed Inform* 2020;**105**:103410. doi:10.1016/j.jbi.2020.103410

3 Islam MdM, Nasrin T, Walther BA, *et al.* Prediction of sepsis patients using machine learning approach: A meta-analysis. *Comput Methods Programs Biomed* 2019;**170**:1–9. doi:10.1016/j.cmpb.2018.12.027

4 Ge W, Huh J-W, Park YR, *et al.* An Interpretable ICU Mortality Prediction Model Based on Logistic Regression and Recurrent Neural Networks with LSTM units. *AMIA Annu Symp Proc* 2018;**2018**:460–9.

5 Awad A, Bader-El-Den M, McNicholas J, *et al.* Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *Int J Med Inf* 2017;**108**:185–95. doi:10.1016/j.ijmedinf.2017.10.002

6 Marafino BJ, John Boscardin W, Adams Dudley R. Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to ICU risk stratification from nursing notes. *J Biomed Inform* 2015;**54**:114–20. doi:10.1016/j.jbi.2015.02.003

7 McMillan S, Chia C-C, Esbroeck AV, *et al.* ICU Mortality Prediction using Time Series Motifs. ;:4.

8 Taori RN, Lahiri KR, Tullu MS. Performance of PRISM (Pediatric Risk of Mortality) Score and PIM (Pediatric Index of Mortality) Score in a Tertiary Care Pediatric ICU. *Indian J Pediatr* 2010;**77**:5.

9 Zhu Y, Fan X, Wu J, *et al.* Predicting ICU Mortality by Supervised Bidirectional LSTM Networks. ;:12.

10 Luo Y, Xin Y, Joshi R, *et al.* Predicting ICU Mortality Risk by Grouping Temporal Trends from a Multivariate Panel of Physiologic Measurements. ;:9.

11 Yu R, Zheng Y, Zhang R, *et al.* Using a Multi-Task Recurrent Neural Network With Attention Mechanisms to Predict Hospital Mortality of Patients. *IEEE J Biomed Health Inform* 2020;**24**:486–92. doi:10.1109/JBHI.2019.2916667

12 Marafino BJ, Park M, Davies JM, *et al.* Validation of Prediction Models for Critical Care Outcomes Using Natural Language Processing of Electronic Health Record Data. *JAMA Netw Open* 2018;**1**:e185097–e185097. doi:10.1001/jamanetworkopen.2018.5097

13 Nachimuthu SK, Haug PJ. Early Detection of Sepsis in the Emergency Department using Dynamic Bayesian Networks. *AMIA Annu Symp Proc* 2012;**2012**:653–62.

14 Barton C, Chettipally U, Zhou Y, *et al.* Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med* 2019;**109**:79–84. doi:10.1016/j.compbiomed.2019.04.027

15 Knaus WA, Draper EA, Wagner DP, *et al.* APACHE II: A severity of disease classification system. *Crit Care Med* 1985;**13**:818–29.

16 Aczon M, Ledbetter D, Ho L, *et al.* Dynamic Mortality Risk Predictions in Pediatric Critical Care Using Recurrent Neural Networks. *ArXiv170106675 Cs Math Q-Bio Stat* Published Online First: 23 January 2017.http://arxiv.org/abs/1701.06675 (accessed 4 Sep 2020).

17 Alves T, Laender A, Veloso A, *et al.* Dynamic Prediction of ICU Mortality Risk Using Domain Adaptation. *ArXiv191210080 Cs Stat* Published Online First: 20 December 2019. doi:10.1109/BigData.2018.8621927

18 HOURLY PREDICTION OF ORGAN FAILURE AND OUTCOME IN INTENSIVE CARE BASED ON DATA MINING TECHNIQUES: In: *Proceedings of the 12th International Conference on Enterprise Information Systems*. Funchal, Madeira, Portugal: : SciTePress - Science and and Technology Publications 2010. 270–7. doi:10.5220/0002903802700277

19 Yu K, Zhang M, Cui T, *et al.* Monitoring ICU Mortality Risk with A Long Short-Term Memory Recurrent Neural Network. *Pac Symp Biocomput Pac Symp Biocomput* 2020;**25**:103–14.

20 Desautels T, Calvert J, Hoffman J, *et al.* Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform* 2016;**4**:e28. doi:10.2196/medinform.5909

21 Johnson AEW, Mark RG. Real-time mortality prediction in the Intensive Care Unit. *AMIA Annu Symp Proc* 2018;**2017**:994–1003.

22 An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. - Abstract - Europe PMC. https://europepmc.org/article/PMC/5851825 (accessed 3 Sep 2020).

23 Mao Q, Jay M, Hoffman JL, *et al.* Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open* 2018;**8**:e017833. doi:10.1136/bmjopen-2017-017833

24 Shashikumar SP, Li Q, Clifford GD, *et al.* Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol Meas* 2017;**38**:2235–48. doi:10.1088/1361-6579/aa9772

25 Goldstein BA, Pencina MJ, Montez-Rath ME, *et al.* Predicting mortality over different time horizons: which data elements are needed? *J Am Med Inform Assoc* 2017;**24**:176–81. doi:10.1093/jamia/ocw057

26 Schvetz M, Fuchs L, Novack V, *et al.* Outcomes prediction in longitudinal data: Study designs evaluation, use case in ICU acquired sepsis. *J Biomed Inform* 2021;**117**:103734. doi:10.1016/j.jbi.2021.103734

27 van Wyk F, Khojandi A, Mohammed A, *et al.* A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. *Int J Med Inf* 2019;**122**:55–62. doi:10.1016/j.ijmedinf.2018.12.002

28 van Wyk F, Khojandi A, Kamaleswaran R. Improving Prediction Performance Using Hierarchical Analysis of Real-Time Data: A Sepsis Case Study. *IEEE J Biomed Health Inform* 2019;**23**:978–86. doi:10.1109/JBHI.2019.2894570

29 Johnson AEW, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;**3**:160035. doi:10.1038/sdata.2016.35

30 The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. ;:4.

31 Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons 2004.

32 Wong A, Otles E, Donnelly JP, *et al.* External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med* 2021;**181**:1065–70. doi:10.1001/jamainternmed.2021.2626