

A Deep Learning Post-Discharge Mortality Prediction Method Considering Diagnosis Information for ICU Patients

Mucan Liu

Institute of Systems Engineering, Dalian University of Technology

Chonghui Guo (✉ dlutguo@dlut.edu.cn)

Institute of Systems Engineering, Dalian University of Technology, PR China

Shaoze Cui

Institute of Information and Decision Technology, Dalian University of Technology

Research Article

Keywords: Intensive care units, Mortality prediction, Deep learning, Diagnosis information, Shapley values

Posted Date: November 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1027700/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A Deep Learning Post-Discharge Mortality Prediction Method Considering Diagnosis Information for ICU Patients

Mucan Liu¹, Chonghui Guo^{1*} and Shaoze Cui²

^{1*}Institute of Systems Engineering, Dalian University of Technology, No.2 Linggong Road, Dalian, 116024, Liaoning, China.

²Institute of Information and Decision Technology, Dalian University of Technology, No.2 Linggong Road, Dalian, 116024, Liaoning, China.

*Corresponding author(s). E-mail(s): dlutguo@dlut.edu.cn;
Contributing authors: liumucan@mail.dlut.edu.cn;
shaoze-cui@foxmail.com;

Abstract

Background and Objective: Mortality prediction is widely used to stratify patients into different risk categories and to provide prognosis evaluation. Nowadays, scoring systems, which predict mortality with some scores reflecting the severity of diseases and physiological states of patients in ICU, have been widely applied for in-hospital mortality prediction. Many research works which focus on designing better machine learning models and algorithms for mortality prediction also have achieved great performance. However, it is not enough to make post-discharge prognosis of mortality only with the aid of better models and algorithms while richer patient related information can make big differences. In this study, we propose a deep learning method considering patient diagnosis information for post-discharge prognosis of mortality. This method can help to significantly improve the performance of prediction. Further more, we propose a method of calculating disease Shapley values to evaluate the mortality risk brought by disease factors. *Methods:* Deep learning models including long short-term memory (LSTM) and temporal convolutional network (TCN) are trained with patient physiological time series data and diagnosis information of

different prevalence to predict post-discharge mortality risk of different time windows. Disease Shapley values to evaluate the mortality risk brought by disease factors are the weighted average of marginal contributions of diseases to patient mortality. Experiments of several post-discharge mortality prediction tasks of different time windows are conducted on the large freely accessible MIMIC-III dataset. To provide more sufficient comparison, the diagnosis information is also introduced for traditional machine learning models. *Results:* In our experiment, LSTM achieves highest AUROC and the improvements of which are 8.67%, 9.68%, 13.33%, 12.32% and 12.25% with the help of diagnosis information for five post-discharge mortality prediction tasks of different time window. Several patient examples are shown to present the mortality risk brought by disease factors, of which the analysis results are in line with clinical experiences. *Conclusions:* In general, our proposed method can improve performance of ICU patient post-discharge mortality prediction and help to evaluate how much do different kinds of diseases which a patient suffers from increase his mortality, thus providing support for clinical decisions.

Keywords: Intensive care units, Mortality prediction, Deep learning, Diagnosis information, Shapley values

1 Introduction

In the medical system, intensive care unit (ICU) is specially used to treat severely ill or unconscious patients. To provide life monitoring and support for these critically ill patients, ICUs are equipped with the most comprehensive treatment and monitoring equipment, such as ventilators, electrocardiogram monitors, blood gas analyzers, etc. Even with advanced equipment and professional medical staff, the mortality rate of ICU patients is still very high. Mortality prediction, which is an important patient outcome prediction task, could help to quantify the severity of the patient's physiological condition [1]. By reflecting the severity of diseases or the prognosis of patients, mortality prediction can help clinicians to make better clinical decisions and help patients to better know the physiological status of themselves [2]. Hence, ICU patient mortality prediction has been a very hot research topic in the field of medical informatics.

In early studies, experts developed a series of scoring systems to assess the in-hospital mortality of ICU patients, such as Logistic Organ Dysfunction system (LODS) [3], Sequential Organ Failure Assessment (SOFA) [4], Acute Physiology, Chronic Health Evaluation (APACHE) [5], etc. However, due to the non-specific characteristic and stringent linear constraint of model, predictions made by scoring systems can be inaccurate for individual patient [6]. Nowadays, massive volumes of data recorded in electronic health records (EHRs) also supported researchers to design models and algorithms for

in-hospital mortality prediction which aim at improving the predicting performance and facilitating the clinical decision making. Outperformance of mortality prediction methods based on machine learning models have been shown by many works [6–13]. Some of them show better prediction performance of machine learning models than traditional scoring systems [7, 8] and some of them develop various models and machine learning algorithms for mortality prediction [6, 9–13]. Further more, the deep learning models achieve particularly satisfactory performance in ICU mortality prediction tasks due to their strong ability of capturing non-linear patterns hidden in data [14]. Many researchers develop deep learning models which can deal with various modals of data and provide valuable clinical decision support [15–20].

Studies mentioned above have made progress in in-hospital mortality prediction, while post-discharge mortality is also of concern to both clinicians and patients [21]. Because of the complex illness and disease history of patients, the prognosis of post-discharge long-term mortality is a big challenge [22]. However, without the limitation of making prediction in hospital, the post-discharge mortality prediction models can take more patient related information into consideration. For example, Grnarova et al. propose an automatic mortality prediction method based on the unstructured textual content of clinical notes, which brings improvement to the difficult problem of post-discharge mortality prediction [23]. Although the extent to which disease events before ICU admission affect prognosis has been debated [24], the information about the diseases patients suffering from is of potentially value for post-discharge mortality prediction. Christiansen et al. reveal that morbidities have impact on mortality among ICU patients by analysing the mortality of patient groups with Charlson Comorbidity Index (CCI) levels [22]. Nielsen et al. investigate how the disease history of different length of years affects mortality prediction and reveal that taking account of long-term and short-term disease history can give more precise prognostic estimates than scoring systems in which only a small number of comorbidities are included in the computation [25]. Dai et al. conduct stastical analysis with MIMIC-III database, revealing closely correlations between diseases patients suffer from and mortality of patients [26]. For those reasons, we consider investigating the effect of diagnosis information during the ICU admissions for post-discharge mortality prediction task with deep learning models. The mian difference between our work and [25] is that they aggregate a long history of diseases while we consider the diagnosis information during admissions for mortality prediction. And they investigate effect of different length of disease history while we innvestigate the diagnosis information of different prevalence for post-discharge mortality prediction. Additionally, because the causes of death of post-discharge patients are cared about by clinicians [27], we provide a method to evaluate mortality risk brought by disease factors based on each individual’s physiological condition and the diagnosis information.

In this article, we propose a deep learning post-discharge mortality prediction method considering diagnosis information for ICU patients. In our

proposed method, easily available patient vital signs data and diagnosis information are fused to predict post-discharge mortality for ICU patients, which can greatly improve the predicting performance of the deep learning models. Besides, the computation of disease Shapley values is introduced to evaluate mortality risk brought by different disease factors.

The main findings and contributions are summarized as follows:

(1) The proposed post-discharge mortality prediction method considering patient diagnosis information during the admissions can significantly improve the prediction performance of deep learning models and traditional machine learning models.

(2) The method of computing disease Shapley values, is used to evaluate the hazard the diseases cause to a single patient. With the aid of disease Shapley values, individual patient's disease condition can be analyzed to help doctors clarify the priority of multiple diseases and facilitate subsequent treatment.

(3) The proposed methods are tested on MIMIC-III dataset on the following five post-discharge mortality prediction tasks: will a patient die within 30 days, within 90 days, within 180 days, within 365 days and within 5 years. Our method is presented to improve the prediction performance for both deep learning models and traditional machine learning models. The evaluation results with disease Shapley values are in line with clinical experiences.

The remainder of this study is organized as follows. In Section 2, the materials and methods part, we first expound detailed process of data extraction and preprocessing. Then we introduce our deep learning method of predicting ICU patient post-discharge mortality considering diagnosis information and the calculation of disease Shapley values. The experimental results and analysis are presented in Section 3. Finally, some conclusions, limitations and future work discussions are presented in Section 4.

2 Materials and methods

In this section, the preparation of data and the proposed method are introduced in detail. An overview of our proposed method is presented in Figure 1. In the following part, we first give an introduction of vital signs and diagnosis information data extraction and preprocessing with MIMIC-III database in Section 2.1. In Section 2.2, the method of fusing patient physiological data and diagnosis information is introduced first. Then, deep learning models including long short-term memory (LSTM) and temporal convolutional network (TCN) are introduced. In Section 2.3, we introduce the method of computing disease Shapley value.

2.1 Data extraction and preprocessing

2.1.1 Data extraction

In this study, we conduct our experiments and analysis on the publicly available medical information mart for intensive care III (MIMIC-III) dataset [28],

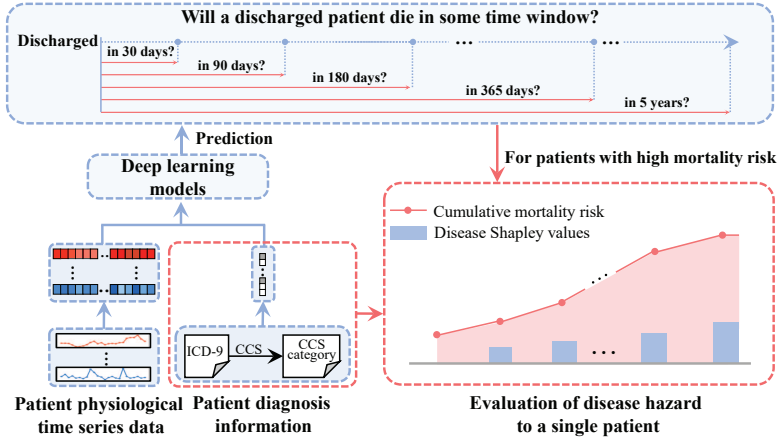


Fig. 1 An overview of the proposed ICU patient post-discharge mortality prediction method considering patient diagnosis information

which can be used to conduct research after application. This dataset contains information from the electronic health record (EHR), including patient demographic information, vital signs, lab events, diagnosis information, etc. In MIMIC-III, each patient is unique with a SUBJECT_ID. Each SUBJECT_ID corresponds to one or more HADM_IDs which means a patient has one or more hospital admissions. And for each HADM_ID, one or more ICUSTAY_IDs can be matched, which means a patient may be admitted to the ICU one time or several times within an admission. Beginning with exhaustive clinical data of over 60,000 ICU stays of 40,000 patients, we apply a similar patient cohort selection inclusion criteria following previous benchmark research [29]. First, admissions with multiple ICU stays or transfers between different ICU units or wards are excluded, which aims to reduce the ambiguity of outcomes associated with hospital admissions rather than ICU stays. Then, we exclude admissions which correspond to patients died in hospital because we aim at training the model for patients who discharge alive. And admissions which have ICU stays less than 48 hours are excluded because we will extract data of 48 hours for mortality prediction. Finally, considering the differences between adults and pediatric physiology, we drop the patients younger than 18. After above process, we get a cohort with 18,324 ICU stays.

Then, we extract basic vital sign data and diagnosis information for patients who meet the selection inclusion criteria. The vital sign data of each admission are represented as a sequence of EHR events with timestamps in MIMIC-III database. There are three most important elements in each event: a timestamp at which the event is recorded; the name of the event; and a numerical value of the event. We group the events of each patient into one-hour intervals by timestamp. The time window of patient data is set as 48 hours, which means for each physiology variables, 48 values (may include missing values) are used. The diagnosis information in MIMIC-III is recorded in the form

of International Classification of Diseases (ICD). In MIMIC-III database, the code version is ICD-9. For each patient, we extract a list of ICD-9 codes from DIAGNOSES_ICD table.

2.1.2 Patient mortality label preparing

After identifying the patient cohort for modeling and analysis, we prepare the mortality labels for all of them. According to the discharged time and death time recorded in the database, we calculate the time interval between discharge and death of patients. The prediction time window in our study is set as 30 days, 90 days, 180 days, 1 year and 5 years referencing some studies [30–32]. For each patient, we assign 5 labels which mean whether a discharged patient will die in 30 days or not, in 90 days or not and so on. The binary labels of each patient i can be represented as y_i^{30d} , y_i^{90d} , y_i^{180d} , y_i^{1y} and y_i^{5y} , respectively.

2.1.3 Preprocessing of patient vital sign data and diagnosis information

Vital sign multivariate time series. The vital sign raw data are extracted as form of EHR events with timestamp. In this study, we select 17 predictors, including capillary refill rate, diastolic blood pressure, fraction inspired oxygen and so on, for mortality prediction. The statistical summary of the predictors is shown in Table 1. With observation window of 48 hours opted, all the events occur in which are partitioned into 48 one-hour periods according to the timestamps. The last observation value is kept if several values of some predictor exist in the same one-hour period. The missing values are imputed with the most recent value of the predictor if there is one and pre-specified normal values otherwise. Binary mask inputs for each predictor at each timestamp are also provided to represent whether a value is a real one or an imputed one. And the categorical variables are processed into one hot vector form. The numeric predictors are dealt with standard normalization. After above process, a matrix $\mathbf{X}_i = (x_{nt}^{(i)})_{N \times T}$ is used to represent values of the N variables during the T hours of patient i . N and T are 76 and 48 in our study, respectively. The element $x_{nt}^{(i)}$ in the matrix is the n th variable value at timestamp t . For deep learning models such as LSTM and TCN, the matrix of shape time-steps \times features (48 \times 76 in our study) can be directly received as input data.

Diagnosis information. In our study, the diagnosis information of patients is used to facilitate the post-discharge mortality prediction of ICU patients. After the data extraction process, we get a list of ICD-9 codes for each patient admission. There are more than 14,000 diagnosis codes in ICD-9. Representing each disease of patients as binary variable will cause the problem of extremely high dimension and sparsity. To overcome that problem, we use the clinical classification software (CCS) developed by healthcare cost and utilization project (HCUP) to cluster diagnoses into manageable number of clinically meaningful categories [33]. We use single-level diagnosis classification scheme to achieve the

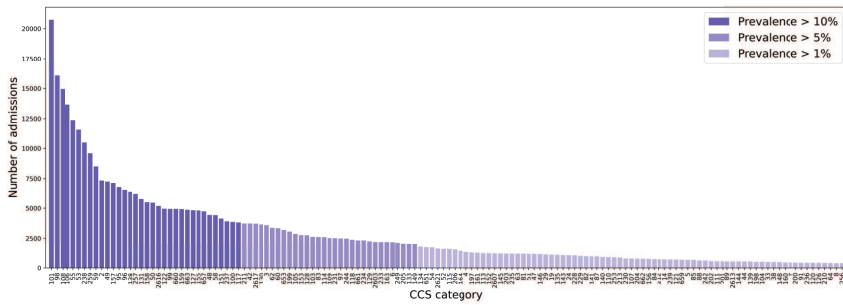
Table 1 Summary of statistical information of the overall patient cohort, survival patient cohort and patient cohort died in different time window. The median values [first quartile, third quartile] are presented for variables except for the number of admissions.

	Patient cohort						
	Overall	Survived	(, 30 days]	(30 days, 90 days]	(90 days, 180 days]	(180 days, 365 days]	(365 days, 5 years]
# admissions	18324	11672	1081	982	830	936	2823
Capillary refill rate	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.00]	[0.00, 1.00]	[0.00, 1.00]	[0.00, 0.00]	[0.00, 0.00]
Diastolic blood pressure	59.00	60.00	57.00	58.00	57.00	58.00	57.00
	[51.00, 69.00]	[52.00, 70.00]	[49.00, 67.00]	[49.00, 67.00]	[49.00, 67.00]	[49.00, 67.00]	[49.00, 67.00]
Fraction inspired oxygen	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	[0.40, 0.50]	[0.40, 0.50]	[0.40, 0.60]	[0.40, 0.60]	[0.40, 0.50]	[0.40, 0.60]	[0.40, 0.60]
Glasgow coma scale eye opening	4.00	4.00	4.00	4.00	4.00	4.00	4.00
	[3.00, 4.00]	[3.00, 4.00]	[3.00, 4.00]	[3.00, 4.00]	[3.00, 4.00]	[3.00, 4.00]	[3.00, 4.00]
Glasgow coma scale motor response	6.00	6.00	6.00	6.00	6.00	6.00	6.00
	[6.00, 6.00]	[6.00, 6.00]	[5.00, 6.00]	[5.00, 6.00]	[5.00, 6.00]	[6.00, 6.00]	[6.00, 6.00]
Glasgow coma scale total	14.00	15.00	13.00	14.00	14.00	14.00	14.00
	[10.00, 15.00]	[10.00, 15.00]	[9.00, 15.00]	[9.00, 15.00]	[10.00, 15.00]	[10.00, 15.00]	[10.00, 15.00]
Glasgow coma scale verbal response	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	[4.00, 5.00]	[4.00, 5.00]	[2.00, 5.00]	[1.00, 5.00]	[1.00, 5.00]	[2.00, 5.00]	[1.00, 5.00]
Glucose	128.00	128.00	128.00	128.00	128.00	129.00	129.00
	[106.00, 161.00]	[106.00, 159.00]	[104.00, 167.50]	[103.00, 165.00]	[103.00, 164.00]	[103.00, 166.00]	[105.00, 164.00]
Heart Rate	85.00	85.00	86.00	85.00	84.00	84.00	83.00
	[73.00, 97.00]	[74.00, 97.00]	[74.00, 100.00]	[73.00, 97.00]	[74.00, 97.00]	[72.00, 97.00]	[72.00, 95.00]
Height	170.00	170.00	166.50	170.00	170.00	165.00	168.00
	[163.00, 178.00]	[163.00, 178.00]	[157.00, 176.50]	[163.00, 178.00]	[163.00, 177.25]	[157.00, 176.50]	[160.00, 175.00]
Mean blood pressure	77.00	77.67	74.00	75.00	74.33	75.33	76.00
	[68.00, 87.67]	[68.67, 88.00]	[65.00, 85.00]	[66.00, 85.67]	[66.00, 85.00]	[66.67, 86.33]	[67.00, 87.00]
Oxygen saturation	98.00	98.00	98.00	98.00	98.00	98.00	98.00
	[96.00, 100.00]	[96.00, 100.00]	[96.00, 100.00]	[96.00, 100.00]	[96.00, 100.00]	[96.00, 100.00]	[96.00, 100.00]
Respiratory rate	18.00	18.00	19.00	19.00	19.00	19.00	18.00
	[15.00, 22.00]	[15.00, 22.00]	[16.00, 23.00]	[16.00, 23.00]	[15.00, 23.00]	[15.00, 22.00]	[15.00, 22.00]
Systolic blood pressure	118.00	118.00	114.00	116.00	115.00	117.00	118.00
	[104.00, 134.00]	[104.00, 134.00]	[100.00, 131.00]	[102.00, 133.00]	[102.00, 132.00]	[103.00, 134.00]	[104.00, 136.00]
Temperature	37.00	37.10	36.72	36.78	36.78	36.89	37.00
	[36.50, 37.56]	[36.56, 37.60]	[36.22, 37.28]	[36.28, 37.30]	[36.30, 37.30]	[36.39, 37.39]	[36.44, 37.50]
Weight	80.00	82.10	73.90	73.35	74.18	76.10	78.00
	[67.00, 96.00]	[68.95, 98.60]	[62.28, 88.05]	[61.77, 87.41]	[62.40, 89.03]	[62.87, 90.70]	[65.54, 92.30]
pH	7.38	7.38	7.37	7.37	7.36	7.37	7.37
	[7.32, 7.42]	[7.32, 7.42]	[7.30, 7.42]	[7.30, 7.43]	[7.30, 7.42]	[7.30, 7.42]	[7.31, 7.42]

clustering, which aggregates illness and conditions into 275 mutually exclusive categories. Most of the categories are clinically homogenous. The descending numbers of patients of CCS categories (which we deem as disease in the following part) with prevalence more than 1% in our dataset are shown in Figure 2 and the information of top 10 CCS categories with highest prevalence is presented in Table 2. The complete correspondence between ICD-9 codes and CCS categories are presented in Supplementary material. There are 34 diseases with prevalence more than 10%, 65 diseases with prevalence more than 5% and 140 diseases with prevalence more than 1%. Figure 3 shows top 20 diseases with highest prevalence in patient cohort who died in different time window and the prevalence of those diseases in survival patients in the dataset are also presented for comparison. It can be found that the prevalence of most diseases in died patients (red bar) are higher than survival patients (blue bar). From Figure 3 we can find that the categories of top 20 diseases are similar among the five mortality time windows but prevalence of diseases shows some differences. Figure 4 shows top 20 diseases (within the 140 diseases) with highest mortality rate in different time window. We can find that the patients who suffer from critical illness such as secondary malignancies (CCS 42), leukemias (CCS 39), cancer of bronchus and lung (CCS 19) and son on, are more likely to die in a year from discharge. The diseases of highest after-one-year mortality are quite different from those in-one-year. For example, diseases like other diseases of bladder and urethra (CCS 162), Parkinson's disease (CCS 79), nephritis, nephrosis, renal sclerosis (CCS 156) and so on, have a highest after-one-year mortality. These diseases are chronic diseases which are not so fatal in a short term but also affect the live quality of patients.

Table 2 Presentation of top 10 CCS categories with highest prevalence

CCS category	CCS category description	ICD-9 code	ICD-9 code description
101	Coronary atherosclerosis and other heart disease	411	Postmyocardial infarction syndrome
		⋮	⋮
		V45.82	Percutaneous transluminal coronary angioplasty status
98	Essential hypertension	401.1	Benign essential hypertension
		401.9	Unspecified essential hypertension
		398.91	Rheumatic heart failure (congestive)
108	Congestive heart failure; nonhypertensive	⋮	⋮
		428.9	Heart failure, unspecified
		427	Paroxysmal supraventricular tachycardia
106	Cardiac dysrhythmias	⋮	⋮
		785.1	Palpitations
		276	Hyperosmolality and/or hypernatremia
55	Fluid and electrolyte disorders	⋮	⋮
		995.1	Angioneurotic edema, not elsewhere classified
		272	Pure hypercholesterolemia
53	Disorders of lipid metabolism	⋮	⋮
		272.4	Other and unspecified hyperlipidemia
		276.61	Transfusion associated circulatory overload
238	Complications of surgical procedures or medical care	⋮	⋮
		V9009	Other retained radioactive fragments
		302	Ego-dystonic sexual orientation
259	Residual codes; unclassified	⋮	⋮
		V8909	Other suspected maternal and fetal condition not found
		280	Iron deficiency anemia secondary to blood loss (chronic)
59	Deficiency and other anemia	⋮	⋮
		285.9	Anemia, unspecified
		003.1	Salmonella septicemia
2	Septicemia (except in labor)	⋮	⋮
		995.92	Severe sepsis

**Fig. 2** Number of patients of each disease with different prevalence threshold

As patients with different life length suffer from different kinds of diseases and patients suffering from different kinds of diseases have different mortality rate, we consider taking diagnosis information as part of patient features for mortality prediction. After clustering the large number of ICD-9 codes with CCS, we can use a vector with binary entries to represent patient diseases, in which each entry represents whether a patient suffers from certain disease or not. In this study, we select diseases which meet some prevalence threshold, thus obtaining disease vectors of different length. The threshold of prevalence is set as 10%, 5% and 1%, the corresponding disease vectors of patient i are $\mathbf{x}_i^{dis(pre>10\%)} = (x_i^{dis_1}, x_i^{dis_2}, \dots, x_i^{dis_{34}})^T$, $\mathbf{x}_i^{dis(pre>5\%)} =$

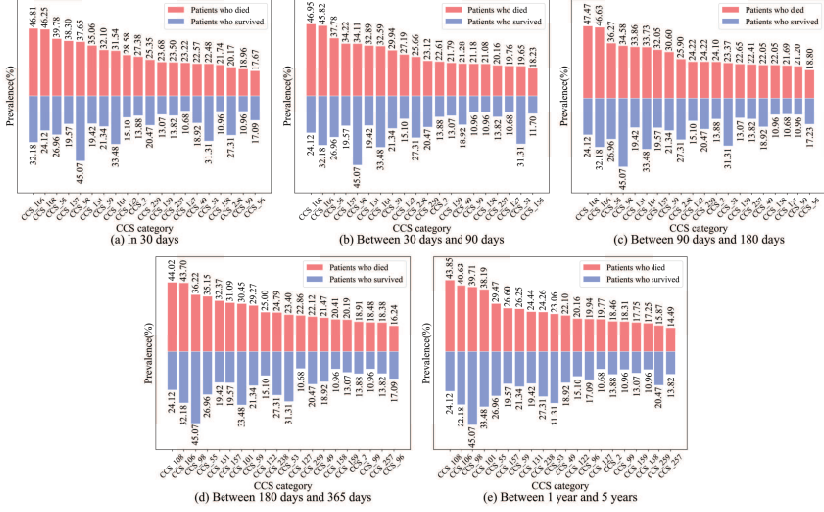


Fig. 3 Top 20 CCS categories with highest prevalence of patients died in different time window

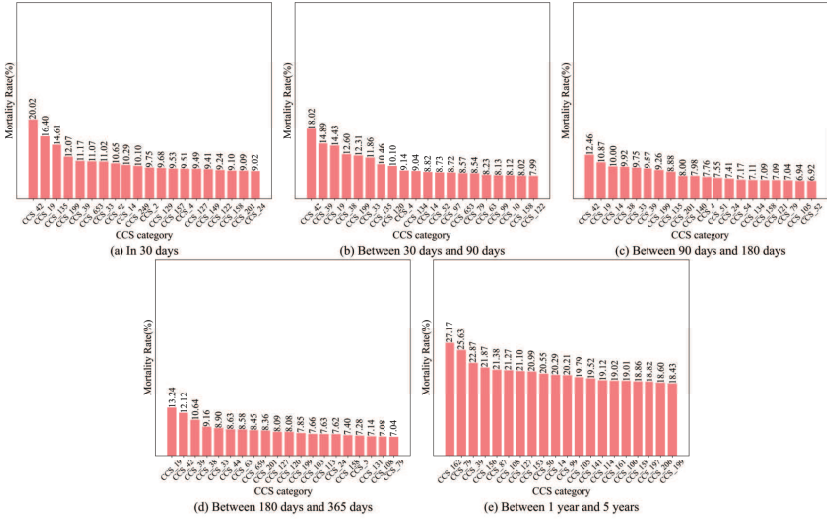


Fig. 4 Top 20 CCS categories with highest mortality rate in different time window

$(x_i^{dis_1}, x_i^{dis_2}, \dots, x_i^{dis_{65}})^T$ and $\mathbf{x}_i^{dis(pre>1\%)} = (x_i^{dis_1}, x_i^{dis_2}, \dots, x_i^{dis_{140}})^T$, in which each binary entry $x_i^{dis_j}$ represents whether a patient suffers from some disease dis_j or not.

2.2 Deep learning post-discharge mortality prediction method considering diagnosis information for ICU patients

2.2.1 Fusion of patient physiological data and diagnosis information

After the patient vital sign data preprocessing, each patient gets a vital sign multivariate time series matrix for deep learning mortality prediction models. And after the preprocessing of patient diagnosis information, each patient gets a binary vector representing the diseases he suffers from. In this study, we adopt an easy way to fuse the patient physiological data and diagnosis information data. We pad the same disease vector to each timestamp, thus obtaining three matrices $\mathbf{X}_i^{dis(pre>10\%)} = (x_{nt}^{(i)})_{34 \times 48}$, $\mathbf{X}_i^{dis(pre>5\%)} = (x_{nt}^{(i)})_{65 \times 48}$ and $\mathbf{X}_i^{dis(pre>1\%)} = (x_{nt}^{(i)})_{140 \times 48}$, which can be concatenated with vital signs multivariate time series matrices through time axis.

2.2.2 Mortality predicting deep learning models

Our proposed mortality prediction method is based on deep learning models. Recurrent neural network (RNN) and convolutional neural networks (CNN) are two outstanding types of deep learning models which have been widely applied to many clinical studies such as outcome prediction [15], readmission prediction [34], critical illness prediction [35] and disease progression prediction [36]. The former one is good at data with temporal properties and the latter is good at data with spatial properties. We implement two representative models long short-term memory (LSTM) [37] and temporal convolutional network (TCN) [38] which can both capture patterns in sequential data. For LSTM, the ability of capturing patterns in sequential data is achieved by recurrently updating the hidden layer state. The hidden layer state can be calculated with the following equations:

$$\mathbf{f}_t = \sigma(\mathbf{W}_{fh}\mathbf{h}_{t-1} + \mathbf{W}_{fx}\mathbf{x}_t + \mathbf{b}_f) \quad (1)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_{ih}\mathbf{h}_{t-1} + \mathbf{W}_{ix}\mathbf{x}_t + \mathbf{b}_i) \quad (2)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_{ch}\mathbf{h}_{t-1} + \mathbf{W}_{cx}\mathbf{x}_t + \mathbf{b}_c) \quad (3)$$

$$\mathbf{c}_t = \mathbf{f}_t \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad (4)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{oh}\mathbf{h}_{t-1} + \mathbf{W}_{ox}\mathbf{x}_t + \mathbf{b}_o) \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (6)$$

where \mathbf{x}_t is variable vector of \mathbf{X}_i at time point t . The initial $\mathbf{h}_0 = \mathbf{0}$. The σ (sigmoid) and \tanh functions are applied element-wise. The \mathbf{W} matrices and \mathbf{b} vectors are parameters determined by the error back propagation during the neural network training stage. The hidden state of timestamp t can be represented as $\mathbf{h}_t = LSTM(\mathbf{x}_t, \mathbf{h}_{t-1})$. The final hidden state \mathbf{h}_T can be transformed

into the mortality probability through a dense layer and sigmoid function. The architecture of our LSTM model is shown in Figure 5, which consists of 2 hidden layers with 256 units each. The dropout rate and recurrent drop out rate are set as 0.2 in both hidden layers. For TCN, its special architecture can take a sequence of any length and map it to an output sequence of the same length just as with an RNN. The basis of TCN is dilated convolution, which can increase the receptive field of model and even with fewer parameters. The detailed structure of our TCN is presented in Figure 6. For both LSTM and TCN, the loss function is binary cross entropy represented as follows:

$$L(y, \hat{y}) = -(y \cdot \log(\hat{y})) + (1 - y) \cdot \log(1 - \hat{y}) \quad (7)$$

Hyper-parameter estimation is conducted via a trial and error way. We optimize the parameters aiming at minimizing the prediction error. We start with initial range of coarse values, measure prediction error, and then adjust hyper-parameters in order to get the least loss on the validation dataset. Finally, the number of epochs is set as 100 (early stopping is also set based on validation loss with patience of 20 to avoid over-fitting), batch size is set as 64 and Adam (adaptive moment estimation) [39] is chosen as optimizer. And because the labels of all five tasks are im-balanced, we conduct over sampling for training data.

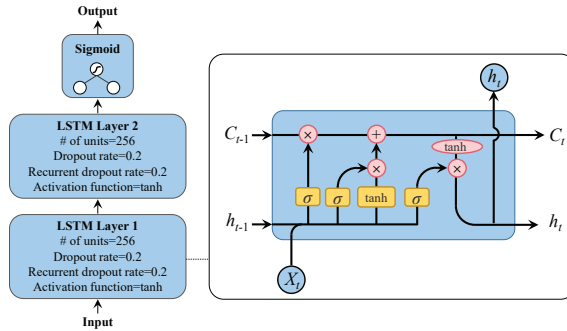


Fig. 5 Architecture of our long short-memory neural network

2.3 Evaluate the mortality risk brought by disease factors through disease Shapley values

To evaluate how diseases of a single patient increase his mortality, we compute disease Shapley values referring to a machine learning interpreting technique which uses the idea of Shapely values to measure how input values of features impact on the output of a model [40]. Shapley value is a concept coming from game theory which aims to allocate the coalition output in a game to all members according to their contributions in a most fair way [41]. By computing expectations of contributions over the member set of all possible permutations

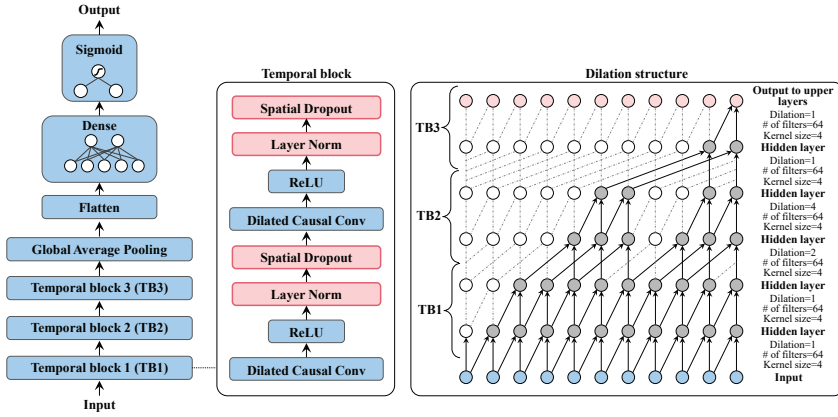


Fig. 6 Architecture of our temporal convolutional network

and taking all possible interactions between players into consideration, Shapley values can achieve a fair way of allocating the total output of coalition. In our study, Shapley value is used to evaluate the mortality risk in different time window brought by disease factors. Under this circumstance, the output of the coalition is the mortality of a single patient and the members in the coalition are diseases the patient suffering from. The disease Shapley values $\phi_i^j(f(\cdot), \mathbf{X}_i, d_i)$, which represent the impact of disease d_i^j for patient i , can be calculated with trained models $f(\cdot)$, the vital sign matrix \mathbf{X}_i and disease set d_i as follows:

$$\phi_i^j(f(\cdot), \mathbf{X}_i, d_i) = \sum_{d_i^{sub} \subseteq d_i \setminus d_i^j} \frac{|d_i^{sub}|!(|d_i| - |d_i^{sub}| - 1)!}{|d_i|!} [f(g(\mathbf{X}_i, \mathbf{x}_i^{j+})) - f(g(\mathbf{X}_i, \mathbf{x}_i^{j-}))] \quad (8)$$

the \mathbf{x}_i^{j+} and \mathbf{x}_i^{j-} are multi-hot disease vectors in which the entries are set as one if the corresponding diseases are in $d_i^{sub} \cup d_i^j$ and d_i^{sub} and the rest of entries are set as zero. The $g(\cdot)$ is the operation of padding \mathbf{x}_i^{j+} or \mathbf{x}_i^{j-} and concatenating them with \mathbf{X}_i . The disease Shapley values are calculated as expected contribution to the mortality of some disease over all permutations of the rest. To provide a more intuitive view of the cumulative effect of patient diseases, we also calculate the patient mortality in a way of generating cumulative disease sets. With patient vital sign multivariate time series and a series of cumulative disease sets in which the diseases are added according to the ascending disease Shapley values, a series of mortalities for a single patient can be calculated. The whole process of computing disease Shapley values and cumulative mortality is presented in Figure 7 by taking a dummy patient with disease set of $d_i = \{d_i^1, d_i^2, \dots, d_i^{m-1}, d_i^m\}$ as example. The number of diseases patient i suffers from is d_i (which is m in the example) and

d_i^1, d_i^2 are the second and the fifth diseases in the CCS disease category list, respectively. We can conduct a loop process to calculate each disease Shapley value for each single patient. In the figure, we only present the process in the condition that $j = 1$. When calculating the Shapley value of disease j , we first generate 2^{m-1} disease subsets from d_i in which d_i^j is excluded. Then we obtain 2^{m-1} corresponding subsets by adding d_i^j to each subset generated above. Then a series of disease vectors \mathbf{x}_i^{j-} and \mathbf{x}_i^{j+} can be obtained through the mapping from generated disease subsets. Finally, the Shapley value of disease j can be calculated with Eq. (7). By repeated above process, a disease Shapley value list $[\phi_i^1, \phi_i^2, \dots, \phi_i^{m-1}, \phi_i^m]$ can be obtained. According to the ascending disease Shapley values, a cumulative disease subset list $[d_i^{cum_0}, d_i^{cum_1}, \dots, d_i^{cum_{m-1}}, d_i^{cum_m}]$ is generated for outputting the cumulative mortality $[y_i^{cum_0}, y_i^{cum_1}, \dots, y_i^{cum_{m-1}}, y_i^{cum_m}]$ to present the cumulative effect of diseases to patient mortality.

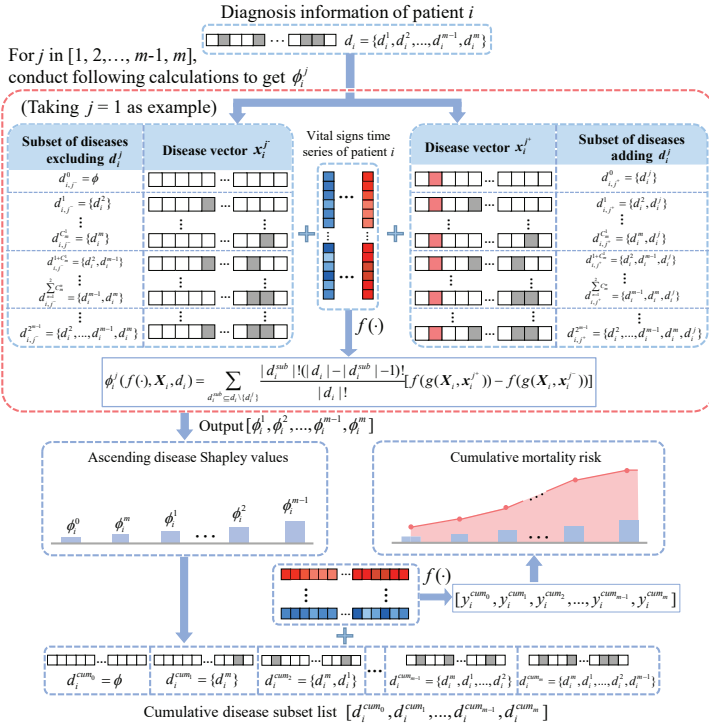


Fig. 7 Calculating process of disease Shapley values and cumulative mortality

3 Experimental results and analysis

In this section, we conduct some experiments on MIMIC-III dataset to investigate the effect of taking consideration of diagnosis information in patient

post-discharge mortality prediction. Besides, we take some patients with high mortality predicted by the model as example to present the hazard diseases bring to patient mortality.

3.1 Design of experiments

In this study, there are 5 tasks separately to predict whether a patient will die in 30 days, in 90days, in 180 days, in 365 days and in 5 years. Mainly two part of experiments are conducted on the several tasks.

Firstly, we compare the mortality prediction performance of traditional scoring systems, traditional machine learning models and deep learning models. Scoring systems implemented include SOFA and LODS. Traditional machine learning models including logistic regression (LR), support vector machine (SVM) [42], random forest (RF) [43] and XGBoost [44] are implemented. Those models can not take multivariate time series data as input, so we extract hand-engineered statistical features of patient vital signs. The statistical features including mean, standard deviation, maximum, minimum, skewness and kurtosis are calculated within 7 sub timeseries including full 48 hours, first 5 hours, first 12 hours, first 24 hours, last 24 hours, last 12 hours and last 5 hours. For each patient i , a vector $(x_{i,\text{mean}_1}, x_{i,\text{std}_1}, \dots, x_{i,\text{max}_n}, x_{i,\text{min}_n}, x_{i,\text{kurt}_n}, x_{i,\text{skew}_n})^T$ is obtained. The length of the vector is 714 (generated from 17 predictors, 7 sub timeseries and 6 statistical features) in our study. And the data in training set are also dealt with missing value imputation, standard normalization and over sampling.

Then, we present the results of deep learning models taking diagnosis information of different prevalence into consideration. To further prove the effect of adding diagnosis information for mortality prediction, we also implement traditional machine learning models taking diagnosis information into consideration. By changing the threshold of prevalence of diseases in our dataset, we can adjust the number of disease categories (namely the length of disease vector) added as patient features. In this study, we set the threshold values of prevalence as 10%, 5% and 1% which correspond to disease vectors of size (34, 1), (65, 1) and (140, 1).

To avoid accidental results of experiment, we conduct 5-fold cross-validation. The original dataset is divided into 5 partitions of roughly equal size and the ratio of positive and negative classes in each partition is roughly the same as original dataset. Random seeds are kept to eliminate the randomness brought by the splits of dataset. The overall performance is obtained by averaging the results of each fold, i.e.,

$$\bar{P} = \frac{1}{K} \sum_{k=1}^K P_K \quad (9)$$

where P_K is the performance metric, which in our study is area under the receiver operating characteristic curve (AUROC) score and area under

precision-recall curve (AUPRC) score, which are adopted to evaluate the discrimination ability of the model in the condition that data are imbalanced labeled. The higher the AUROC and the AUPRC are, the better the model is; K is the number of folds, which is 5 in our study.

Our experiments are conducted with Python 3.7.3. The logistic regression, support vector machine and random forest are implemented with scikit-learn library of 0.23. XGBoost is implemented with xgboost library of 1.3.3. The LSTM and TCN models are implemented with Keras 2.3.1 and Tensorflow 2.1.0.

3.2 Results and analysis

In this section, we first present the results of the several mortality prediction tasks with different models and different volume of diagnosis information. Then, several patients died in five time windows are shown to present the evaluation of a single patient mortality risk brought by disease factors.

3.2.1 Prediction performance comparison

The experimental result comparison of traditional scoring systems, traditional machine learning models and deep learning models are shown in Table 3. We present the mean value and standard deviation of AUROC and AUPRC of the 5 folds of each task. For each task, the highest two AUROC and AUPRC are marked in bold. From the table, we can find that most machine learning models have better performance than scoring systems. SVM shows weak performance in the 30-days, 90-days and 180-days mortality prediction, but performs well in the rest tasks. Random forest performs well in the 180-days, 1-years and 5-years mortality prediction. As for deep learning models, TCN performs well in 30-days and 90-days mortality prediction and LSTM achieves the best results except for 5-years mortality prediction.

The experimental results of traditional machine learning models and deep learning models using different volume of diagnosis information are compared in Table 4. Instead of marking the highest AUROC and AUPRC, we mark the highest AUROC and AUPRC of the best two models in bold. For example, the best models for 30-days mortality prediction task are TCN and LSTM and the highest AUROC of the two models are TCN and LSTM using diagnosis information of prevalence more than 5% and 10%, respectively. Besides, for each model, the best result are marked with an asterisk. For example, the best result of logistic regression on the task of 30-days mortality prediction is obtained when using diagnosis information of prevalence more than 1%. From the table, we can find that adding diagnosis information improves the AUROC and AUPRC for both traditional machine learning models and deep learning models. For 30-days mortality prediction task, the AUROC scores are improved up to about 9.74%, 6.56%, 4.22%, 6.66% and 8.67% for logistic regression, SVM, RF, XGBoost, TCN and LSTM by using disease information, respectively. There are also improvements for AUPRC scores for models

using diagnosis information. As for longer term post-discharge mortality prediction tasks, the effect of adding diagnosis information is more significant both in AUROC and AUPRC for all models, which means vital sign data from relatively short period of time can not reflect the physiological state of patients sufficiently in longer term, while the diseases patients suffer from can provide more information about it. Additionally, diagnosis information helps to improve the performance more for deep learning models than traditional machine learning models, which is probably due to the strong ability of dealing with the unstructured time series data and capturing potential non-linear relationships between information of patient vital signs and diseases of deep learning models. With the help of diagnosis information, deep learning models can outperform all traditional machine learning models including SVM and random forest which even has better performance than deep learning models without diagnosis information.

In general, our method can improve the post-discharge mortality prediction performance to a greater extent, which means patient diagnosis information can help to better evaluate the physiological state of patients. We believe methods focusing on designing better model structure and algorithm for mortality prediction are likely to achieve better prediction performance with the help diagnosis information.

3.2.2 Evaluation of mortality risk brought by disease factors

For each task, we take one fold of experiment using LSTM as predicting model and the length of disease vector is determined according to the AUROC scores on the validation set to conduct the evaluation of mortality risk brought by disease factors. For each task, we present three patients who are predicted with low mortality without diagnosis information and high mortality with added diagnosis information in Figure 8-Figure 12. All of the patients presented actually died in the corresponding time window.

In Figure 8-Figure 12, the bar plot represents the computed disease Shapley values, which reflect the expected change value in the predicted risk of the model when a patient suffers from a disease versus when he doesn't. The line chart is predicted mortality with generated patient disease vectors to reflect the cumulative effect of diseases. The y -value of each dot represents the cumulative mortality. It is predicted with the cumulative disease set which includes the corresponding disease of the dot and those on the left. In Figure 8, patient 1 and 2 suffer from acute illness such as acute myocardial infarction and acute and unspecified renal failure. the disease Shapley values of these diseases for these patients are high, which means the model predicts these patients of high mortality largely due to these diseases. As for patient 1~3, the predicted mortality is relatively low without diagnosis information, which means their vital signs don't show very bad physiological condition. The diagnosis information associated with vital signs reveal the high mortality risk of the patients. As for patient 13 in Figure 12, the predicting mortality is relatively high without the diagnosis information, which means his vital signs data reflects a not optimistic

Table 3 Experimental results comparison of traditional scoring systems, traditional machine learning models and deep learning models.

Methods	30-days mortality		90-days mortality		180-days mortality		1-year mortality		5-years mortality	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
SOFA	0.6079±0.0205	0.0936±0.0090	0.6047±0.0139	0.1657±0.0093	0.6109±0.0138	0.2255±0.0060	0.6118±0.0108	0.2903±0.0076	0.6191±0.0095	0.4737±0.0147
LODS	0.6559±0.0165	0.1053±0.0074	0.6505±0.0133	0.1880±0.0098	0.6517±0.0092	0.2559±0.0108	0.6514±0.0090	0.3295±0.0051	0.6532±0.0174	0.5089±0.0183
LR	0.6611±0.0215	0.1052±0.0088	0.6799±0.0089	0.2031±0.0096	0.6830±0.0108	0.2771±0.0115	0.6890±0.0098	0.3532±0.0128	0.7096±0.0097	0.5664±0.0096
SVM	0.6406±0.0157	0.0944±0.0033	0.6502±0.0087	0.1843±0.0085	0.6588±0.0150	0.2572±0.0113	0.6723±0.0064	0.3454±0.0102	0.7206±0.0103	0.5927±0.0108
RF	0.6771±0.0083	0.1162±0.0070	0.6854±0.0098	0.2147±0.0086	0.6979±0.0136	0.2970±0.0163	0.7076±0.0047	0.3784±0.0121	0.7235±0.0120	0.5929±0.0181
XGBoost	0.6785±0.0159	0.1158±0.0076	0.6921±0.0111	0.2186±0.0129	0.6898±0.0088	0.2849±0.0163	0.6897±0.0046	0.3637±0.0110	0.7059±0.0083	0.5684±0.0098
TCN	0.6876±0.0175	0.1219±0.0162	0.6969±0.0173	0.2207±0.0261	0.6965±0.0189	0.3027±0.0316	0.6952±0.0303	0.3591±0.0303	0.6959±0.0234	0.5584±0.0200
LSTM	0.7013±0.0248	0.1395±0.0205	0.7076±0.0140	0.2297±0.0124	0.6993±0.0112	0.2994±0.0148	0.7064±0.0189	0.3800±0.0224	0.7095±0.0167	0.5788±0.0164

Table 4 Experimental results of traditional machine learning models and deep learning models considering diagnosis information. The baseline models without diagnosis information and models with different length of disease vectors are compared. The $+m\%$ D means the threshold of prevalence of diseases. The highest AUROC and AUPRC of the best two models in bold and the best result of each model are marked with an asterisk

Methods	30-days mortality		90-days mortality		180-days mortality		1-year mortality		5-years mortality	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
LR	0.6611±0.0215	0.1052±0.0088	0.6799±0.0089	0.2031±0.0096	0.6830±0.0108	0.2771±0.0115	0.6890±0.0098	0.3532±0.0128	0.7096±0.0097	0.5664±0.0096
LR+10%D	0.7155±0.0200	0.1365±0.0141	0.7419±0.0084	0.2632±0.0148	0.7493±0.0078	0.3633±0.0182	0.7587±0.0061	0.4542±0.0058	0.7699±0.0063	0.6450±0.0061
LR+5%D	0.7202±0.0197	0.1402±0.0115	0.7477±0.0119	0.2677±0.0177	0.7550±0.0101	0.3689±0.0188	0.7623±0.0059	0.4571±0.0088	0.7751±0.0069	0.6508±0.0067
LR+1%D	0.7255±0.0186*	0.1443±0.0142*	0.7560±0.0091*	0.2768±0.0101*	0.7632±0.0071*	0.3762±0.0158*	0.7725±0.0055*	0.4693±0.0114*	0.7838±0.0071*	0.6618±0.0066*
SVM	0.6406±0.0157	0.0944±0.0033	0.6502±0.0087	0.1843±0.0085	0.6588±0.0150	0.2572±0.0113	0.6723±0.0064	0.3454±0.0102	0.7206±0.0103	0.5927±0.0108
SVM+10%D	0.6755±0.0153	0.1078±0.0054	0.6941±0.0069	0.2142±0.0085	0.7128±0.0130	0.3113±0.0164	0.7284±0.0067	0.4093±0.0123	0.7555±0.0087	0.6315±0.0116
SVM+5%D	0.6793±0.0161	0.1102±0.0050	0.6994±0.0059	0.2167±0.0087	0.7172±0.0128	0.3165±0.0164	0.7310±0.0068	0.4123±0.0129	0.7581±0.0088	0.6340±0.0117
SVM+1%D	0.6826±0.0161*	0.1120±0.0051*	0.7043±0.0062*	0.2209±0.0089*	0.7224±0.0121*	0.3229±0.0156*	0.7371±0.0065*	0.4213±0.0140*	0.7612±0.0088*	0.6377±0.0118*
RF	0.6771±0.0083	0.1162±0.0070	0.6854±0.0098	0.2147±0.0086	0.6979±0.0136	0.2970±0.0163	0.7076±0.0047	0.3784±0.0121	0.7235±0.0120	0.5929±0.0181
RF+10%D	0.7024±0.0256	0.1273±0.0182	0.7294±0.0103	0.2523±0.0106*	0.7417±0.0126	0.3393±0.0074	0.7493±0.0056	0.4266±0.0137	0.7551±0.0084	0.6252±0.0123
RF+5%D	0.7013±0.0210	0.1234±0.0111	0.7288±0.0090	0.2490±0.0057	0.7455±0.0147	0.3448±0.0206	0.7507±0.0061	0.4309±0.0116	0.7549±0.0086	0.6262±0.0080*
RF+1%D	0.7057±0.0243*	0.1286±0.0157*	0.7323±0.0102*	0.2487±0.0051	0.7481±0.0109*	0.3514±0.0151*	0.7556±0.0059*	0.4353±0.0218*	0.7557±0.0063*	0.6244±0.0075
XGBoost	0.6785±0.0159	0.1158±0.0076	0.6921±0.0111	0.2186±0.0129	0.6898±0.0088	0.2849±0.0163	0.6897±0.0046	0.3637±0.0110	0.7059±0.0083	0.5684±0.0098
XGBoost+10%D	0.7152±0.0204	0.1432±0.0185	0.7385±0.0058	0.2768±0.0162	0.7462±0.0136	0.3729±0.0208	0.7602±0.0089	0.4631±0.0105	0.7599±0.0044	0.6405±0.0085
XGBoost+5%D	0.7234±0.0167	0.1454±0.0180*	0.7447±0.0186	0.2720±0.0210	0.7534±0.0132	0.3793±0.0200	0.7621±0.0059*	0.4626±0.0166	0.7664±0.0082	0.6432±0.0117
XGBoost+1%D	0.7239±0.0208*	0.1424±0.0140	0.7484±0.0110*	0.2780±0.0178*	0.7536±0.0055*	0.3826±0.0120*	0.7598±0.0084	0.4649±0.0183*	0.7717±0.0050*	0.6545±0.0075*
TCN	0.6876±0.0175	0.1219±0.0162	0.6969±0.0173	0.2207±0.0261	0.6965±0.0189	0.3027±0.0316	0.6952±0.0303	0.3591±0.0303	0.6959±0.0234	0.5584±0.0200
TCN+10%D	0.7219±0.0214	0.1556±0.0215	0.7654±0.0088	0.2986±0.0148*	0.7837±0.0103*	0.3997±0.0235*	0.7821±0.0086	0.4787±0.0204	0.7859±0.0048	0.6702±0.0106
TCN+5%D	0.7334±0.0308*	0.1643±0.0259*	0.7611±0.0174	0.2805±0.0253	0.7768±0.0034	0.3891±0.0152	0.7827±0.0055	0.4792±0.0100	0.7890±0.0051	0.6714±0.0076
TCN+1%D	0.7189±0.0253	0.1574±0.0226	0.7717±0.0062*	0.2909±0.0253	0.7823±0.0058	0.3951±0.0107	0.7875±0.0096*	0.4832±0.0164*	0.7948±0.0058*	0.6782±0.0106*
LSTM	0.7013±0.0248	0.1395±0.0205	0.7076±0.0140	0.2297±0.0124	0.6993±0.0112	0.2994±0.0148	0.7064±0.0189	0.3800±0.0224	0.7095±0.0167	0.5788±0.0164
LSTM+10%D	0.7621±0.0132*	0.1775±0.0204*	0.7761±0.0115*	0.3032±0.0177	0.7815±0.0108	0.3962±0.0193	0.7832±0.0087	0.4818±0.0109	0.7890±0.0055	0.6730±0.0111
LSTM+5%D	0.7471±0.0155	0.1668±0.0139	0.7732±0.0067	0.2958±0.0110	0.7888±0.0086	0.4033±0.0207	0.7851±0.0101	0.4839±0.0124	0.7919±0.0055	0.6778±0.0142
LSTM+1%D	0.7513±0.0107	0.1666±0.0130	0.7751±0.0043	0.3077±0.0174*	0.7925±0.0090*	0.4116±0.0114*	0.7921±0.0099*	0.5048±0.0168*	0.7964±0.0077*	0.6861±0.0126*

condition. When considering the diseases he suffered from, the physiological status assessed by the model gets worse. With the aid of the disease Shapley values, we can learn which diseases bring high mortality risk to patients and get better assessment on patient physiological state, which might be helpful for therapeutic interventions.

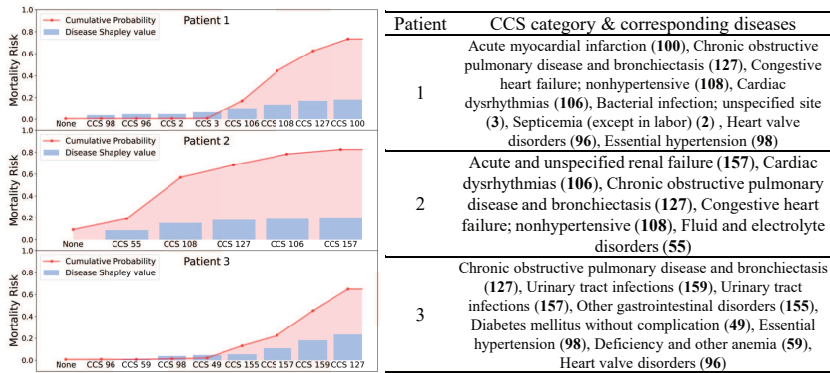


Fig. 8 Disease hazard evaluation examples of patients died in 30 days

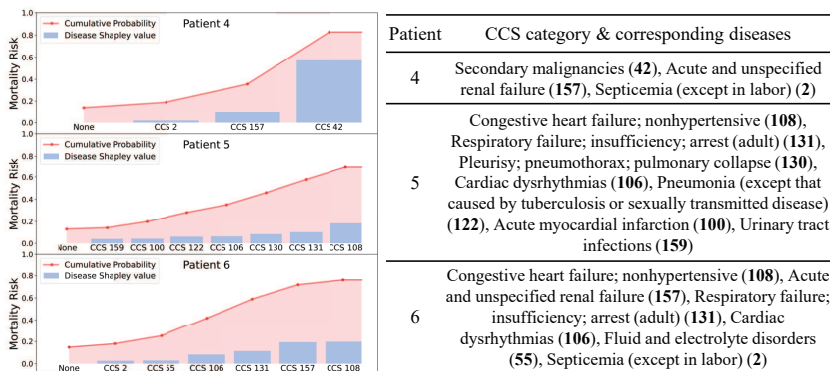


Fig. 9 Disease hazard evaluation examples of patients died between 30 days and 90 days

3.2.3 Discussion

A series of experiments are conducted and the results are presented in this section. We find that our proposed method can improve the mortality prediction performance tasks for both traditional machine learning models and deep learning models. The degree of prediction performance improvement varies with the tasks and model chosen. In general, the prediction performance improvement is more significant for deep learning models such as TCN and

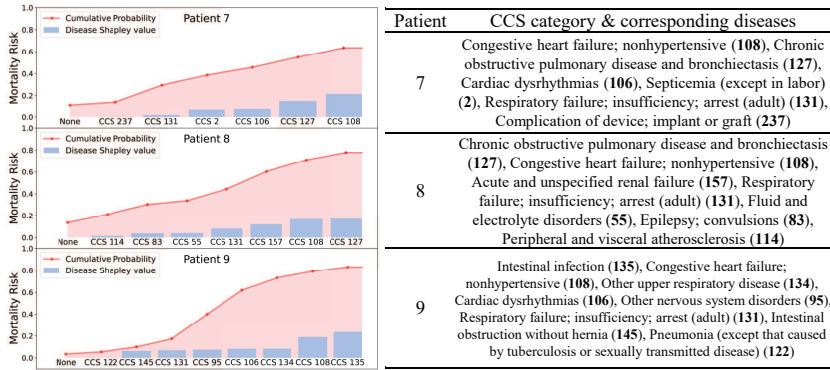


Fig. 10 Disease hazard evaluation examples of patients died between 90 days and 180 days

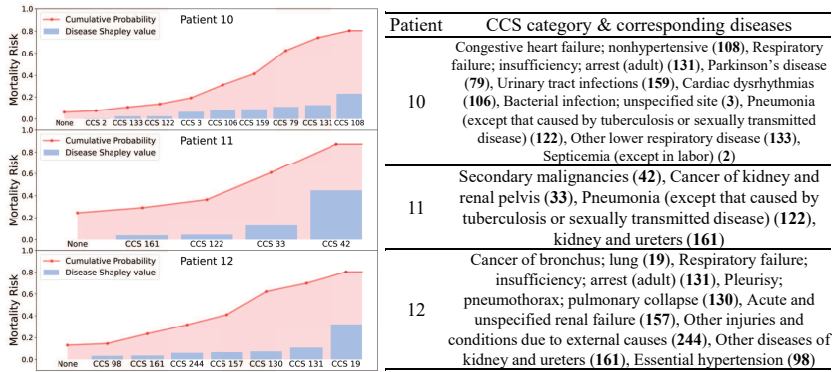


Fig. 11 Disease hazard evaluation examples of patients died between 180 days and 1 year

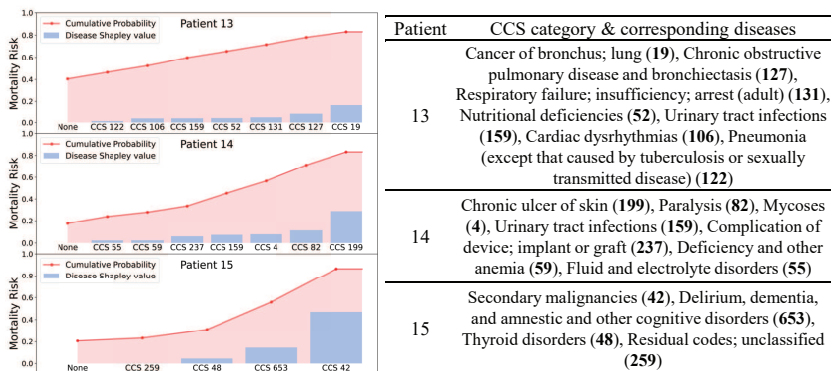


Fig. 12 Disease hazard evaluation examples of patients died between 1 year and 5 years

LSTM than traditional machine learning models such as logistic regression, SVM, random forest and XGBoost, which is probably because deep learning

models have stronger ability of capturing potential non-linear relationships between patient vital signs and disease information. The prediction performance improvement is obviously significant for longer term tasks such as 1-year and 5-years post-discharge mortality prediction, which is because the snapshot of patient vital signs are not enough to reflect his longer term physiological state while the diseases patients suffer from can provide more information about it. The experimental results are in line with clinical experience that the patient physiological state is closely related with the diseases he suffers from. In addition to using diagnosis information to predict the mortality of patients, we also provide a way of evaluating the hazard of diseases brought to a single patient mortality. By computing disease Shapley values, we can learn to what extent different diseases affect different patient's physiological state. As for the patient examples shown in our study, acute critical illness such as acute myocardial infarction, acute and unspecified renal failure and so on bring significant risk to patient mortality in one year. And illness such as secondary malignancies, various kinds of cancers also have serious bad effect on patients mortality. Although our method can not provide evaluation in a causal way, we can better predict post-discharge mortality and better evaluate the long term patient physiological state, which could help doctors to adopt therapeutic interventions and help patients to better know the condition of their health status.

4 Conclusions, limitations and future work

In this study, we propose a deep learning post-discharge mortality prediction method considering diagnosis information, which is tested on MIMIC-III dataset and proven to improve the performance in several mortality prediction tasks for deep learning models. This method also works for traditional machine learning models. Besides, we also provide a way of evaluating mortality risk brought by disease factors to a single patient by disease Shapley values. With the aid of the disease Shapley values, individual patient's disease condition can be analyzed to help doctors clarify the priority of multiple diseases and facilitate subsequent treatment.

There are some limitations and future work in our study. Firstly, we only represent the patient diagnosis information with a binary vector in a simplest way, which can not reflect the severity of patient diseases and relations between diseases. We will try to represent the diseases patients suffer from referring to methods which can better represent the relationships between diseases. With more detailed information about the severity of diseases patients suffer from, the physiological state of patients can be evaluated more accurately. Then, the mining of unstructured textual content of clinical notes of patients are not investigated in this study, which may further facilitate the post-discharge mortality prediction. Although there are some limitations of our research yet, we think this study is valuable for ICU clinical post-discharge mortality prediction and for the following up studies on ICU mortality prediction.

Acknowledgments. This research is supported by the National Natural Science Foundation of China (No. 71771034) and the Fundamental Research Funds for the Central Universities (No. DUT21YG108). The authors would also like to thank the MIT Laboratory for Computational Physiology and collaborating research groups for providing the freely available database (MIMIC-III).

Declarations

Conflict of interest. The authors declare that they have no conflict of interest in the conduct of this research.

Ethical approval. This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent. Informed consent was obtained from all individual participants included in the study.

Authorship Contributions

Mucan Liu: Conceptualization, Methodology, Software, Validation, Visualization, Writing - original draft. **Chonghui Guo:** Supervision, Writing - review & editing, Funding acquisition. **Shaoze Cui:** Conceptualization, Methodology, Writing - review & editing.

References

- [1] Power, G.S., Harrison, D.: Why try to predict ICU outcomes? Current opinion in critical care **20**, 544–549 (2014)
- [2] Suter, P., Armaganidis, A., Beaufils, F., Bonfill, X., Burchardi, H., Cook, D., Fagot-Largeault, A., Thijs, L., Vesconi, S., Williams, A., Gall, J., Chang, R.: Predicting outcome in ICU patients. Intensive Care Medicine **20**, 390–397 (1994)
- [3] Le Gall, J.-R., Klar, J., Lemeshow, S., Saulnier, F., Alberti, C., Artigas, A., Teres, D.: The logistic organ dysfunction system: a new way to assess organ dysfunction in the intensive care unit. JAMA: The Journal of the American Medical Association **276**, 802–810 (1996)
- [4] Vincent, J.-L., Moreno, R., Takala, J., Willatts, S., Artigas, A., Bruining, H., Reinhart, C.K., Suter, P., Thijs, L.G.: The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive Care Medicine **22**, 707–710 (1996)

- [5] Knaus, W., ZIMMERMAN, J., WAGNER, D., DRAPER, E., LAWRENCE, D.: APACHE-Acute physiology and chronic health evaluation: A physiologically based classification system. *Critical Care Medicine* **9**, 591–597 (1981)
- [6] Pirracchio, R., Petersen, M.L., Carone, M., Rigon, M.R., Chevret, S., van der Laan, M.J.: Mortality prediction in intensive care units with the super ICU learner algorithm (SICULA): a population-based study. *The Lancet Respiratory Medicine* **3**, 42–52 (2015)
- [7] Dybowski, R., Weller, P., Chang, R., Gant, V.: Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet* **347**, 1146–1150 (1996)
- [8] Kim, S., Kim, W., Park, R.W.: A comparison of intensive care unit mortality prediction models through the use of data mining techniques. *Healthcare informatics research* **17**, 232–243 (2011)
- [9] Krishnan, G., Kamath S, S.: A novel GA-ELM model for patient-specific mortality prediction over large-scale lab event data. *Applied Soft Computing* **80**, 525–533 (2019)
- [10] Guo, C., Lu, M., Chen, J.: An evaluation of time series summary statistics as features for clinical prediction tasks. *BMC Medical Informatics and Decision Making* **20**, 48 (2020)
- [11] Guo, C., Liu, M., Lu, M.: A dynamic ensemble learning algorithm based on k-means for ICU mortality prediction. *Applied Soft Computing* **103**, 107166 (2021)
- [12] Ding, Y., Wang, Y., Zhou, D.: Mortality prediction for ICU patients combining just-in-time learning and extreme learning machine. *Neuro-computing* **281**, 12–19 (2017)
- [13] Ghorbani, R., Ghousi, R., Makui, A., Atashi, A.: A new hybrid predictive model to predict the early mortality risk in intensive care units on a highly imbalanced dataset. *IEEE Access* **8**, 141066–141079 (2020)
- [14] Yann LeCun, Y.B., Hinton, G.: Deep learning. *Nature* **521**, 436–444 (2015)
- [15] Caicedo, W., Gutierrez, J.: ISeeU: Visually interpretable deep learning for mortality prediction inside the ICU. *Journal of Biomedical Informatics* **98**, 103269 (2019)
- [16] Purushotham, S., Meng, C., Che, Z., Liu, Y.: Benchmark of deep learning models on large healthcare MIMIC datasets. *Journal of Biomedical*

Informatics **83**, 112–134 (2017)

- [17] Thorsen-Meyer, H.-C., Nielsen, A., Nielsen, A., Kaas-Hansen, B., Toft, P., Schierbeck, J., Strøm, T., Chmura, P., Heimann, M., Dybdahl, L., Spangsege, L., Hulsén, P., Belling, K., Brunak, S., Perner, A.: 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**, 179–191 (2020)
- [18] Shickel, B., Loftus, T., Adhikari, L., Ozrazgat Baslanti, T., Bihorac, A., Rashidi, P.: DeepSOFA: A continuous acuity score for critically ill patients using clinically interpretable deep learning. *Scientific Reports* **9**, 1879 (2019)
- [19] Liu, L., Liu, Z., Wu, H., Wang, Z., Shen, J., Song, Y., Zhang, M.: Multi-task learning via adaptation to similar tasks for mortality prediction of diverse rare diseases. *CoRR* **abs/2004.05318v2** (2020) <https://arxiv.org/abs/2004.05318>. <https://doi.org/https://arxiv.org/abs/2004.05318v2>
- [20] Si, Y., Roberts, K.: Deep patient representation of clinical notes via multi-task learning for mortality prediction. *AMIA Joint Summits on Translational Science proceedings* **2019**, 779–788 (2019)
- [21] Azoulay, E., Adrie, C., De Lassence, A., Pochard, F., Moreau, D., Thiery, G., Cheval, C., Moine, P., Garrouste-Orgeas, M., Alberti, C., *et al.*: Determinants of postintensive care unit mortality: a prospective multicenter study. *Critical care medicine* **31**(2), 428–432 (2003)
- [22] Christiansen, C., Christensen, S., Johansen, M., Larsen, K., Tønnesen, E., Sørensen, H.: The impact of pre-admission morbidity level on 3-year mortality after intensive care: a danish cohort study. *Acta anaesthesiologica scandinavica* **55**, 962–970 (2011)
- [23] Grnarova, P., Schmidt, F., Hyland, S.L., Eickhoff, C.: Neural document embeddings for intensive care patient mortality prediction. *CoRR* **abs/1612.00467** (2019) <https://arxiv.org/abs/1612.00467>
- [24] Esper, A.M., Martin, G.S.: The impact of comorbid conditions on critical illness. *Critical care medicine* **39**(12), 2728–2735 (2011)
- [25] Nielsen, A., Thorsen-Meyer, H., Belling, K., Nielsen, A., Thomas, C., Chmura, P., Lademann, M., Moseley, P., Heimann, M., Dybdahl, L., *et al.*: Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the danish national patient registry and electronic patient records. *Lancet Digit Health* **1**, 78–89 (2019)

- [26] Dai, Z., Liu, S., Wu, J., Li, M., Liu, J., Li, K.: Analysis of adult disease characteristics and mortality on MIMIC-III. *PloS one* **15**, 0232176 (2020)
- [27] Braber, A., van Zanten, A.R.: Unravelling post-ICU mortality: predictors and causes of death. *European Journal of Anaesthesiology* **27**(5), 486–490 (2010)
- [28] Johnson, A., Pollard, T., Shen, L., Lehman, L.-w., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L., Mark, R.: MIMIC-III, a freely accessible critical care database. *Scientific Data* **3**, 160035 (2016)
- [29] Harutyunyan, H., Khachatrian, H., Kale, D., Galstyan, A.: Multitask learning and benchmarking with clinical time series data. *Scientific Data* **6**, 96 (2019)
- [30] Haapio, M., Helve, J., Grönhagen-Riska, C., Finne, P.: One- and 2-year mortality prediction for patients starting chronic dialysis. *Kidney International Reports* **2**, 1176–1185 (2017)
- [31] Fried, L., Kronmal, R., Newman, A., Bild, D., Mittelmark, M., Polak, J., Robbins, J., Gardin, J., Group, f.: Risk factors for 5-year mortality in older adults: The cardiovascular health study. *JAMA* **279**, 585–592 (1998)
- [32] de Rooij, S., Govers, A., Korevaar, J., Abu-Hanna, A., Levi, M., Jonge, E.: Short-term and long-term mortality in very elderly patients admitted to an intensive care unit. *Intensive care medicine* **32**, 1039–1044 (2006)
- [33] Elixhauser A, S.C., L, P.: Clinical classifications software (CCS). Agency for Healthcare Research and Quality (2015). <https://doi.org/https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>
- [34] Loreto, M., Lisboa, T., Moreira, V.P.: Early prediction of ICU readmissions using classification algorithms. *Computers in biology and medicine* **118**, 103636 (2020)
- [35] Lauritsen, S., Kristensen, M., Olsen, M., Larsen, M., Lauritsen, K., Jørgensen, M., Lange, J., Thiesson, B.: Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nature Communications* **11**, 3852 (2019)
- [36] Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J.: Doctor AI: Predicting clinical events via recurrent neural networks. In: *Machine Learning for Healthcare Conference*, pp. 301–318 (2016)
- [37] Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* **9**, 1735–1780 (1997)

- [38] Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modelings. CoRR **abs/1803.01271** (2018) <https://arxiv.org/abs/1803.01271>
- [39] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
- [40] Lundberg, S., Lee, S.-I.: A unified approach to interpreting model predictions share on. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768–4777 (2017)
- [41] Shapley, L.S.: A value for n-persons games. *Annals of Mathematics Studies* **28**, 307–318 (1953)
- [42] Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support vector machines. *IEEE Intelligent Systems and their applications* **13**(4), 18–28 (1998)
- [43] Breiman, L.: Random forests. *Machine learning* **45**(1), 5–32 (2001)
- [44] Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)