MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231057

Treatment Prediction in the ICU Using a Partitioned, Sequential, Deep Time Series Analysis

Michael SHAPIRO[a,1] and Yuval SHAHAR[b]
[a] *Department of Internal Medicine T, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel*
[b] *Department of Software and Information Systems Engineering (SISE), Ben-Gurion University*
Be'er Sheva, Israel
ORCiD ID: Michael Shapiro https://orcid.org/0000-0001-5943-6974, Yuval Shahar https://orcid.org/0000-0003-0328-2333

Abstract: We have developed a time-oriented machine-learning tool to predict the binary decision of administering a medication and the quantitative decision regarding the specific dose. We evaluated our tool on the MIMIC-IV ICU database, for three common medical scenarios. We use an LSTM based neural network, and considerably extend its use by introducing several new concepts. We partition the common 12-hour prediction horizon into three sub-windows. Partitioning models the treatment dynamics better, and allows the use of previous sub-windows' data as additional training data with improved performance. We also introduce a sequential prediction process, composed of a binary treatment-decision model, followed, when relevant, by a quantitative dose-decision model, with improved accuracy. Finally, we examined two methods for including non-temporal features, such as age, within the temporal network. Our results provide additional treatment-prediction tools, and thus another step towards a reliable and trustworthy decision-support system that reduces the clinicians' cognitive load.

Keywords. Treatment prediction, ICU, time series analysis, deep learning.

1. Introduction

Physicians deal with complex medical situations requiring fast decisions with significant repercussions. The current solution is a set of rigid and simple evidence-based rules, called *clinical guidelines*. However, this assumes the physician has made the right diagnosis, does not specify the appropriate dose, and ignores the disease progression pattern and the patients' response to the initial treatment. These limitations increase the variance in care, and thus leading to what Kahneman et al. consider as increased decision-making *noise* [1]. A computational solution in the form of a complete, context-sensitive decision support system can help inexperienced medical residents in guidance, error detection and education, *nudging* young clinician towards the right path [2].

[1] Corresponding Author: Michael Shapiro MD MSc, Internal medicine T, Sourasky Medical Center, 6 Weizmann St, Tel Aviv, 6423906, Israel. email: mikehpg@gmail.com.

In this study, we present a solution by extending a known deep learning methodology (LSTM) by several new considerations for predicting what treatment will be provided by an *average physician*, while accounting to the patient's unique characteristics and disease course. To evaluate our approach, we chose three common medical scenarios in the *Intensive Care Unit* (ICU) setting, for which several causes and temporal trajectories might be possible, potentially modifying the course of therapy.

2. Methods

2.1. Setting

This project was conducted using *the Medical Information Mart for Intensive Care* (MIMIC) IV database that contains de-identified data regarding 53,000 patients hospitalized in the ICU of Beth Israel Deaconess Medical Center between 2008 and 2019. It includes demographics, past medical history, laboratory results, and vital signs reported on an hourly basis, at the highest temporal granularity resolution in the database.

2.2. Time windows for treatment prediction

Medical treatment is a continuous process that takes time to fully manifest itself. As a logical window treatment, we chose a 12-hour horizon that covers both the immediate response and additional follow-up treatment. We then split it into three partitions of four hours, each providing sufficient time for administration of treatment and for measuring its effect. For training, the previous 24 hours' data were used as input, and the treatment given in the next four hours was the target (Figure 1). Each patient could have several clinical events, but they had to be a least 12 hours apart. Using a 4 hours sliding window, we "crawled" along the timeline, training our models.

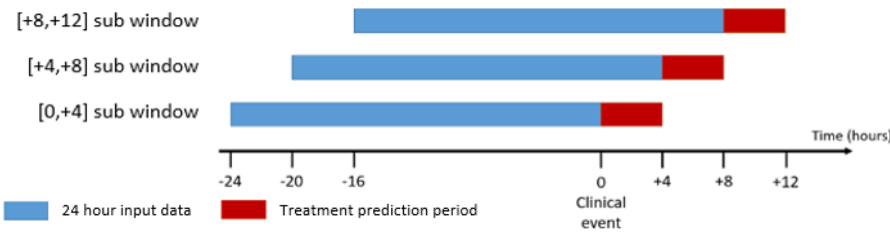


Figure 1. The input and prediction time periods for the three partitions of the treatment horizon.

2.3. Predictive and target variables

The input data variables used for prediction were chosen by a medical domain expert and included both temporal data - laboratory tests and vital signs, as well as constants - demographics and past medical history. Linear imputation was used to fill the missing values for achieving hourly granularity for the temporal data. Our goal was to predict the action performed by the actual Beth Israel Medical Center's attending physician (whoever that was), as recorded in the database - both the binary decision of prescribing any treatment, evaluated using AUC (area under curve) and the dose decision, evaluated using the mean square error (MSE) and mean absolute error (MAE). A multi-outcome architecture was used when several treatments were predicted. We evaluated the algorithm in three medical scenarios: 1. Hypoglycemia, low blood sugar, with the goal to predict intravenous (IV) glucose treatment; 2. Hypokalemia, low blood potassium, with the goal to predict oral (PO) and IV potassium treatment; 3. Hypotension, low blood pressure, with the goal to predict IV treatment with Fluids, Dopamine and Norepinephrine.

2.4. The predictive architecture and extensions

The architecture we designed for this study employs a recurrent neural network, specifically a bidirectional Long-Short Term Memory (LSTM) layer, followed by two fully connected layers. We extended this architecture by introducing several new concepts. First, we compared two methods for handling, by the augmented LSTM network, the integration of constant, a-temporal (e.g., demographic) data with temporal, dynamic data (e.g., results of a laboratory test): [1] An “External model” that analyzed the constants in a separate branch of the neural network and propagated changes in weights, as a result of the feedback, throughout both branches, with [2] an “Internal model” that incorporated the constants into the input of the LSTM by repeatedly adding them to the temporal features in each time-step. Second, we evaluated the use of earlier sub-windows' data (e.g. during hours [0-4] following the detection of the clinical state) as additional training data for the later sub-windows (e.g., for hours [4-8]) prediction, in addition to the use of training data from the later period. Naturally, the evaluation used test data belonging only to the later sub-window. Third, we proposed and evaluated a sequential prediction process composed of two classifiers: A classifier trained to make the binary decision whether to treat at all; followed, when relevant, by the application of a classifier trained separately for the quantitative decision of determining the magnitude of therapy. Both models were trained independently on all of the available data, rather than only on data of subjects receiving treatment, due to superior results we noted when learning from the full data set.

3. Results

3.1. Predicting treatment decisions across the partitioned 12-hours horizon

The three sub-windows differed in their treatment rate and dose, especially after a hypokalemia state, in which the treatment prevalence dropped, starting with the first time window following the detection of hypokalemia, from 52.6% to 15.4%. For other domains the reduction was milder, while the treatment dose dropped by 20% to 30%. Usually, the prediction accuracy for the first sub-window was higher than for the following two sub-windows. Using data from previous sub-windows resulted in up to 2% to 3% boost in AUC and 10% reduction in MAE and MSE. This effect was more significant for hypoglycemia (Table 1) and hypotension, in which the change across time windows was mild, compared to the variation in the domain of hypokalemia.

Table 1. Treatment prediction using previous sub-window data - hypoglycemia domain.

	sub-window data only[+4,+8]	**and [+4,+8] sub window data[0,+4]**
	Binary treatment prediction model	
Accuracy	0.66	0.681
AUC	0.727	0.746
	Dose prediction model	
MAE	11.739	10.726
MSE	21.584	17.226

AUC = Area under the ROC curve, MAE – Mean absolute error, MSE = Mean square error

3.2. Optimal use of constant features

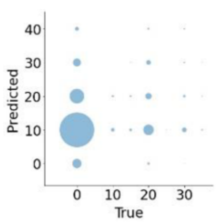
Using constants as part of the LSTM input ("*Internal model*"), led to a 1.5% improvement in AUC compared to the use of a branched network ("*External model*"), and to a 3.5% improvement compared to not using the constant data at all. Due to the superiority of the Internal model, it was used during the described experiments.

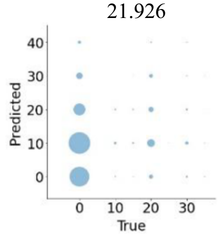
3.3. Sequential dose prediction

Using the sequential process resulted in a four to seven times increase in the sensitivity to "no treatment" decision in the hypokalemia domain (Table 2) and a lesser increase in other domains. In addition, using the sequential model resulted in a reduction of 9% to 39% in the medication-dose MAE in all scenarios, with a similar MSE. We believe this divergence occurs as MSE is disproportionally affected by large errors, due to squaring.

Table 2. Hypokalemia treatment prediction for IV treatment using the sequential model.

	Base model	**Sequential prediction**
MAE	16.888	15.273
MSE	21.702	21.926





MAE – Mean absolute error, MSE = Mean square error

4. Discussion

Medical treatment is a complex endeavor requiring the analysis of large quantities of data and devising a treatment plan according to guidelines and the elusive "gut feeling" gained by experience. In this study we constructed a temporal neural network framework that can model the complex time dependent nature of the treatment process and output the suitable common practice of typical Beth Israel Deaconess ICU clinician.

Treatment prediction is a hard task to crack even for IBM's Watson system, reported to have provided incorrect cancer treatment recommendations. Other notable treatment prediction include prediction of sepsis treatment in the ICU setting using reinforcement learning [3] and CURATE.AI, that predicts chemotherapy in prostate cancer. However, except for specific scenarios, it is hard to ascertain the effect of a single response on the patient's prognosis. Nevertheless, a decision support tool for clinicians is highly needed.

Thus, to provide a solution for the general scenario we postulate that mimicking the average physician's response will provide a highly useful tool for decision support, education and error detection. We used the common LSTM based architecture to analyze the prior 24 hours and issue a binary prediction of treatment and continuous dose prediction in a multi-outcome form. We proceeded to examine several applicational considerations to extend the basic architecture. First, we found that integrating constant features as part of the input into the LSTM performs better, perhaps because the temporal dynamic data is examined in the context of the constant data items. Then, we partitioned the 12-hour treatment window into three equal sub-windows and showed that the first four hours are the most active and treatment slows, sometimes extremely so, afterwards. Despite this, the use of data from previous sub-windows to train models for future periods resulted in improved results, thus alleviating a major problem in the healthcare domain of insufficient training data. Finally, we showed that a sequential model significantly improves the "No Treatment" prediction, as well as the mean accuracy of absolute dose prediction, both important aspects for making a decision support system trustworthy.

This study has several limitations. First, further research is needed to evaluate the effect of deviation from the proposed prediction on the *outcome* of the patients. Second, our algorithms were evaluated on one large database; assessing them on an additional database can provide better insight about the generalizability of the results.

5. Conclusions

This study examined treatment prediction in three scenarios and showed the successful extension of a recurrent neural network with constant features for mimicking the action of ICU physicians. We believe that using this architecture can be a stepping stone to a decision support system for a young clinician, or provide an alarm for mistreatment.

Acknowledgements: This work was partially supported by ONR award No. N629091912124, by the IDF Medical Corps, and by the Kahn-Orion scholarship.

References

[1] D. Kahneman, O. Sibony, and C.R. Sunstein, Noise: A Flaw in Human Judgment, Little, Brown Spark, 2021.

[2] R.H. Thaler, and C.R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness, Penguin, 2009.

[3] M. Komorowski, L.A. Celi, O. Badawi, A.C. Gordon, and A.A. Faisal, The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care, *Nat. Med.* **24** (2018) 1716–1720. doi:10.1038/s41591-018-0213-5.