ICU Outcome Predictions Using Real-Time Signals with Wavelet-Transform-based Convolutional Neural Network

Yiqun Jiang
Iowa State University
yiqunj@iastate.edu

Shaodong Wang
Iowa State University
shaodong@iastate.edu

Qing Li
Iowa State University
qlijane@iastate.edu

Wenli Zhang
Iowa State University
wlzhang@iastate.edu

Abstract

Intensive care units (ICUs) serve patients with life-threatening conditions. The limited ICU resources cause severe economic and healthcare burdens worldwide. It is critical to conduct ICU outcome predictions at an early stage and promote efficient use of ICU resources. However, all the current prediction methods have limitations such as unsatisfactory accuracy and depending on resource-demanding laboratory tests or expert domain knowledge. In this research, we design a wavelet-transformed-based convolutional neural network, WTCNN, which only requires patients' vital sign series and information at ICU admission for real-time ICU outcome predictions. The model is evaluated using a large real-world ICU database and outperforms state-of-art baselines on both ICU mortality and length-of-stay prediction tasks. We conduct LIME for model interpretation and prescriptive analysis. Our work provides an efficient tool for ICU outcome predictions, allowing healthcare providers to take action promptly on patients at risk and reduce the negative impacts on patient outcomes.

Keywords: Healthcare analytics, wavelet transform, deep learning, CNN, LIME

1. Introduction

The intensive care unit (ICU) provides constant care to patients with life-threatening illnesses and injuries. ICU outcome predictions, in terms of both mortality and length of stay, can reduce the economic and healthcare costs due to critical care needs (Zimmerman et al., 2006; Almashrafi et al., 2016). From the perspective of mortality rates, ICUs are the highest mortality units in almost all healthcare institutions because ICU patients tend to be extremely susceptible to negative outcomes as a result of their severe medical conditions. Among the 4 million ICU admissions in the United States (US) every year, the reported average mortality rates range from 8% to 19% (Halpern & Pastores, 2010). ICU outcome predictions can determine the severity of the illness and the efficacy of treatments and interventions (Pirracchio et al., 2015). From the perspective of patients' length of stay, prolonged ICU stays explain a significant fraction of the healthcare cost. Intensive care accounts for 13.7% of hospital costs and 4.1% of national health expenditures (Halpern & Pastores, 2010). ICU outcome predictions are useful to guide medical-resource allocation, improve care quality, and decrease hospitalization costs (Almashrafi et al., 2016). During public health threats such as the COVID-19 pandemic, the need for ICU outcome predictions has been amplified because the hospitals have been overburdened with patients, and many of them require critical care.

Because of the importance of ICU outcome predictions, much research has been done in this field over the last two decades. There are three major types of methods currently available to medical practitioners: severity scoring systems, machine learning models, and deep learning models. All these methods, however, have their limitations. For the severity score systems, researchers and healthcare practitioners have questioned their accurateness and consistency, complained about the long wait for laboratory results as predictor variables (Goswami et al., 2010), and expressed dissatisfaction with the fact that they need expert assessments as input (Reith et al., 2017). Therefore, our first research question is how we can design a new ICU outcome prediction model that only uses real-time information without the requirement of laboratory results and expertise from intensivists, and is more accurate.

Over the last ten years, electronic-ICU has been adopted to capture the status of patients' vital functions using bedside monitors, which sheds new light on ICU outcome prediction problems. The eICU system provides enormous volumes of time series of vital sign data (Pollard et al., 2018), which have revealed various dynamical patterns (Lehman et al., 2015). These dynamic patterns can be used to inform prognosis, provide early warnings for life-threatening

URI: https://hdl.handle.net/10125/103091
978-0-9981331-6-4

conditions, and predict outcomes for ICU patients (Lehman et al., 2015). For example, the vital sign series (also referred to as time-series data in this research) provides information about basic body functions from different aspects with different time spans. The short-time changes such as abrupt increases or decreases in the vital signs are usually critical messages for the changes in patients' health conditions. The long-time trends of the series indicate the stability of the patients' health conditions.

However, due to their great number and diversity of dispersion, these dynamic patterns are difficult to detect. More and more researchers turn to machine learning and deep learning models. Despite their potential, both machine learning and deep learning-based models have limitations on their performance due to feature extraction issues or limitations on their applicability for interpretability issues. The next generation of ICU outcome prediction models, according to researchers' and practitioners' expectations, should be more accurate and timely (Kramer et al., 2014). Hence, our second research question is how we can effectively extract accurate features from ICU bedside monitoring data that is readily accessible.

We find that convolutional neural networks (CNNs) and wavelet transforms are well-suited approaches to address our research questions. CNNs, as types of deep learning models, automatically and adaptively learn spatial hierarchies of features from low to high-level patterns, which are capable of capturing the features in time-series data in terms of both short-time changes and long-time trends (Li & Guan, 2021). Thus, CNNs are effective deep learning model structures in ICU outcome predictions relying on time-series data. However, existing research finds that the performance of CNNs is negatively affected by the great amount of noise in time-series data (Habibi Aghdam et al., 2016). The wavelet transform, a powerful signal processing technique, can effectively enhance the signals in time-series data by eliminating noise and strengthening signals, such as short-time changes and long-time trends, in the time series. Therefore, the wavelet transform can further enhance the predictive power of CNN models.

Combining CNN and wavelet transform, we design a wavelet-transform-based convolutional neural network (WTCNN) with a specific structure dealing with time-series data by adding a wavelet convolutional layer in CNN. The WTCNN can reduce the negative impact of noises on ICU patients' time-series of vital signs and improve the ICU outcome prediction accuracy. Meanwhile, the proposed model only requires patients' vital sign series and information at ICU admission as inputs for real-time predictions and does not depend on laboratory test results or medical domain knowledge.

To demonstrate the effectiveness of the proposed model, we evaluate the model on five datasets extracted from the eICU database–a large multi-center critical care database for ICU outcome predictions by using the five most common admission diagnoses. We compared the results with the state-of-art baselines. The proposed method outperforms the baselines by a large margin on four datasets and achieves comparable accuracy on the rest of one dataset, demonstrating that the proposed model effectively predicts the outcomes of ICU patients. Besides, we interpret the prediction results using local interpretable model-agnostic explanations (LIME) (Ribeiro et al., 2016) for prescriptive analysis, by which we visualize the prediction results and provide actionable suggestions for ICU healthcare practitioners.

The contributions of this work are two-fold. First, it is the first design that incorporates the wavelet transform into a convolutional neural network intended for time series analysis in the medical field, providing a novel tool for medical time series analysis. The superior performance in ICU outcome predictions provides strong evidence that the model structure is well-suited for extracting useful signals from time-series data. Second, from the perspective of societal impacts, the proposed model predicts ICU outcomes with solely the real-time vital sign series and information gathered at admission, thus can deliver crucial guidance for the healthcare professionals and medical institutions in a timely manner. The prediction results can further be used for hospital management to reduce healthcare costs.

2. Background

2.1. Related work

ICUs serve patients with high-severity illnesses. In the United States, one in every five fatalities happens during or immediately after ICU treatment (Angus et al., 2004). Meanwhile there are 4 million ICU admissions each year (Halpern & Pastores, 2010), with the average length of stay for terminal ICU hospitalization being 12.9 days and the cost of $24,541 (Angus et al., 2004). There are at least two motives for early ICU outcome predictions. First, early ICU outcome predictions facilitate early interventions which can decrease the mortality risk for patients in ICUs (Doiron et al., 2018). Second, the scarcity and high cost of intensive care resources make it crucial to conduct ICU outcome predictions and promote efficient use of such resources.

Three major types of methods currently available to intensive care practitioners are severity scoring systems, machine learning models, and deep learning models.

In the last decade, severity scoring systems have been widely applied for ICU outcome predictions, such as SAPS II, MPM0-II and APACHE IV, among which APACHE IV is a widely used benchmark in related research (Zimmerman et al., 2006). These models have high explainability, but the presumption of the linear relationship between predictors and the response variable limits them from handling complex datasets.

Machine learning models step on the stage afterward to focus on accuracy improvement in predictions. Pirracchio et al. (2015) and Kim et al. (2019) use classic machine learning models such as random forest and support vector machine (SVM) in ICU mortality predictions. These studies require manual feature engineering. The performance of such studies is unsatisfactory due to the lack of efficient feature extraction techniques.

Recently, deep learning models have been developed to further improve ICU outcome prediction performance. Deep learning structures, including CNN (Caicedo-Torres & Gutierrez, 2019; Kim et al., 2019), recurrent neural network (RNN) (Thorsen-Meyer et al., 2020), and bidirectional long short-term memory (Bi-LSTM) (Sheikhalishahi et al., 2020) have been applied for ICU mortality and length of stay predictions. However, the existing work simply applies common deep learning models without specific structural adjustments to the structures according to ICU outcome predictions.

Most existing research makes predictions for patient outcomes at 24H after ICU admissions (Pirracchio et al. 2015). Several other studies offer predictions at multiple time points, including at the time of admissions and 48 hours after admissions (Sheikhalishahi et al., 2020; Thorsen-Meyer et al., 2020). Specifically, real-time ICU outcomes are difficult to predict. Harutyunyan et al., (2019) have made such an attempt but without success. In this study, we seek to offer a model that can predict ICU outcomes both in real-time and at different time points.

Most studies for ICU outcome predictions have used features extracted from multiple data modalities, including medical domain knowledge, demographic data, vital signs, and lab results (Harutyunyan et al., 2019). Among these data, lab results usually take hours to days to obtain, while acquiring domain knowledge requires medical expertise (Goswami et al., 2010). As a result, using these two types of data for real-time prediction is not feasible. On the contrary, the demographic data and vital signs are easily accessible. The vital signs can be monitored in real-time and obtained by electrical monitors, which do not require expertise. Nevertheless, existing studies normally rely on data from multiple sources to achieve better results which hinders the possibility of timely or real-time ICU outcome predictions. Furthermore, the existing studies fail to extract the most important features of patients' vital sign data. Current studies usually simplify patients' vital sign data by categorizing or taking a long-period average, such as a one-hour average (Zhang et al., 2020), which results in information loss on the vital signs.

In this study, we present a new model with real-time prediction capability to address the research gaps mentioned above. Our model efficiently extracts information from vital signs through wavelet transforms and condenses the information by taking advantage of CNN.

2.2. Wavelet transform

A wavelet is an oscillation with wave-like characteristics localized in time, which has an amplitude that begins and ends at zero. The wavelet transform changes the original signal in different time extensions by choosing different wavelets for corresponding frequency analysis while retaining the shape of a signal. In other words, the wavelet transform decomposes the original signals into wavelets with different stretched rates. This decomposition provides both frequency-domain and time-domain information, which can reveal critical patterns and has been widely used in signal processing (Orlando Oliveira et al., 2019).

Using small and large wavelet functions, the wavelet transform can identify fine and coarse details in signals, respectively. This allows the wavelet transform to represent various patterns by separating the different details in data, such as long-term trends, breakdown points, and discontinuities in higher derivatives (Orlando Oliveira et al., 2019). This is important to our study because both long-term trends and sudden changes in the vital signs reflect the crucial health conditions of the patients.

In healthcare, the wavelet transform has been successful in extracting features from several forms of complex biosignals, such as electrocardiograms, medical sounds and images, blood flow, and blood pressure signals (Addison, 2005). However, no existing study uses wavelet transform to represent vital sign data in ICU outcome predictions. We expect the wavelet transform can also extract useful information from vital sign series.

2.3. Convolutional neural network (CNN)

As we mentioned in Section 2.1, deep learning has been recently introduced to ICU outcome predictions because of its capability to handle complex patterns, thus improving prediction accuracy. CNNs are one of the most powerful tools in pattern recognition (Kim et al., 2019). Through the use of the appropriate filters, a CNN is able to successfully capture the hidden spatial and temporal relationships of features that are otherwise difficult to capture (Szegedy et al., 2015).

This research proposes a novel structure based on CNNs with a wavelet convolutional layer, denoted as WTCNN. We design the structure of WTCNN to enhance useful signals of time series data and efficiently extract features. The wavelet convolutional layer offers a simultaneous localization in both the time and frequency domains, projecting the time series data to a two-dimensional output. Multiple convolutional layers after the wavelet convolutional layer are used to capture local relationships between frequency characteristics in the adjacent time periods, condense the information and compute new features.

Compared to traditional CNNs, the proposed WTCNN has distinct advantages in terms of better fitting the temporal vital sign data. Combining signal processing techniques with deep learning models, the proposed model can extract important features more efficiently. In other words, during the training of a standard CNN, the filters can only capture information from the time domain of the time-series data. WTCNN includes a wavelet transform layer that adds frequency-domain knowledge as important characteristics, improving ICU outcome predictions' performance.

3. Research Design

This research proposes a new model, WTCNN, as illustrated in Figure 4, by adding a wavelet convolutional layer to the traditional CNN with specific structures designed to process ICU patients' time-series of vital signs effectively. The wavelet convolutional layer in WTCNN takes advantage of the discrete wavelet transform to enhance the vital sign signals. In this section, we first introduce the wavelet convolutional layer and then present the structure of the proposed model.

3.1. Discrete Wavelet Transform (DWT)

As explained in Section 2.2, the unique ability of simultaneous spectral and temporal analysis makes the wavelet transform a powerful tool in signal processing. The wavelet transform calculates how much of a wavelet is present in a signal at a given scale and time. If the wavelet matches the shape of the signal well at a given scale and time location, as seen in Figure 1, a relatively high transform value is produced. The transformed value is small otherwise. The process which measures how a wavelet modifies the shape of a signal is a convolution operation. In the wavelet transform, the signal is convoluted with a wavelet of a specific scale at each time location, as shown in Figure 1.

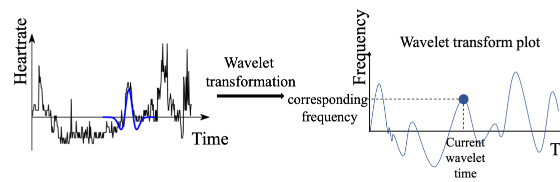

Figure 1. An example of the wavelet transform on vital sign series

The DWT provides a sparse representation for signals to compress the signal. In actual practice, it is computed using a filter bank. To be more specific, a one-level filter bank passes a signal through both a high pass filter and a low pass filter to decompose it into low and high frequencies correspondingly, which are called approximation coefficients (ACs) and detail coefficients (DCs), respectively. For example, when a vital sign series v_i, $i = 1, ..., t$ are being passed through a low-pass filter with a wavelet function g, the convolution result of v_i and g is AC, calculated by

$$f_{WT}(v_i) = (v_i * g)[t] = \sum_{k=-\infty}^{\infty} v_i[k]g[t-k].$$

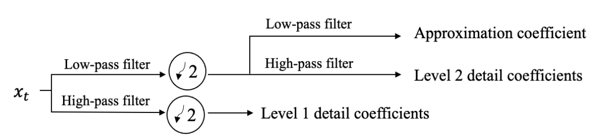

Figure 2. A two-level filter bank

An example of a two-level filter bank is shown in Figure 2. Generally, an n-level filter bank repeats the process on the approximation coefficients from the one-level filter bank n times and gets the final approximation coefficients and detail coefficients at different levels.

Figure 3 presents an example of a 2-level bank decomposition of heart rate series. The ACs catch the low-frequency information of the series or "trend." The levels 1 and 2 DCs are highpass representations, which are noise in this case. Therefore, we use only the ACs for further analysis.

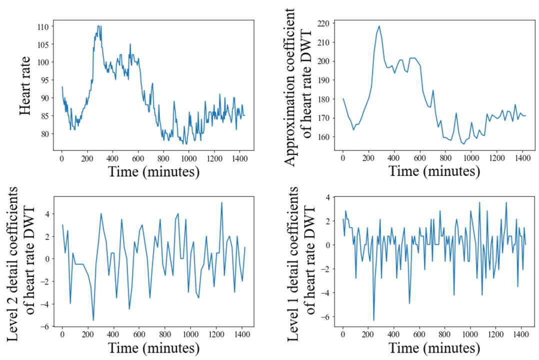

Figure 3. A 2-level bank decomposition of heart rate

3.2. Network Structure

The ICU outcome prediction task is defined as follows: for an ICU admission, there are records of n vital sign series $v_1, v_2, ..., v_n$ over the time interval $[0, t]$ for the patient. The ICU outcomes include patient discharge status $y_{mortality}$ and length of stay y_{los}. In this research, we define binary classification problems to predict if the patient's discharge status is expired ($y_{mortality} = 1$) or alive ($y_{mortality} = 0$), and if the length of stay is longer than a specific time period ($y_{los} = 1$) or not ($y_{los} = 0$). WTCNN is designed to give the probability of each class, thus providing critical information of ICU patients' health conditions. As shown in Figure 4, the first layer of WTCNN is a wavelet convolutional layer, which takes vital sign series v_i as inputs. The outputs of this layer are in the format of matrices, denoted by $X_i = f_{WT}(v_i) \in R^{1\times m_i n}, i = 1, 2, ..., n$ and m_i is the number of the time stamps for v_i.

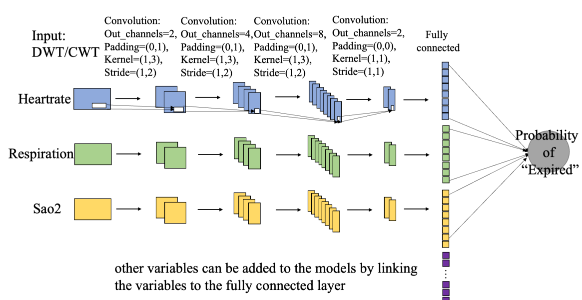

Figure 4. The structure of WTCNN

Given the input matrix obtained from the wavelet convolutional layer of each vital sign, an output of the convolution layer is $u_i = xReLU(BN(W_{c_i} X_i))$ with the weight matrix $W_{c_i} \in R^{c_i \times 1}$ ((c_i, 1) is the convolutional kernel size), the rectified linear unit $ReLU(x) = max(x, 0)$, and the batch normalization function $BN(x) = \frac{x - E[x]}{\sqrt{Var[x] + \varepsilon}} * \gamma + \beta$ (γ and β are the learnable parameter vectors. The batch normalization and rectified linear unit help with faster, stable and effective model training.

4. Experiments and Evaluation

In this section, we evaluated the proposed model in the experiments on mortality and length-of-stay predictions. First, we introduced the experiment setting. Second, we refined the model and determined the parameters for the proposed model. Third, we presented the model performance by comparing it with five baselines. Last, we interpreted the model through a local interpretable model-agnostic explanations algorithm (LIME).

4.1. Experiment Setting

Since the wavelet transform could not be performed on time series with missing values, we filled the missing values in a series with values at their nearest time location. If a patient's entire vital sign series was never recorded, we filled the values with the mean of the population in the dataset.

The data were split into train/validation/test sets with proportions of 75%, 18.75%, and 6.25%, respectively, by stratified sampling based on ICU outcomes because of the imbalance of positive vs. negative cases in the data. The parameter tuning was performed on the training and validation sets. The model evaluation was performed on untouched testing sets to obtain a reliable assessment. To eliminate the randomness of data splitting, we repeated the whole training, validation, and testing process ten times. The results were evaluated by an average area under the curve (AUC) of the ten trials.

A weighted loss function, weighted cross entropy $l(y, \hat{y}) = - y * log(\hat{y}) * w - (1 - y) * log(1 - \hat{y}) * (1 - w)$ was used to accommodate outcome imbalances during optimization with the prediction to be $\hat{y}$ and the ground truth y. The weight w was chosen as the proportion of $y = 1$.

We implemented the model in Pytorch using the Adam optimizer with batch size 16 and learning rate 0.001. The model was trained using 20 epochs with early stopping based on AUC of the validation set.

4.1.1. Data description. The testbed comes from eICU Collaborative Research Database, a multi-center intensive care unit database containing over 200,000 admissions across the US. We retrieved the datasets of the five most frequent admission diagnoses, which were 'Sepsis, pulmonary'(SP), 'Infarction, acute myocardial (MI)', 'cerebrovascular accident (CVA)', 'Congestive heart failure' (HF), and 'Sepsis, renal/UTI' (SR). Among these five admission diagnoses, the ratio of patients expired was 12% for SP and was around 5% for the other four. The length of stay distribution showed a left-skewed characteristic, with most patients discharged within 3 days. Table 1 displayed the information for the patients of each dataset.

In this research, we adopted the vital signs with a missing rate of less than 50% to limit the influence of missing data. The vital signs used in this research included heart rate, respiration, sao2 (oxygen saturation), st1, st2, and st3 (estimated ST-segment level x of the ECG, $x \in \{1, 2, 3\}$). The vital signs were initially recorded at 1-minute intervals and stored as 5-minute medians in the database. The vital signs were useful in body condition assessment, thus providing useful information in ICU outcome predictions.

Table 1. Descriptive statistics of the datasets

admission diagnoses	HF	CVA	MI	SP	SR
# patient	4,840	5,284	5,919	6,823	4,284
# admission	6,522	6,552	7,159	8,761	5,235
expire rate	0.054	0.050	0.032	0.122	0.064
length of stay (los)	>3 days: 32.40% >5 days: 15.81% >7 days: 9.18% >10 days: 4.28%	>3 days: 23.35% >5 days: 11.28% >7 days: 6.75% >10 days: 3.91%	>3 days: 15.72% >5 days: 6.62% >7 days: 3.45% >10 days: 1.649%	>3 days: 41.56% >5 days: 24.88% >7 days: 16.38% >10 days: 9.38%	>3 days: 30.36% >5 days: 13.85% >7 days: 7.28% >10 days: 3.36%
heart rate	84.11	79.69	79.23	88.87	88.20
respiration	21.05	19.55	19.79	21.80	20.82
sao2	95.56	96.61	95.99	95.27	96.34
st1	1.35	1.94	2.42	1.71	0.71
st2	2.51	2.91	4.33	2.47	1.14
st3	2.35	2.15	4.15	2.09	0.90

4.1.2. Prediction tasks. The objective of ICU mortality predictions was represented as a binary classification problem. As shown in Table 1, the majority of admission records for all five datasets had discharged status of alive, with only a small portion having expired, indicating an imbalanced classification problem. We evaluated the classification model with AUC values and compared the model performances with state-of-art baselines.

Existing studies on length-of-stay predictions mainly focused on normal vs. prolonged length of stay (Almashrafi et al., 2016). Hence it was more meaningful to model the length-of-stay prediction task as binary classification. However, the cut-off points used to define the prolonged lengths of stay vary. Since the length of stay distribution was extremely left-skewed and most admission records had a length of stay less than 10 days, we chose 3, 5, 7, and 10 days as criteria. In addition, we compared the results with APACHE IV, a "golden standard" in ICU outcome predictions, by estimating the AUC values of its regression results of the length of stay.

4.2. Model refinement

We refined the model by selecting the most effective vital signs and adding patients' information at ICU admissions to our model.

4.2.1. Vital sign contribution analysis. In this subsection, we used the ACs of a single vital sign as the model input for investigating each vital sign's contribution in mortality predictions. The results were concluded in Table 2. With the ACs of heart rate, respiration, or sao2 as inputs, the proposed model achieved high AUC values ranging from 0.67 to 0.82. The AUC values of the model using the ACs of st1, st2, or st3 as inputs ranged from 0.52 to 0.55. Considering the fact that the AUC value of 0.5 indicated the model had no discriminative ability, these models showed poor prediction performance. Therefore, in this study, we adopted heart rate, respiration, or sao2 as model inputs.

Table 2. Performance (AUC) using a single vital sign

Vital sign	Disease				
	HF	CVA	MI	SP	SR
heart rate	0.673	0.721	0.820	0.713	0.717
respiration	0.710	0.716	0.765	0.690	0.707
sao2	0.679	0.750	0.780	0.727	0.770
st1	0.528	0.545	0.567	0.524	0.532
st2	0.519	0.532	0.559	0.528	0.540
st3	0.518	0.533	0.540	0.519	0.550

4.2.2. Addition of patients' information at ICU admissions. In the experiment above, we only used the vital signs as the model inputs. In the following experiments, we added patients' information at ICU admissions to the fully connected layer in WTCNN. These variables contained patients' demographics and statuses documented at ICU admissions, all of which were available at admissions and would not affect the model's real-time prediction ability.

4.3. Model performance

We evaluated the model performance from two aspects. First, we compared WTCNN with five state-of-art baselines on ICU mortality predictions and three baselines on length-of-stay predictions using 24H vital sign series and patients' information at ICU admissions. Then we evaluated the model's real-time prediction ability by comparing its performance with the golden standard of ICU outcome predictions, APACHE IV, with vital sign series of different time spans from 3H to 24H on ICU outcome predictions.

4.3.1. Performance on mortality predictions at 24H after admissions. We compared WTCNN with five baselines: APACHE IV (Zimmerman et al., 2006), a benchmark method widely applied in ICU outcome predictions, and four state-of-art deep learning models: CNN_Torres (Caicedo-Torres & Gutierrez, 2019), CNN_Kim (Kim et al., 2019) , LSTM (Thorsen-Meyer et al., 2020), and BiLSTM (Sheikhalishahi et al., 2020). As shown in Table 3, with patients' information at ICU admission and three vital signs as inputs, WTCNN reached the best AUC compared with all the five baselines for HF, MI, SP, and SR. The patients' demographics enrich the information of patients' heterogeneity and thus help to improve the model performance by a large margin. WTCNN achieved comparable performance to APACHE IV (i.e., highest AUC) on the CVA. Of particular note was that the AUC of WTCNN was larger than the best result of the five baselines by more than 3% for HF and 5% for SR.

Table 3. Performance on mortality predictions

	HF	CVA	MI	SP	SR
CNN_Torres	0.538	0.657	0.561	0.649	0.622
CNN_Kim	0.722	0.855	0.85	0.795	0.758
LSTM	0.641	0.691	0.834	0.718	0.716
BiLSTM	0.677	0.797	0.836	0.81	0.746
APACHE	0.75	**0.88**	0.88	0.78	0.81
WTCNN	**0.780**	0.875	**0.897**	**0.823**	**0.863**

4.3.2. Performance on length-of-stay predictions at 24H after admission. We applied WTCNN to predict if the patient was discharged within 3, 5, 7, and 10 days. In the experiments of length-of-stay predictions, we compared WTCNN with three baselines, including APACHE IV, LSTM, and BiLSTM.

Table 4. Performance on length-of -stay predictions

		HF	CVA	MI	SP	SR
3 days	**WTCNN**	**0.702**	**0.725**	**0.788**	**0.759**	**0.701**
	APACHE	0.660	0.664	0.697	0.702	0.694
	LSTM	0.610	0.684	0.706	0.642	0.677
	BiLSTM	0.622	0.674	0.714	0.710	0.682
5 days	**WTCNN**	**0.688**	**0.750**	**0.777**	**0.739**	**0.708**
	APACHE	0.660	0.664	0.697	0.702	0.697
	LSTM	0.625	0.700	0.670	0.672	0.683
	BiLSTM	0.617	0.680	0.750	0.724	0.697
7 days	**WTCNN**	**0.693**	**0.775**	**0.805**	**0.747**	**0.732**
	APACHE	0.660	0.664	0.697	0.702	0.697
	LSTM	0.600	0.739	0.687	0.662	0.721
	BiLSTM	0.590	0.752	0.757	0.714	0.706
10 days	**WTCNN**	**0.669**	**0.757**	**0.822**	**0.734**	**0.732**
	APACHE	0.660	0.664	0.697	0.702	0.697
	LSTM	0.592	0.674	0.729	0.725	0.662
	BiLSTM	0.538	0.743	0.684	0.718	0.656

As Table 4 illustrated, WTCNN outperformed all the baselines in all cases in terms of AUC scores. The AUC values of WTCNN for CVA, MI, SP, and SR were all larger than 0.7, while the AUC values of HF under different conditions are at least 0.67, indicating the strong predictive power for the length of stay predictions. Among all the ICU admission diagnoses, the proposed model achieved the best performance on the dataset of MI, with AUC ranging from 0.78 to 0.82.

4.3.3. Real-time prediction results compared to APACHE IV. In this subsection, we compared the results of WTCNN with APACHE IV, which was a golden standard in ICU outcome predictions that predicted ICU outcomes, including mortality and length of stay, with 24H records of patient data. We made comparisons on mortality and length of stay predictions on different time spans.

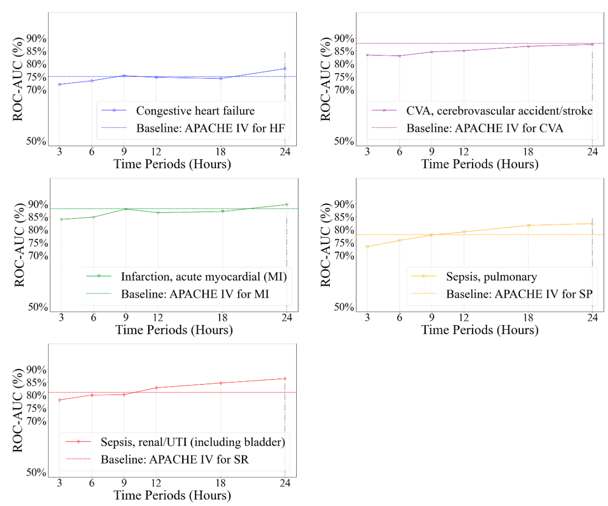

Figure 5. Real-time ICU mortality prediction

For real-time mortality predictions, we applied WTCNN for mortality predictions on different time spans of 3H, 6H, 9H, 12H, 18H, and 24H, respectively. As shown in Figure 5, the AUC score had an increasing trend with a prolonged time span. For CVA, MI, and SR, the AUC values with 9H records were larger than 0.8. On most datasets, the model outperformed APACHE IV using data within 9H after

ICU admission, indicating a strong ability to early determine the patients' health status. The AUC of WTCNN using 24H records had an average of a 3% increase compared to APACHE IV.

For the real-time length of stay predictions, we used three days as the prediction outcome (i.e., $y_{los>3} = 1$ or $y_{los>3} = 0$) because that was the patients' average length of stay in the ICUs. Then we used different lengths of patients' vital signs (i.e., from 3H to 24H) as input for analysis. As Figure 6 showed, the AUC of the model increased with more time series of vital signs. On the dataset MI, the model showed a strong ability to give early predictions as soon as 3H after ICU admissions. The model achieved better performance on all the five datasets with 24H worth of data compared to APACHE IV.

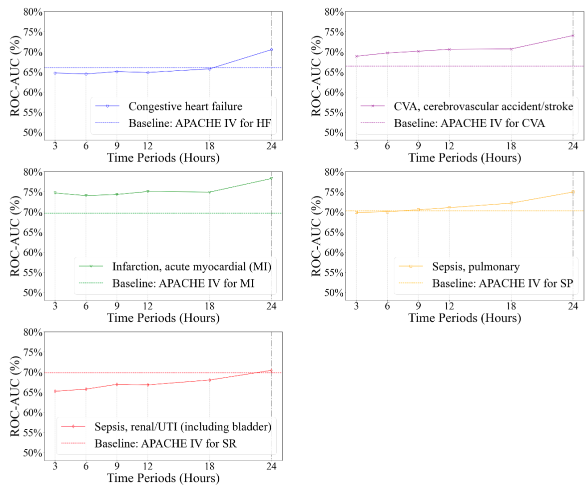

Figure 6. Real-time ICU length of stay prediction

4.4. Prescriptive analysis based on model interpretation

In this section, we illustrated the effectiveness of vital signs in mortality predictions through local interpretable model-agnostic explanations (LIME). LIME interpreted the inputs' contribution to prediction results by fitting an approximate linear model on the inputs. With LIME, we were able to provide prescriptive suggestions for intensive care practitioners.

We took the last layer of WTCNN as the vector representation of each vital sign. Then LIME fitted a linear model on the vector representations, where the weights reflected the feature contribution. In this case, a negative weight indicated a tendency to predict the patient as "alive" while a positive weight indicated a tendency to predict the patient as "expired." The absolute value of the weight determined the impact of changes in features on the predictions.

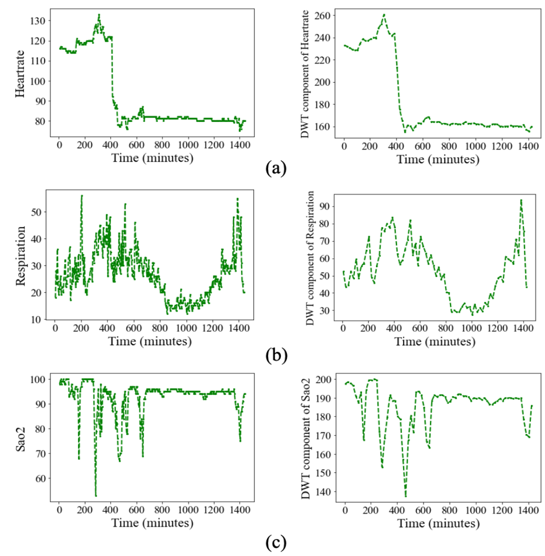

Figure 7. Vital signs & corresponding AC (alive)

Figure 7 plotted the three vital sign series and the corresponding ACs of an exemplary patient with discharge status "alive." The ACs showed the low-frequency trends of the vital sign series, thus providing a clear view of how vital sign series change. As Figure 7 showed, the patient had a steady value in the normal ranges of heart rate and sao2 after 6.6 hours in ICU, indicating strong evidence of health status. The respiration was not stable, but there were no long-time fluctuations.

Figure 8. LIME coefficients of vital signs (alive)

Figure 8 presented the LIME weights of WTCNN representations based on the three vital signs. Most of the weights were negative, which implied a strong tendency to predict the patient to be alive. The absolute feature weight value based on respiration representations was smaller, which resulted from the instability of the series.

The out-of-range values of the vital signs were clear signs to alert the intensive care provider to take prompt action on the patients. If the interventions stabilized the vital signs, it indicated the treatment was effective and could increase the patient's survival rate.

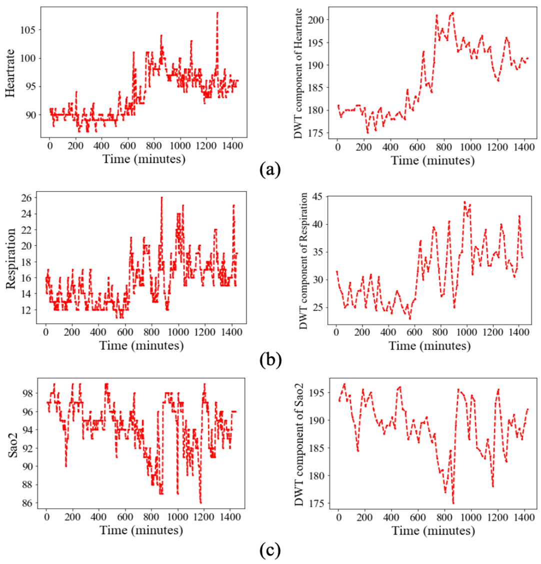

Figure 9. Vital signs & corresponding AC (expired)

Figure 9 plotted the vital sign series and the corresponding approximation coefficients of an exemplary patient with discharge status "expired." Unlike patient 1, none of the three vital sign series of patient 2 had reached a steady state during the first 24 hours in ICU, indicating a large probability of "Expired" in the end. The linear approximation model by LIME gave probability 1 for "Expired."

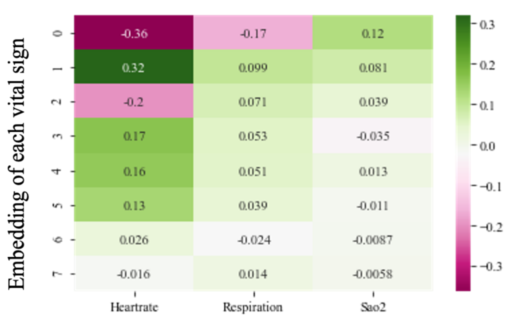

Figure 10. LIME coefficients of vital signs (expired)

For all three vital signs' representations, the weights given by LIME were mostly positive, as shown in Figure 10, which was also indicative of a high "Expired" possibility. The large fluctuations of vital signs were a strong indication of high-risk patients, which required the medical provider to step in and provide timely treatment.

Overall, our analysis could direct intensive care providers' attention to patients whose vital signs were out of the normal ranges and had long-time fluctuations, prompting them to take early interventions and eventually improve care quality and decrease the hospitalization costs.

5. Discussion and Conclusion

Limited resources and the high mortality risk of ICU necessitate immediate actions. Correspondingly, undertaking ICU outcome predictions is critical in encouraging the effective use of ICU resources. The emergence of electronic-ICU brings enormous informative vital sign series, benefiting ICU outcome predictions. However, the patterns in the vital sign series are difficult to discern due to their large quantity and diversity of dispersion. Based on these facts, we investigated two research questions. One, how to design a new ICU mortality prediction model that uses real-time information, requires no lab tests or expertise from intensivists, and is more accurate. Two, how to effectively extract accurate and interpretable features from ICU bedside monitoring data that is readily accessible.

We propose a novel model, WTCNN, by incorporating signal processing techniques into a deep learning model. The proposed model enhances the useful signals in time-series data by a wavelet convolutional layer and successfully extracts and condenses the features using the CNN structure. The model has been demonstrated to have high accuracy in ICU outcome predictions on a real-world multi-center ICU database. On mortality predictions, the model outperforms the state-of-the-art baselines by a large margin. In addition, the model can achieve similar accuracy of predictions with merely 3 to 9 hours of data to generally acknowledged approaches like APACHE IV, which uses 24 hours of data. On length-of-stay predictions, the model achieves better AUC values on all the prediction tasks compared to APACHE IV. More importantly, the extracted and condensed features from the model are interpretable. We conduct prescriptive analysis based on the local interpretability weight of each vital sign by analyzing the features with LIME.

The major contributions of this research are two-fold. From the aspect of methodology innovation, first, the model requires only three ICU bedside monitoring vital sign series which are easy to obtain. Second, we are one of the first studies to incorporate signal processing techniques into deep learning models for ICU outcome predictions. Third, the proposed model condenses the vital sign series to vector representations, which can be further combined with other information for ICU outcome predictions. From the aspect of societal impacts, our research provides an efficient and high-accuracy tool for ICU outcome predictions, allowing healthcare providers to take action promptly on patients at risk and reduce the negative impacts on patient outcomes. Furthermore, the ICU resources can be prioritized to offer earlier interventions to patients with impending risk for failure, and later, slots to patients with deferred risk.

6. References

Addison, P. S. (2005). Wavelet transforms and the ECG: A review. *Physiological Measurement*, *26*(5), R155–R199.

Almashrafi, A., Elmontsri, M., & Aylin, P. (2016). Systematic review of factors influencing length of stay in ICU after adult cardiac surgery. *BMC Health Services Research*, *16*(1), 318.

Angus, D. C., Barnato, A. E., Linde-Zwirble, W. T., Weissfeld, L. A., Watson, R. S., Rickert, T., ... & Robert Wood Johnson Foundation ICU End-Of-Life Peer Group. (2004). Use of intensive care at the end of life in the United States: an epidemiologic study. Critical care medicine, 32(3), 638-643.

Caicedo-Torres, W., & Gutierrez, J. (2019). ISeeU: Visually interpretable deep learning for mortality prediction inside the ICU. *Journal of Biomedical Informatics*, *98*, 103269.

Doiron, K. A., Hoffmann, T. C., & Beller, E. M. (2018). Early intervention (mobilization or active exercise) for critically ill adults in the intensive care unit. Cochrane Database of Systematic Reviews, (3).

Goswami, B., Singh, B., Chawla, R., Gupta, V. K., & Mallika, V. (2010). Turn Around Time (TAT) as a Benchmark of Laboratory Performance. *Indian Journal of Clinical Biochemistry*, *25*(4), 376–379.

Habibi Aghdam, H., Jahani Heravi, E., & Puig, D. (2016). Analyzing the Stability of Convolutional Neural Networks against Image Degradation: *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 370–382.

Halpern, N. A., & Pastores, S. M. (2010). Critical care medicine in the United States 2000–2005: An analysis of bed numbers, occupancy rates, payer mix, and costs*: *Critical Care Medicine*, *38*(1), 65–71.

Harutyunyan, H., Khachatrian, H., Kale, D. C., Ver Steeg, G., & Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. *Scientific Data*, *6*(1), 96.

Kim, S. Y., Kim, S., Cho, J., Kim, Y. S., Sol, I. S., Sung, Y., Cho, I., Park, M., Jang, H., Kim, Y. H., Kim, K. W., & Sohn, M. H. (2019). A deep learning model for real-time mortality prediction in critically ill children. *Critical Care*, *23*(1), 279.

Kramer, A. A., Higgins, T. L., & Zimmerman, J. E. (2014). Comparison of the Mortality Probability Admission Model III, National Quality Forum, and Acute Physiology and Chronic Health Evaluation IV Hospital Mortality Models: Implications for National Benchmarking*. *Critical Care Medicine*, *42*(3), 544–553.

Lehman, L. H., Adams, R. P., Mayaud, L., Moody, G. B., Malhotra, A., Mark, R. G., & Nemati, S. (2015). A Physiological Time Series Dynamics-Based Approach to Patient Monitoring and Outcome Prediction. *IEEE Journal of Biomedical and Health Informatics*, *19*(3), 1068–1076.

Li, H., & Guan, Y. (2021). DeepSleep convolutional neural network allows accurate and fast detection of sleep arousal. *Communications Biology*, *4*(1), 18.

Orlando Oliveira, M., Horacio Reversat, J., & Alberto Reynoso, L. (2019). Wavelet Transform Analysis to Applications in Electric Power Systems. In D. Baleanu (Ed.), *Wavelet Transform and Complexity*. IntechOpen.

Pirracchio, R., Petersen, M. L., Carone, M., Rigon, M. R., Chevret, S., & van der Laan, M. J. (2015). Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *The Lancet Respiratory Medicine*, *3*(1), 42–52.

Pollard, T. J., Johnson, A. E. W., Raffa, J. D., Celi, L. A., Mark, R. G., & Badawi, O. (2018). The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific Data*, *5*(1), 180178.

Reith, F. C., Synnot, A., van den Brande, R., Gruen, R. L., & Maas, A. I. (2017). Factors Influencing the Reliability of the Glasgow Coma Scale: A Systematic Review. *Neurosurgery*, *80*(6), 829–839.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Sheikhalishahi, S., Balaraman, V., & Osmani, V. (2020). Benchmarking machine learning models on multi-centre eICU critical care dataset. *PLOS ONE*, *15*(7), e0235424.

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9.

Thorsen-Meyer, H.-C., Nielsen, A. B., Nielsen, A. P., Kaas-Hansen, B. S., Toft, P., Schierbeck, J., Strøm, T., Chmura, P. J., Heimann, M., Dybdahl, L., Spangsege, L., Hulsen, P., Belling, K., Brunak, S., & Perner, A. (2020). Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: A retrospective study of high-frequency data in electronic patient records. *The Lancet Digital Health*, *2*(4), e179–e191.

Zhang, D., Yin, C., Zeng, J., Yuan, X., & Zhang, P. (2020). Combining structured and unstructured data for predictive models: A deep learning approach. *BMC Medical Informatics and Decision Making*, *20*(1), 280.

Zimmerman, J. E., Kramer, A. A., McNair, D. S., & Malila, F. M. (2006). Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today’s critically ill patients*: *Critical Care Medicine*, *34*(5), 1297–1310.