Journal of Biomedical Informatics 109 (2020) 103518

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Endpoint prediction of heart failure using electronic health records

Jiebin Chu[a], Wei Dong[b], Zhengxing Huang[a,*]

[a] College of Biomedical Engineering and Instrument Science, Zhejiang University, Key Lab for Biomedical Engineering of Ministry of Education, and Zhejiang Provincial Key Laboratory for Network Multimedia Technologies, China
[b] Department of Cardiology, Chinese PLA General Hospital, China

ARTICLE INFO

Keywords:
Endpoint prediction
Heart failure
Recurrent neural network
Generative adversarial network
Multi-task learning
Electronic health records

ABSTRACT

Background: Heart failure (HF) is a serious condition associated with high morbidity and mortality rates. Effective endpoint prediction in patient treatment trajectories provides preventative information about HF prognosis, guides decision making about the type and intensity of care, and enables better understanding of provider performance.
Objective: We explored the potential of a large volume of electronic health records (EHRs) for endpoint prediction of HF. Specifically, a suite of patient features observed at the prediction time point were utilized as the auxiliary information during the training of the prediction model.
Material and method: We extract the latent representation of patient treatment trajectory by equipping a recurrent neural network (RNN) with two learning strategies, namely adversarial learning and multi-task learning. As for the adversarial learning strategy, an adversarial learning scheme is used to differentiate the generated feature vector from the real one, while in the multi-task learning strategy, we consider the prediction of patient feature vector as an auxiliary task other than endpoint prediction. With such learning strategies, the extracted representation of patient treatment trajectory is particularly optimized for predicting HF endpoint, including HF-readmission, all-cause mortality and their combination (i.e., composite endpoint).
Results and discussion: We evaluate the proposed approach on a real clinical dataset collected from a Chinese hospital. The experimental dataset contains 2102 HF patient treatment trajectories with 13,545 visits on the hospital. The area under the ROC curve (AUC) achieved by our best model in predicting composite endpoint is 0.744, which is better than that of state-of-the-art models, including the standard Long Short Term Memory (0.727), Gated Recurrent Unit (0.732), RETAIN(0.730). With respect to the prediction of HF-readmission and all-cause mortality, our method also shows better performance than benchmark models.
Conclusion: The experimental results show that the proposed model can achieve competitive performance over state-of-the-art models in terms of endpoint prediction for HF, and reveal some suggestive hypotheses that could be validated by further investigations in the medical domain.

1. Background and significance

Heart failure (HF) is a complex and long-term cardiac circulatory disorder syndrome caused by dysfunction of heart, which pumps insufficiently to keep up with the demands on it. 26 million adults globally are diagnosed with HF, while 3.6 million are newly diagnosed every year [1]. 35% of patients suffering from HF die within the first year and the remaining die within 5 years [1]. In China, 0.9% of the population suffered from heart failure in 2000 [2], and the proportion increased rapidly to 1.3% in 2014 [1]. Patients with HF usually present a wide range of risk of two primary endpoints: unexpected HF-readmission after discharge, and all-cause mortality [3–5]. Therefore, accurate endpoint prediction of HF could play an essential role in supporting clinical decisions that allow precise prognosis and timely intervention for HF patients in their treatment trajectories [3,6–8].

In recent years, with the widespread adoption of electronic health records (EHRs), the information of visits on medical facilities, including physical examinations, lab test results, and medications, etc., along with the appropriate context, have been regularly documented in healthcare information systems. Apparently, it provides huge potential for HF endpoint prediction by utilizing a large volume of EHR data. Although valuable, many EHR-driven prediction models rely on the aggregate information captured at a certain time point, and less consider the sequential information in the entire patient treatment trajectory.

* Corresponding author.
E-mail address: zhengxinghuang@zju.edu.cn (Z. Huang).

https://doi.org/10.1016/j.jbi.2020.103518
Received 10 March 2020; Received in revised form 10 June 2020; Accepted 19 July 2020
Available online 25 July 2020
1532-0464/

To alleviate this problem, recurrent neural network (RNN), as a class of deep learning models exhibiting sequential data, has been applied on EHR data, and has shown great achievements in medical applications [9–11]. Specifically, two typical RNN variants, namely long short-term memory (LSTM) [12] and gated recurrent unit (GRU) [13], which are developed to hold the long-term dependency and solve the gradient vanishing and explosion problem, have been applied to many clinical prediction and prognosis tasks [9,14], with more promising performance than conventional machine learning models. Recently, context-aware embedding and attention-based mechanisms have been widely adopted in combined with deep neural networks to tackle sequential healthcare data. The context-aware embedding utilizes contextual information in sequences to enhance the latent representations, which finally benefit the prediction performance in medical tasks [36–39]. The attention mechanism is able to assign attentional signals to inputs, and thus not only makes the learning models focus on the essential parts, but also provides interpretability of models [39–41].

Although remarkable, there is still room for the further improvement on performance. We notice that both the endpoint and associated patient features at the time instant of the endpoint occurrence are faithfully documented in EHR, and thus are available in most cases. However, many previous studies only utilize the information of the endpoint occurrence as labels of training samples, while the associated patient features, as auxiliary information for learning, are not well leveraged in model training. It inevitably limits the sufficient utilization of the potential of EHR data.

In view of the above analysis, we proposed a deep adversarial learning model and a multi-task learning model for HF endpoint prediction, namely DAL-EP and MTL-EP, by equipping the standard RNN with the generative adversarial learning strategy and multi-task learning strategy, respectively. During the training process, the proposed models can utilize patient features recorded at the prediction time instant, as auxiliary information, to benefit RNN in extracting latent representations of patient treatment trajectories, and therefore improve the performance of HF endpoint prediction. In detail, the proposed DAL-EP has four main components: RNN, endpoint predictor, patient feature generator and discriminator. RNN is adopted to extract the latent representation from the past treatment trajectory of a HF patient, and the learned representation is then fed into the predictor to anticipate the endpoint occurrence at the next time instant of the HF patient treatment trajectory. Meanwhile, the extracted latent representation by RNN is also used to generate patient features at the prediction time instant. For DAL-EP, the generated feature vector is mixed with the actual feature vector of the patient and to be judged by the discriminator. The discriminative results are fed back to help RNN extract more representative features of patient treatment trajectory. As for MTL-EP, it has the same components with DAL-EP except for the discriminator, whose function is replaced by multi-task classification.

To our best knowledge, it is the first study to utilize a suite of patient features observed at the prediction time in the HF endpoint prediction problem. Specifically, we propose using two learning strategies, i.e., adversarial learning and multi-task learning, to incorporate the multiple auxiliary patient features into the model training, which benefits extracting latent representations of patient treatment trajectories, and thereafter improve the performance of HF endpoint prediction. The experimental results on a real clinical dataset demonstrate that our approach can achieve competitive performance over state-of-the-art models.

2. Materials and method

In this section, we first introduce the problem definition and notations used for concepts. After that, we present a detailed description of the proposed model.

2.1. Problem formulation

Formally, let $\mathscr{D} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_D\}$ be a clinical dataset consisting of D samples, and each $\boldsymbol{S} = \{\boldsymbol{X}, \boldsymbol{x}_{T+1}, y\}$ in $\mathscr{D}$ is a single patient sample, which consists of three components: (1) $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_T]$ denotes a patient treatment trajectory consisting of T visits, in which each $\boldsymbol{x}_t$ represents the feature vector documented at the tth time step (i.e., the tth visit), and T denotes the length of the past patient treatment trajectory. In addition, let $\boldsymbol{x}_t = [f_1, f_2, f_3, \cdots, f_F] \in \mathbb{R}^F$ be a patient feature vector, where f_i represents the value of the ith feature and F is the domain of patient features; (2) $\boldsymbol{x}_{T+1}$ represents the patient feature vector at the next time step $T + 1$; and (3) y denotes the occurrence of endpoint in the prediction window, $y \in \{0, 1\}$, where "1" indicates the occurrence of endpoint and "0" otherwise.

The objective of this study is to predict HF endpoint in a certain prediction window after the latest time step T, by using the information of the past patient treatment trajectory $\boldsymbol{X}$. We intend to solve the problem by the proposed models. The training process of the proposed models uses not only the endpoint occurrence as the label, but also the associated patient features at the next time step $\boldsymbol{x}_{T+1}$ as auxiliary information. The details of the proposed models are presented in the next sub-section.

2.2. Model description

In this study, we proposed two models to address the HF endpoint prediction problem by equipping the standard RNN with a generative adversarial learning strategy and multi-task learning strategy, respectively. As shown in Fig. 1, the proposed DAL-EP consists of four main components: RNN, endpoint predictor, generator and discriminator. MTL-EP model has the same first three components with DAL-EP, and replace the discriminator and its loss function with MSE.

2.2.1. Recurrent neural network

As illustrated in Fig. 1, we use RNN to tackle the sequential information of patient treatment trajectory. There are two common variants of RNN: LSTM and GRU. According to the experiment results (As can be seen in Section 3.2, GRU performs closely or better than LSTM), we take GRU cell as the RNN cell in our proposed models. The GRU transition equations are shown as follows:

$$\boldsymbol{r}_t = \sigma(W_r \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{1}$$

$$\boldsymbol{z}_t = \sigma(W_z \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{2}$$

$$\widetilde{h_t} = \backslash;\tanh(W_{\widetilde{h}} \cdot [r_t \odot h_{t-1}, x_t] \tag{3}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h_t} \tag{4}$$

where $\boldsymbol{x}_t$, $\boldsymbol{h}_t$ are the input and the hidden state. $\boldsymbol{r}_t$ and $\boldsymbol{z}_t$ are the values of reset gate and update gate, respectively. W_r, W_z and W_h are weight matrixes. σ represents the sigmoid activation function. The last hidden state, denoted as $\boldsymbol{h}_T$, is considered as the latent representation of patient treatment trajectory. We then input $\boldsymbol{h}_T$ into a classifier for HF endpoint prediction.

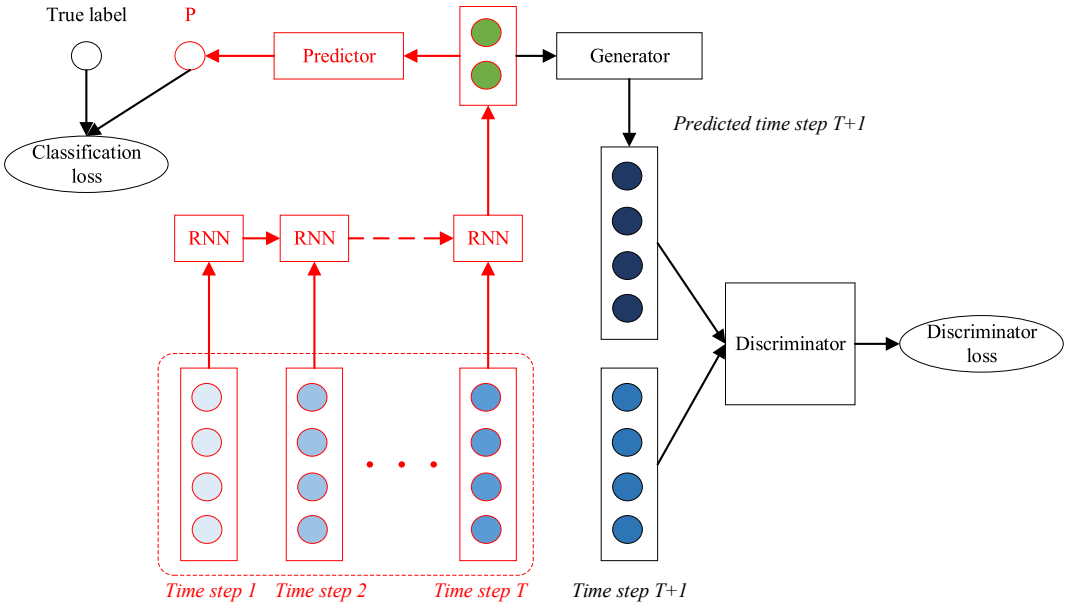
2.2.2. Endpoint predictor

In this study, we adopt a logistic regression as the classifier for HF endpoint prediction, which takes $\boldsymbol{h}_T$ as input and outputs the probability p of endpoint occurrence in a certain prediction window:

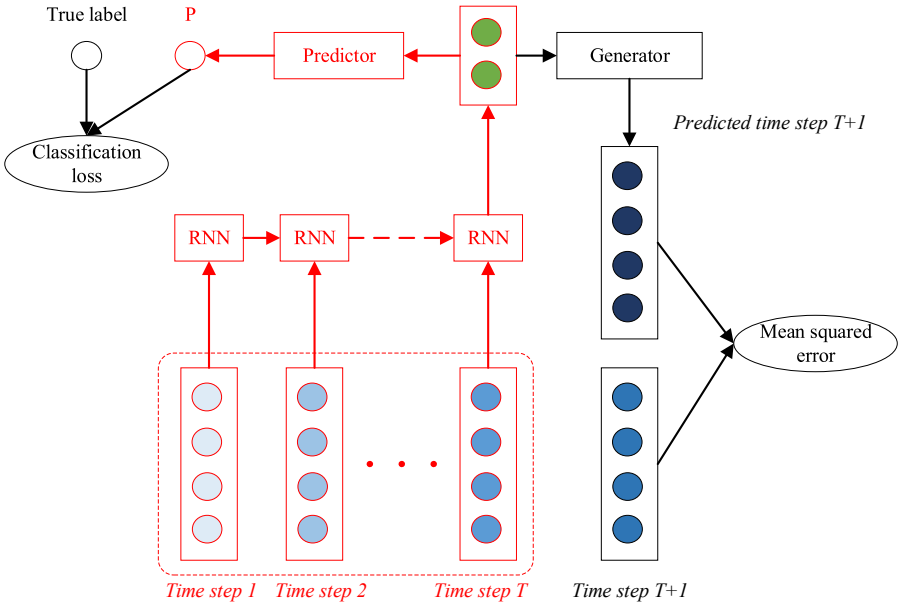
$$p = \sigma(W_{clf} \cdot \boldsymbol{h}_T + b_{clf}) \tag{5}$$

where W_{clf} and b_{clf} are the weight matrix and the bias, respectively. The classification loss is defined as the cross entropy:

$$Loss_{clf} = -\frac{1}{D} \sum_{i}^{D} (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \tag{6}$$



(A)



(B)

Fig. 1. (A) DAL-EP, which consists of four main components: 1) RNN: a neural network for the sequential data processing; 2) Endpoint predictor: a logistic regression to calculate the probability of endpoint occurrence; 3) Generator: a dense layer to predict the patient feature vector at the next time step based on the last hidden state of RNN; and 4) Discriminator: a logistic regression to judge whether the input future vector is real or fake. (B) MTL-EP, which has the same first-three components with DAL-EP, and replace the discriminator and its loss function with MSE. Note that the components in red color work both in the training and testing processes, while the components in black just play roles in the training process.

where i is the index of samples and D is the number of samples.

2.2.3. Generator

The generator takes $\boldsymbol{h}_T$ as input, and generates the fabricated feature vector at the next time step $T + 1$, denoted as $\boldsymbol{x}_{pred}$:

$$\boldsymbol{x}_{pred} = W_{pred} \cdot \boldsymbol{h}_T + \boldsymbol{b}_{pred} \tag{7}$$

where W_{pred} and b_{pred} are the weight matrix and the bias vector, respectively. In this case, the combination of the first three components is similar to a generator of generative adversarial network (GAN), which fabricates the fake patient feature vector to deceive the discriminator.

2.2.4. Discriminator

The discriminator is a logistic regression to judge whether the given feature vector is real or fake. It takes the generated feature vector $\boldsymbol{x}_{pred}$ and real one $\boldsymbol{x}_{T+1}$ as input, and calculates the probability that if $\boldsymbol{x}_{pred}$ is real or not.

$$p_{fake} = \sigma(W_{disc} \cdot \boldsymbol{x}_{perd} + b_{disc}) \tag{8}$$

$$p_{real} = \sigma(W_{disc} \cdot \boldsymbol{x}_{T+1} + b_{disc}) \tag{9}$$

where p_{fake_i} is the probability the discriminator gives for the ith fake sample, and the p_{reali} for the ith real sample, conversely. The loss

functions of the generator and discriminator are defined as follows:

$$Loss_{gen} = -\frac{1}{D}\sum_{i}^{D}(1\log(p_{fake_i}) + (1-1)\log(1-p_{fake_i})) = -\frac{1}{D}\sum_{i}^{D}\log(p_{fake_i}) \quad (10)$$

$$Loss_{disc} = \left(-\frac{1}{D}\sum_{i}^{D}(0\log(p_{fake_i}) + (1-0)\log(1-p_{fake_i}))\right) + \left(-\frac{1}{D}\sum_{i}^{D}(1\log(p_{real_i}) + (1-1)\log(1-p_{real_i}))\right) = -\frac{1}{D}\sum_{i}^{D}(\log(1-p_{fake_i}) + \log(p_{real_i})) \quad (11)$$

Eq. (10) is the cross entropy between the discriminative result of the generated feature vectors and a series of "1″ labels. The more probability the generated samples are judged as true samples, the smaller value the $Loss_{gen}$ is [15].

Similarly, Eq. (11) is the sum of the cross entropies between the discriminative results of the generated/real feature vectors and a series of binary labels, where "1″ represents real and "0" represents fake. The more accurate the discriminator judges, the smaller value the $Loss_{disc}$ is [15].

2.2.5. Training process for DAL-EP

Both the generator and its prepositive components constitute the network structure of DAL-EP, whose loss function is defined as follows:

$$Loss_{DAL} = \alpha_{DAL}Loss_{clf} + \beta_{DAL}Loss_{gen} + \lambda\,||\Theta||_2^2 \quad (12)$$

where α_{DAL} and β_{DAL} are the coefficients of the corresponding losses, Θ is the set of all trainable parameters in the network, $||\Theta||_2^2$ represents the L2-regulation term, and λ is the corresponding coefficient. Note that, $Loss_{DAL}$ and $Loss_{disc}$ iteratively optimized during adversarial learning process of DAL-EP.

2.2.6. Training process for MTL-EP

As for the multi-task learning strategy, we consider the prediction of the feature vector as an auxiliary classification task, and its loss function is defined as mean squared error (MSE):

$$Loss_{MSE} = \frac{1}{D\times F}\sum_{i}^{D}\sum_{j}^{F}(x_{pred_{i,j}} - x_{T+1i,j})^2 \quad (13)$$

The closer the predicted feature vectors with real ones, the smaller value the $Loss_{MSE}$ is.

The loss function for MTL-EP is defined as follows:

$$Loss_{MTL} = \alpha_{MTL}Loss_{clf} + \beta_{MTL}Loss_{MSE} + \lambda\,||\Theta||_2^2,$$

whereα_{MTL} and β_{MTL} are the coefficients of the corresponding losses, Θ is the set of all trainable parameters in the network, $||\Theta||_2^2$ represents the L2-regulation term, and λ is the corresponding coefficient.

2.3. Dataset

To validate the effectiveness of the proposed model, we conduct a case study on a real clinical dataset collected from the cardiology department of Chinese PLA General Hospital. The dataset contains EHRs of 25,402 HF patients with 49,155 visits on the hospital and distributed from 1994 to 2017. Note that we treat each visit as a time step in patient treatment trajectory, and the associated EHR at each visit contains various treatment information, including patient demographics, medication, operation, diagnosis, laboratory test result, vital signs, and physical examination results, etc., of a patient during his or her visit on the hospital.

Preprocessing was performed on the experimental dataset. Patients who were admitted to the hospital more than four times were included in the experimental dataset, to obtain the longitudinal data that can sufficiently depict their treatment trajectories. In addition, patient samples with more than 30% missing values were excluded in the analysis. Finally, the experimental dataset consists of 2102 patient treatment trajectories with 13,545 visits on the hospital.

Based on the suggestion from our clinical collaborators, 197 clinical features were selected to constitute the patient feature vector. Note that categorical features, including gender, discharge diagnosis, operation, medicine and complication, are recorded with discrete values, and continuous features, including age, length of stay, BMI and lab test values, are categorized into three levels: lower than normal, normal and higher than normal, according to the clinical protocol adopted by the hospital, and recorded as one-hot vectors with three dimensions. For example, the value of "serum albumin (ALB)" was separated into ALB_L, ALB_N and ALB_H, corresponding to ALB $<$ 40 g/L, 40–55 g/L and $>$ 55 g/L, respectively.

As for the length of patient treatment trajectory, we set the limitation for the maximum length and truncate the trajectory over the limit. Specifically, we denote the limited maximum length of the trajectory as N_{ptt}, and the length of each trajectory as n. The truncation strategy is in the following:

(1) When $n < N_{ptt}$, we selected the first $n-1$ visits as the input and filled the insufficient sequences with zero-padding. Then$n$th visit was taken as the true label for training.
(2) When $n \geq N_{ptt}$,we selected the first $N_{ptt}-1$ visits as input and the N_{ptt}th visit as the true label for training.

Fig. 2 shows the distribution of n, in which blue bars represent the numbers of total samples with the corresponding lengths of trajectories. Bars in grey, yellow, green and red represent the numbers of positive samples, when the length of prediction window (denoted as L_{pw}) is set to be three months, six months, twelve months and twenty-four months, respectively. Moreover, the light-colored bars correspond to the numbers of samples with HF-readmission, the deep-colored bars correspond to the numbers of samples with all-cause mortality, and the stacked bars of both correspond to numbers of samples with composite endpoint. As can be seen in Fig. 2, the trajectory length of over one third of samples is three. With the increases of n, the number of positive samples decreases rapidly. Moreover, by comparing the numbers of positive samples between different L_{pw}, it can be seen that the number of samples who had endpoint increases remarkably, with the extension of L_{pw}.

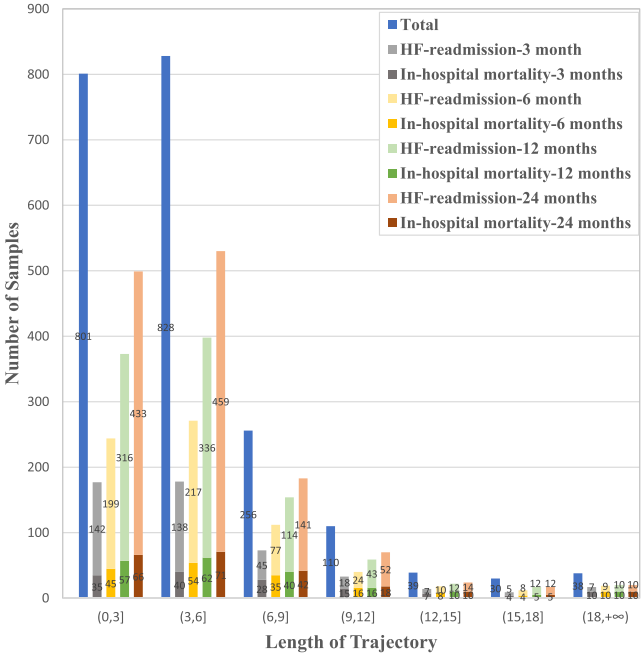


Fig. 2. Number of patient samples with the distribution of n.

Table 1
Basic characteristics of the experimental HF dataset.

General Profile		Value	
# of trajectories		2102	
# of visit		13,545	
Endpoint		Count	Ratio(%)
HF-readmission		355	16.89
All-cause mortality		129	6.14
Composite endpoint		484	23.03
Critical Clinical Feature		Count	Ratio(%)
Sex	Male	6643	59.26
Age	Age_L	807	7.20
	Age_N	2812	25.08
	Age_H	7591	67.72
BMI	BMI_L	643	5.74
	BMI_N	4047	36.10
	BMI_H	6520	58.16
NT-proBNP	NT-proBNP_L	0	0.00
	NT-proBNP_N	2409	21.49
	NT-proBNP_H	8801	78.51
LVEF	HFrEF	827	7.38
	HFmrEF	634	5.66
	HFpEF	9749	86.97
cTnT	cTnT_L	0	0.00
	cTnT_N	10,336	92.20
	cTnT_H	874	7.80
CRP	CRP_L	0	0.00
	CRP_N	8197	73.12
	CRP_H	3013	26.88
eGFR	eGFR_L	1724	15.38
	eGFR_N	4097	36.55
	eGFR_H	5389	48.07
SUA	SUA_L	62	0.55
	SUA_N	8814	78.63
	SUA_H	2334	20.82
ALB	ALB_L	2491	22.22
	ALB_N	8685	77.48
	ALB_H	34	0.30

(BMI, body mass index; NT-proBNP, N-terminal pro-brain natriuretic peptide; LVEF, left ventricular ejection fraction; HFrEF, heart failure with reduced left ventricular ejection fraction (LVEF $\leq$ 40%); HFmrEF, heart failure with mid-range left ventricular ejection fraction (40% < LVEF < 50%); HFpEF, heart failure with preserved left ventricular ejection fraction (LVEF $\geq$ 50%); cTnT, cardiac troponin T; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; SUA, serum uric acid; ALB, serum albumin).

When N_{ptt} is set to be 18, and L_{pw} is set to be three months, of 2102 patient treatment trajectories, 355 (16.89%) HF readmission, 129 (6.14%) all-cause mortality, and 484 (23.03%) composite endpoints occurred. Table 1 lists the basic characteristics of the experimental dataset.

3. Experiment

3.1. Experiment details

3.1.1. Benchmark models

To evaluate the performance of the proposed model, we selected several benchmark models for comparison:

- **LR:** is a Logistic Regression model consisting of $F_{flat} = N_{ptt} * F$ input units and one output value with the sigmoid function. We made the 2-dimensional sequential features $\boldsymbol{X} \in \mathbb{R}^{N_{ptt} \times F}$ flattened to be the one-dimensional input $\boldsymbol{X_{flat}} \in \mathbb{R}^{F_{flat}}$ of LR.
- **RF:** is a Random Forest model with 300 estimators, which l used the flattened input as the same with LR.
- **XGBOOST:** is an eXtreme Gradient Boosting model with 300 estimators, which used the flattened input as the same with LR.
- **SVM**: is a Support Vector Machine model with Radial Basis Function (RBF) as kernel function, which used the flattened input as the same with LR.
- **DAE:** is a combination of both the denoising auto encoder and logistic regression classifier. The deep representation of patient treatment trajectory was learned by DAE from EHR data. Then, the extracted latent representation was fed into the logistic regression for HF endpoint prediction.
- **LSTM:** consists of a basic LSTM and a logistic regression classifier, without the external components mentioned in the proposed model.
- **GRU:** consists of a basic GRU and a logistic regression classifier, without the external components mentioned in the proposed model.
- **RETAIN:** is a highly cited RNN-based predictive model using Reverse Time Attention Mechanism proposed by Choi et al [40].

As for the training of the LR, DAE, LSTM and GRU, we defined their loss functions as the cross entropy, and attached the L2-regulation term on the loss functions to alleviate overfitting. They were optimized by the AdamW optimizer [16] and validated by the five-cross validation strategy. Hyper-parameters of models were selected by using the grid search strategy.

3.1.2. Evaluation metrics and hyper-parameter search

Endpoint prediction is one of the specific EHR-data driven prediction tasks. As indicated in many previous work [17,18], Receiver Operating Characteristic (ROC) and Area Under the ROC Curve (AUC) are widely used for evaluating the performance of clinical prediction tasks. Therefore, we choose ROC and AUC to evaluate the performance of HF endpoint prediction in the experiments. As for the hyper-parameter setting, we adopted the grid search strategy to find out the best values and evaluated the mean AUC performance by using the five-fold cross-validation strategy.

3.1.3. Experimental setup

All of the experiments were performed on a PC running on Microsoft Windows 10 with an Intel Core IV CPU 3.4 GHz and 16 Gigabytes of main memory. The algorithms were implemented using Python 3.7 and the source code of our model can be downloaded from https://github.com/ZJU-BMI/DAL-EP. Prior approval was obtained from the data protection committee of the hospital to conduct the study, and the patient data were anonymized in this study and in this paper.

3.2. Result

To evaluate the effectiveness of the proposed model, we conducted experiments on a real clinical dataset. By default, the prediction window L_{pw} is set to be three months and the maximum length of patient treatment trajectory N_{ptt} is set to be 18, if not specified.

Table 2 shows the performance of HF endpoint prediction using both the proposed model and benchmarks.

Fig. 3 shows the achieved ROC curves of both the proposed models and benchmarks on HF endpoint prediction. The first column (i.e., a.1, b.1, and c.1) of Fig. 3 shows ROC curves of all models for three endpoint prediction cases. To evaluate the efficiency of equipping RNN with the generative adversarial learning strategy or multi-task learning strategy, we plot ROC curves of the proposed DAL-EP vs the standard GRU and RETAIN, the proposed MTL-EP vs the standard GRU and RETAIN, in the second and third columns, respectively.

3.3. Statistical analysis

To investigate if there are significant differences on the performance of HF endpoint prediction between the proposed models and

Table 2
The achieved AUC values of the proposed models and benchmarks in three cases. The best metric values are shown in bold.

Models	Composite endpoint	HF-readmission	All-cause mortality
DAL-EP	**0.744 ± 0.002**	**0.661 ± 0.005**	**0.867 ± 0.004**
MTL-EP	0.742 ± 0.002	0.658 ± 0.006	0.866 ± 0.006
GRU	0.732 ± 0.002	0.643 ± 0.006	0.858 ± 0.003
LSTM	0.727 ± 0.003	0.637 ± 0.006	0.857 ± 0.004
DAE	0.715 ± 0.002	0.627 ± 0.005	0.831 ± 0.002
LR	0.711 ± 0.001	0.622 ± 0.002	0.810 ± 0.002
RF	0.714 ± 0.002	0.639 ± 0.003	0.822 ± 0.003
XGBOOST	0.719 ± 0.002	0.639 ± 0.006	0.833 ± 0.005
SVM	0.695 ± 0.003	0.615 ± 0.009	0.813 ± 0.011
RETAIN	0.730 ± 0.006	0.644 ± 0.005	0.841 ± 0.005

benchmarks, we conducted the paired *t*-test on the achieved AUC values of model pairs. The calculated P-values are presented in Table 3.

3.4. Convergence analysis

To illustrate the efficiency of the proposed model, we compared the convergences of models. Fig. 4 shows the variation tendency of both AUC and loss in the training process, with respect to the case of composite endpoint prediction. For the sake of fairness, we took the classification loss as the loss of DAL-EP, MTL-EP, as the same with GRU and RETAIN.

3.5. The impacts of N_{ptt} and L_{pw} on the performance tendency

To investigate the impact of changing N_{ptt} and L_{pw} on the variation tendency of the performance of HF endpoint prediction, we set one of them by default, and conducted the experiments on composite endpoint prediction by changing the value of the other one. Fig. 5 (a) shows the achieved AUC values of models for varying N_{ptt} when L_{pw} is set to be three months, and Fig. 5(b) shows the achieved AUC values of models for varying L_{pw} when N_{ptt} is set to be 18.

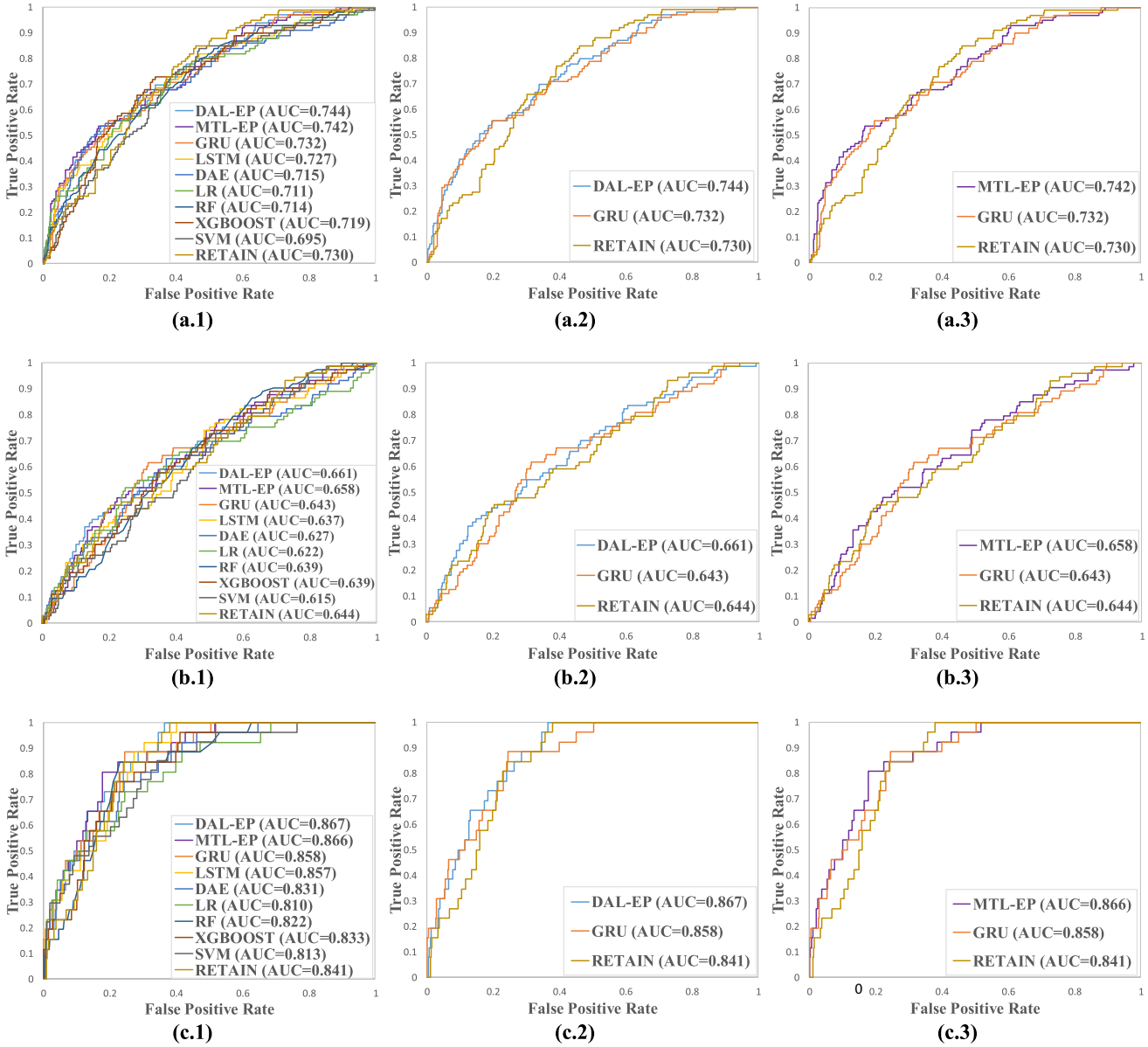


Fig. 3. The obtained ROC curves of the proposed models and benchmarks on the experimental HF dataset. (a) Composite endpoint, (b) HF-readmission, and (c) All-cause mortality.

Table 3
The paired *t*-test results between the proposed models and benchmarks in three HF endpoint prediction cases.

		DAL-EP	MTL-EP	LSTM	GRU	DAE	LR	RF	XGBOOST	SVM	RETAIN
Composite endpoint	DAL-EP	–	9.09E-02$^{\cdot}$	1.75E-10***	1.48E-11***	3.18E-18***	1.12E-14***	8.39E-17***	4.90E-17***	2.51E-20***	2.97E-05***
	MTL-EP	–	–	4.80E-09***	1.60E-10***	2.79E-17***	1.32E-13***	5.13E-16***	4.92E-16***	9.53E-20***	9.37E-05***
	GRU	–	–	–	4.17E-04***	2.51E-14***	1.68E-12***	5.51E-13***	2.29E-12***	3.20E-18***	3.78E-01$^{\cdot}$
	LSTM	–	–	–	–	3.04E-10***	5.01E-09***	1.68E-09***	9.57E-08***	5.94E-16***	1.86E-01$^{\cdot}$
	DAE	–	–	–	–	–	4.02E-06***	7.22E-01$^{\cdot}$	2.91E-05***	4.28E-14***	1.12E-05***
	LR	–	–	–	–	–	–	1.62E-03***	1.06E-10***	1.16E-09***	2.74E-06***
	RF	–	–	–	–	–	–	–	1.89E-04***	7.85E-13***	5.71E-06***
	XGBOOST	–	–	–	–	–	–	–	–	2.22E-15***	1.24E-04***
	SVM	–	–	–	–	–	–	–	–	–	7.42E-10***
	RETAIN	–	–	–	–	–	–	–	–	–	–
HF-readmission	DAL-EP	–	2.74E-01$^{\cdot}$	9.09E-07***	8.77E-09***	4.53E-12***	2.04E-11***	1.64E-10***	5.93E-09***	2.14E-11***	6.56E-07***
	MTL-EP	–	–	2.65E-05***	2.57E-07***	1.43E-10***	1.97E-09***	5.28E-07***	2.20E-07***	1.68E-10***	2.58E-05***
	GRU	–	–	–	4.17E-02*	3.07E-06***	7.98E-07***	7.44E-02$^{\cdot}$	1.19E-01$^{\cdot}$	1.67E-07***	7.68E-01$^{\cdot}$
	LSTM	–	–	–	–	3.23E-04***	1.07E-05***	4.07E-01$^{\cdot}$	4.96E-01$^{\cdot}$	3.47E-06***	1.71E-02*
	DAE	–	–	–	–	–	3.06E-02*	1.56E-06***	3.07E-05***	2.52E-03***	6.89E-07***
	LR	–	–	–	–	–	–	4.08E-12***	2.65E-07***	3.29E-02*	1.41E-07***
	RF	–	–	–	–	–	–	–	9.94E-01$^{\cdot}$	6.36E-06***	2.44E-02*
	XGBOOST	–	–	–	–	–	–	–	–	2.62E-07***	5.54E-02$^{\cdot}$
	SVM	–	–	–	–	–	–	–	–	–	6.70E-08***
	RETAIN	–	–	–	–	–	–	–	–	–	–
All-cause mortality	DAL-EP	–	8.23E-01$^{\cdot}$	1.03E-04***	9.94E-05***	5.69E-12***	2.72E-18***	1.05E-15***	2.56E-12***	5.01E-09***	6.15E-10***
	MTL-EP	–	–	2.25E-03***	1.51E-03***	4.14E-09***	1.82E-11***	1.40E-13***	1.03E-10***	7.41E-11***	1.55E-08***
	GRU	–	–	–	5.86E-01$^{\cdot}$	2.75E-14***	2.38E-18***	4.25E-15***	7.59E-11***	7.45E-08***	9.27E-08***
	LSTM	–	–	–	–	2.69E-10***	7.26E-17***	6.82E-14***	6.28E-10***	4.81E-08***	6.46E-07***
	DAE	–	–	–	–	–	4.04E-14***	8.51E-07***	2.64E-01$^{\cdot}$	5.31E-04***	1.81E-04***
	LR	–	–	–	–	–	–	5.73E-08***	7.20E-09***	3.68E-01$^{\cdot}$	1.24E-09***
	RF	–	–	–	–	–	–	–	1.26E-05***	4.07E-02*	2.13E-08***
	XGBOOST	–	–	–	–	–	–	–	–	1.79E-04***	2.77E-03***
	SVM	–	–	–	–	–	–	–	–	–	9.86E-07***
	RETAIN	–	–	–	–	–	–	–	–	–	–

**: P < 0.01.
*** : P < 0.005.
* : P < 0.05.
$^{\cdot}$:P ≥ 0.05.

3.6. Error analysis

To analyze potential patient features that cause wrongly predictions, we analyze the prediction results on composite endpoint prediction. Specifically, we categorize samples into four groups, namely true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Table 4 presents the top 30-ranked features with their frequencies in the decreasing order for each group. We argue that if the frequencies of a particular feature are significantly variable between in different groups, the feature may be nontrivial for the HF endpoint prediction task.

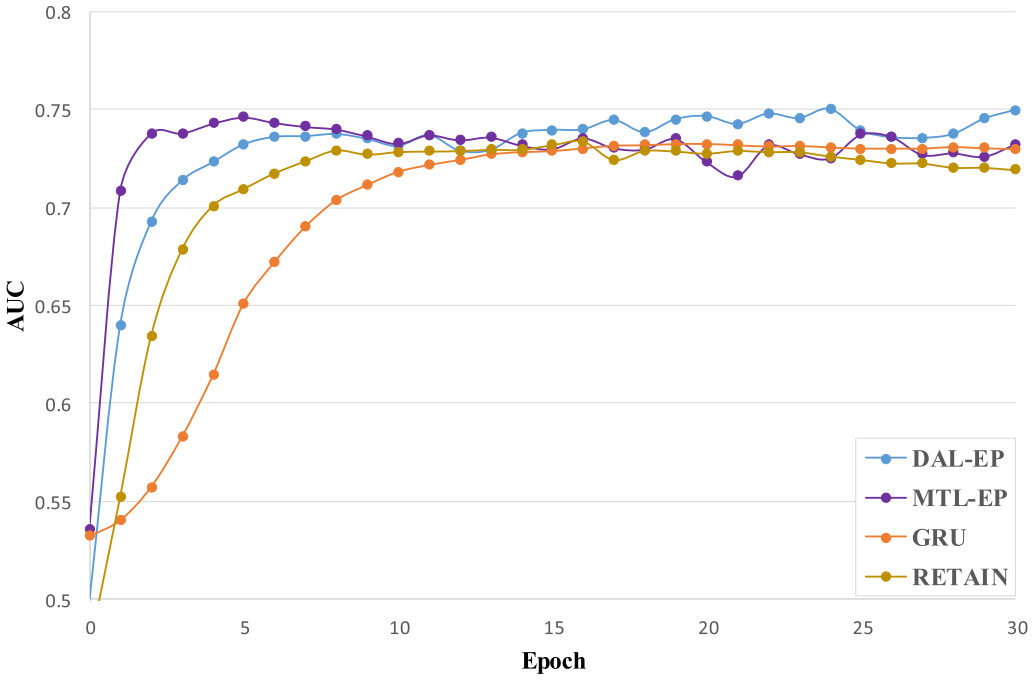
As illustrated in Table 4, although most of the features listed in four groups are coincident to each other, parts of them have varying frequencies within different groups. Taking "HFpEF" for example, its frequencies in both TP (0.820) and FP (0.835) are obviously lower than its values in TN (0.911) and FN (0.921) groups. This finding indicates that our model tends to judge the patient with high LVEF value as negative. Note that LVEF is one of the most important features to indicate the cardiac function. "HFpEF" represents ejection fraction > 50% and HF patients with high LVEF are considered to have better prognosis, as illustrated in many literature [19–21].

Besides, "LDH_N", "D-BIL_N" and "ALP_N" have variable frequencies within different groups. As can be seen in Table 4, "LDH_N", "D-BIL_N" and "ALP_N" have the lowest and highest frequencies in TP and TN, respectively. Note that these three features are essential indicators for liver function. As pointed out in clinical literature [22,23], liver function abnormalities are frequently observed in HF patients, it is therefore no wonder that these features were less appearing with normal values.

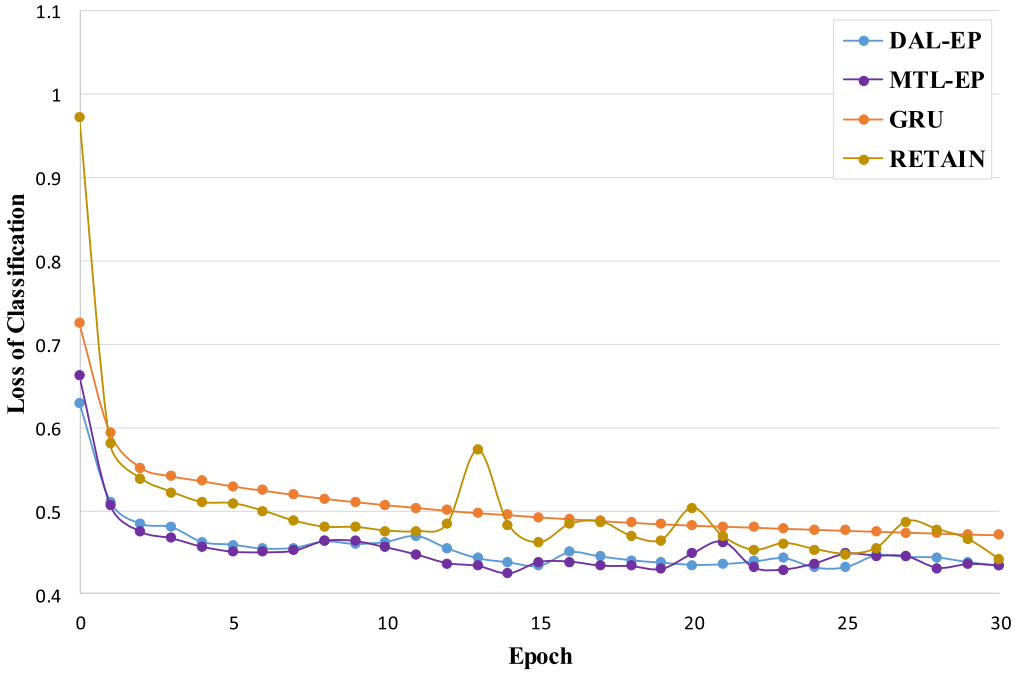
Despite the overlapping features across the four groups, there are some features only listed in parts of groups. For example, "NT-proBNP_H" is presented in TP and FP groups, indicating that the abnormally high value of "NT-proBNP" is easy to induce our model to make a positive judgement of endpoint prediction. Note that this finding fits with the clinical commonsense. As a vital indicator in evaluating the condition of HF patients, "NT-proBNP" is produced when the cardiomyocytes are stimulated by pressure or pull, the ventricular volume is expanded and the pressure load is increased [24,25]. In this sense, the high value of "NT-proBNP" generally indicates the poor prognosis of HF patients [26–28].

In addition, "D-dimer" is worthy of analysis. As can be seen in Table 4, "D-dimer_H" is listed in both TP and FP groups. This finding indicates that a HF patient with the abnormal value of D-dimer would has a very high probability to be recognized as the positive sample. From a clinical point of view, "D-dimer" is the product of plasma fibrin degradation, and reflects the condition of blood clotting [29]. We speculate that the value of "D-dimer" could be caused by the unstable blood flow in the end stage of HF. However, at present, there are few studies on investigating the correlations between "D-dimer" and HF. As interested by our clinical collaborators, this finding is worth for the further clinical investigation.

Moreover, serval feature related to anemia, such as "RDW-CV_N", "RBC_N" and "HGB_N", are only listed in TN or FN groups. It seems that these features with normal values are more frequent in TN and FN groups than that in TP and FP groups. As for the abnormal values of these features, "RDW-CV_H", "RBC_L" and "HGB_L" are essential indicators suggesting the symptom of anemia, in which all of them have higher frequency in TP and FP (as shown in Supplementary Table 1). In clinical practice, anemia is one of the most common complications of HF, and can either be the inducement of HF or be caused by HF [30–32]. Thus, it is no wonder that the anemia-related features with



(a)



(b)

Fig. 4. Variation tendency of AUC and loss in the training process for composite endpoint prediction. (a) Convergence comparison on AUC among DAL-EP, MTL-EP, GRU and RETAIN; (b) Convergence comparison on loss among DAL-EP, MTL-EP, GRU and RETAIN.

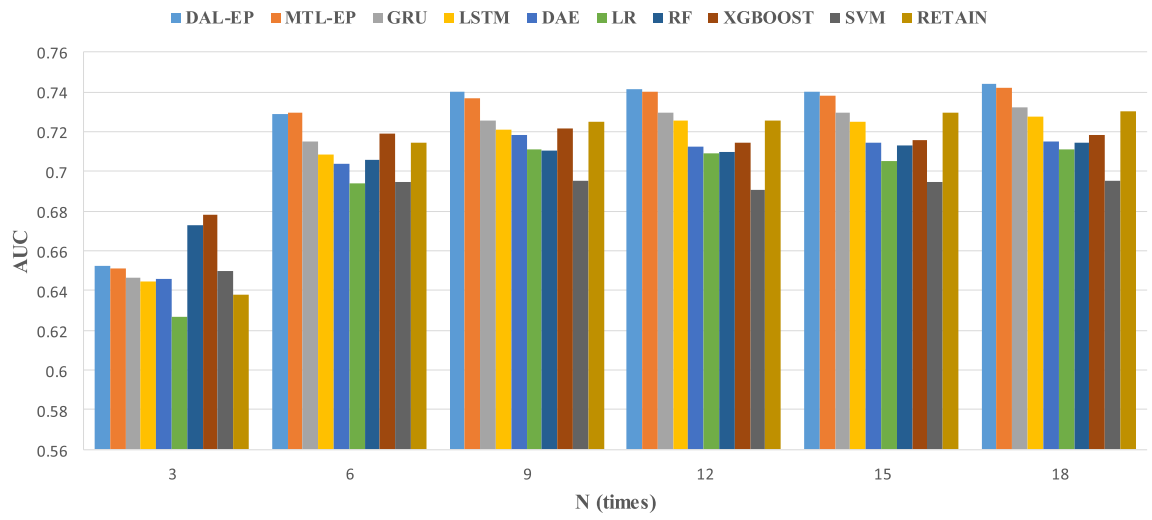
normal values frequently appear in the predicted negative samples.

Beyond that, "CRP_N" is also a feature only listed in TN and FN groups. On the contrary, "CRP_H" has higher frequency in TP and FP. Note that "CRP" is a critical factor for HF prognosis [33]. The high level of CRP reflects an inflammatory response on patient and indicates the poor therapeutic response [34]. In this regard, these findings are consistent with clinical knowledge that higher CRP is related to higher rates of readmission and mortality for HF patients [33,35].

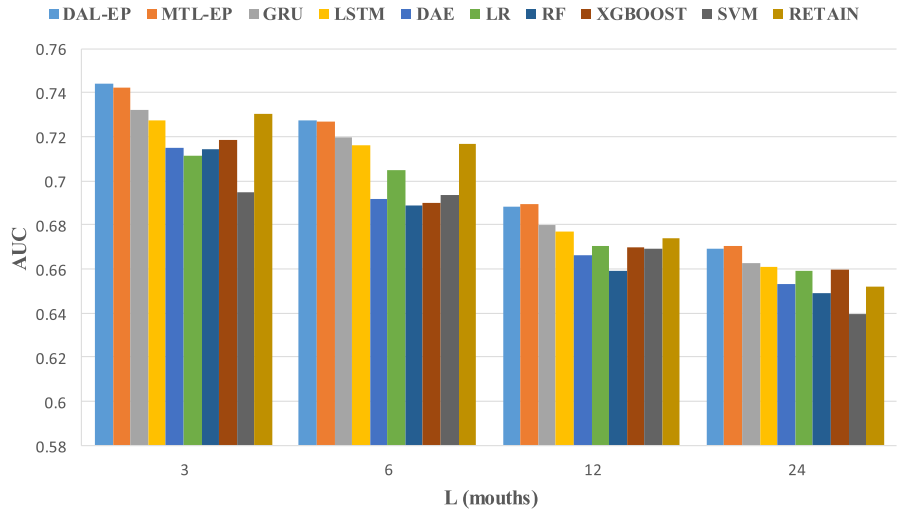
4. Discussion

4.1. Discussion on results

As can be seen in Table 2, the proposed model achieves the best AUC performance in all three cases. Taking the composite endpoint prediction as an example, the proposed DAL-EP performs better than sequential learning models including LSTM, GRU, and RETAIN, by 2.34%, 1.64%, and 1.92%, respectively, in terms of AUC. In addition,



(a)



(b)

Fig. 5. (a) AUC values of models on composite endpoint prediction with the changing N_{ptt}, where L_{pw} is set to be three months; (b) AUC values of models on composite endpoint prediction with the changing L_{pw}, where N_{ptt} is set to be 18.

the proposed model significantly outperforms the state-of-the-art non-sequential learning models including DAE, LR, RF, XGBOOST, and SVM by 4.06%, 4.64%, 4.20%, 3.48%, 7.05%, respectively. MTL-EP also achieve close performance of DAL-EP. These findings sufficiently indicate that the introduction of auxiliary information (i.e., patient features recorded at the prediction time instant) can significantly boost the performance of endpoint prediction.

Fig. 3 shows the achieved ROC curves of models on predicting endpoint when N_{ptt} is 18 times and L_{pw} is three months. Specifically, both DAL-EP and MTL-EP achieve the best performance among all models. Moreover, as can be seen in Table 2 and Fig. 3, all sequential models (LSTM, GRU, DAL-EP, MTL-EP and RETAIN) are remarkably superior to the non-sequential models (LR, DAE, RF, XGBOOST and SVM). Note that patient treatment trajectory is a kind of longitudinal data, and patient features documented on sequential visits in the trajectory have the temporal dependencies to each other. It is no wonder that non-sequential models performed worse than the other models since both of them lack the ability to tackle such sequential dependencies embedded in patient treatment trajectories.

Surprisingly, RETAIN merely achieves close performance compared with basic LSTM and GRU. Of note, RETAIN adopted a reversed-RNN-based attention mechanism to assign weights to visits in sequences. Since the outputs of RNN cells were only used to generate the attention weights, the representations for prediction task were considered to omit the sequential information, and may lead to the unremarkable performance.

As for the proposed DAL-EP model, we incorporate GAN into RNN model to equip our model with the ability of learning the auxiliary information from patient features on the next visit. As can be seen in Table 2 and Fig. 3, the generative adversarial learning strategy boosts the prediction performance. In detail, the generator fabricates a predicted feature vector of the next visit. The generated vector is then fed into the discriminator to be distinguished together with the real feature vector of the next visit, and in turn, the discriminator forces the generator to fabricate more real-like features. Both steps are iteratively conducted to make the model extract latent and representative features from the input EHR data, and to finally improve the prediction performance. The proposed MTL-EP model actually does the similar thing as DAL-EP, which feeds back MSE to RNN to benefit the information extraction, and finally achieve the close performance.

Table 4
Top 30 frequent features in each group.

Rank	TP		FP		TN		FN	
	Feature	Frequency	Feature	Frequency	Feature	Frequency	Feature	Frequency
1	Mg_N	0.989	Mg_N	0.993	Mg_N	0.994	Mg_N	0.998
2	Basophil_N	0.957	Na_N	0.972	Na_N	0.993	Na_N	0.993
3	Na_N	0.954	MPV_N	0.942	ALP_N	0.983	ALP_N	0.975
4	MPV_N	0.950	Basophil_N	0.941	D-BIL_N	0.969	MPV_N	0.957
5	CK_N	0.934	CK_N	0.928	MPV_N	0.968	AST_N	0.954
6	ALP_N	0.910	ALP_N	0.917	AST_N	0.951	Basophil_N	0.952
7	LDL-C_N	0.907	cTnT_N	0.909	Basophil_N	0.950	cTnT_N	0.947
8	AST_N	0.906	AST_N	0.902	T-BIL_N	0.941	D-BIL_N	0.939
9	ALT_N	0.900	ALT_N	0.900	cTnT_N	0.940	TP_N	0.934
10	cTnT_N	0.896	TP_N	0.893	TP_N	0.931	TT_N	0.927
11	PTA_N	0.886	D-BIL_N	0.892	Sphygmus_N	0.929	ALT_N	0.921
12	Sphygmus_N	0.876	Sphygmus_N	0.892	PTA_N	0.926	CK_N	0.921
13	D-BIL_N	0.875	LDL-C_N	0.891	CK_N	0.926	BPC_N	0.921
14	NT-proBNP_H	0.871	PTA_N	0.887	BPC_N	0.923	HFpEF	0.921
15	T-BIL_N	0.866	T-BIL_N	0.876	TT_N	0.915	PTA_N	0.914
16	TP_N	0.860	NT-proBNP_H	0.874	ALT_N	0.913	T-BIL_N	0.913
17	K_N	0.857	TT_N	0.874	HFpEF	0.911	Sphygmus_N	0.909
18	TT_N	0.826	K_N	0.865	LDH_N	0.906	LDH_N	0.899
19	HFpEF	0.820	HFpEF	0.835	WBC_N	0.901	K_N	0.898
20	PT_N	0.793	BPC_N	0.831	PT_N	0.897	LDL-C_N	0.889
21	IPHOS_N	0.789	LDH_N	0.812	RDW-CV_N	0.896	WBC_N	0.878
22	Triglyceride_N	0.774	PT_N	0.811	K_N	0.890	PT_N	0.878
23	LDH_N	0.770	WBC_N	0.797	LDL-C_N	0.876	IPHOS_N	0.868
24	BPC_N	0.764	IPHOS_N	0.788	ALB_N	0.873	ALB_N	0.853
25	TC_N	0.756	TC_N	0.758	IPHOS_N	0.872	RDW-CV_N	0.850
26	D-dimer_H	0.743	SUA_N	0.748	CRP_N	0.854	SUA_N	0.843
27	SUA_N	0.740	Triglyceride_N	0.741	GGT_N	0.840	GGT_N	0.837
28	WBC_N	0.731	Age_H	0.730	SUA_N	0.828	CRP_N	0.817
29	Age_H	0.675	D-dimer_H	0.693	RBC_N	0.807	RBC_N	0.812
30	ALB_N	0.672	ALB_N	0.689	MO_N	0.780	HGB_N	0.787

(Mg, magnesium; Na, sodium; MPV, mean platelet volume; CK, creatine kinase; ALP, alkaline phosphatase; LDL-C, low density lipoprotein cholesterol; AST, aspartic transaminase; ALT, alanine aminotransferase; cTnT, cardiac troponin T; PTA, prothrombin time activity; D-BIL, direct bilirubin; NT-proBNP, N-terminal pro-brain natriuretic peptide; T-BIL, total bilirubin; TP, total protein; K, potassium; TT, thrombin time; HFpEF, heart failure with preserved left ventricular ejection fraction (LVEF $\geq$ 50%); PT, Prothrombin time; IPHOS, inorganic phosphorus; LDH, lactic dehydrogenase; BPC, blood platelet count; TC, total cholesterol; SUA, serum uric acid; WBC, white blood cell count; ALB, serum albumin; RDW-CV, Red cell Distribution Width-Correlation Variance; RBC, Red Blood Cell Count; CRP, C-reactive protein; GGT, gamma-glutamyltransferase; MO, monocytes; HGB, Hemoglobin).

As can be seen in Table 3, the results of the paired *t*-test indicate that there are indeed statistically significant differences between the proposed models and benchmarks in HF endpoint prediction. Combined with the results in Table 2, we can conclude that the proposed models statistically outperform state-of-the-art models in terms of HF endpoint prediction.

4.2. Discussion on convergence

As can be seen in Fig. 4(a), our proposed models achieve better performance in fewer epochs. Moreover, Fig. 4(b) shows that our proposed models converge to a lower level of loss than the standard GRU. In comparison with RETAIN, our proposed model was superior on convergence with higher AUC values in fewer epochs as well as lower but more stable loss functions. These findings indicate that the introduction of auxiliary information (i.e., patient features recorded at the prediction time instant) can significantly benefit the model on fitting capacity and improving the prediction performance.

With respect to DAL-EP and MTL-EP, both of them achieved the approaching performance, but there are some differences between them on convergence. As can be seen in Fig. 4 (a), the achieved AUC values of the proposed MTL-EP reached the peak in 10 epochs, and slightly decrease with the increase of epoch, while the proposed DAL-EP does not show the same tendency. We consider that MSE in MTL-EP is the loss with much strict constraint to make predicted feature vectors closer to the real one. Such strict constraint finally may cause overfitting. By contrast, the loss from the discriminator only tell the generator whether the fabricated feature vector can be distinguished, which is more tolerant and generalized.

4.3. Discussion on the impact of N_{ptt} and L_{pw} on the prediction performance

As can be seen in Fig. 5(a), the achieved AUC values of the proposed models firstly increases gradually with the increase of the maximum length of patient treatment trajectory N_{ptt}, and then become stable with the further increase of N_{ptt}. It is comprehensive since the prediction performance can be significantly improved if there is sufficient information fed into learning.

As can be seen in Fig. 5, although non-sequential models such as RF and XGBOOST, achieved comparative performance with our proposed models when N_{ptt} is small, the performance difference between our models and baselines becomes remarkably with the increase of N_{ptt}. This finding indicates that the utilization of sequential information can indeed improve the performance of prediction.

Fig. 5(b) shows that the performance of all models decreases with the increases of L_{pw}. Note that the dependency between the past visits and the next visit becomes weaker with a larger prediction window, resulting in the deterioration of the prediction performance.

Despite the performance decreases for all models with the increase of L_{pw}, the proposed model outperforms benchmarks consistently. This finding indicates the feasibility of our proposed model for HF endpoint prediction.

4.4. Limitation and future work

Although the proposed models have achieved impressive

performance for HF endpoint prediction, many efforts can be conducted to improve the prediction performance. For example, the proposed models take the documented patient features on each visit of the trajectory as the input of RNN at each time step, while the varied time intervals between visits are not considered. Intuitively speaking, the larger the time intervals, the less the dependency between adjacent visits. While RNN models assume that an event process proceeds by unit time steps. This assumption hinders their ability to capture the heterogeneity of varied time intervals between visits. In patient treatment trajectories, time intervals between adjacent visits are usually distributed irregularly, and this characteristic inevitably impedes the performance of HF endpoint prediction using RNN models. We intend to take advantage of the time interval information to further improve the prediction by introducing time decay factor [41,42].

Moreover, the collected dataset is imbalanced with minor positive samples. Especially for the HF-readmission prediction task, only 6.14% of samples are positive. This data-imbalanced problem can be well tackled by using cost-sensitive methods [38,43] and resampling methods [44], to enrich the prediction performance of our proposed models. We plan to address this issue in our future work.

In addition, our study had the limitations that the dataset was collected from a single-site, and HF-readmissions of patients were not documented by the investigator of purpose but retrospectively based on EHR data collected from the collaborated hospital. Nevertheless, we cannot rule out that some patients in the dataset had readmissions in other medical facilities.

The other limitation is that patient treatment trajectory is partially observed. Restricted by the accessibility to clinical information in the entire treatment trajectory, this study utilizes hospitalized information to learn representative features from EHR data, on which the HF endpoint prediction task is conducted. Note that various outpatient information, and follow up data, are valuable to indicate clinical outcomes and can be leveraged into learning. We plan to remedy this limitation by conducting a clinical trial-based study, whose objective is to collect outpatient information and the daily life information of HF patients. Certainly, by providing more sufficient data into the model, the prediction performance can be further improved.

Beyond that, as deep learning models that are considered as black boxes, our proposed models lack sufficient model interpretability to help clinicians understand the derived prediction results in a convincing and reasonable manner. To remedy this limitation, we intend to equip our proposed models with the attention mechanism such that the interpretability of our models can be enriched on the premise of remaining current prediction performance [40,41,43].

Although this paper focuses on the endpoint prediction problem for HF patients, the proposed models are flexible to be adapted in other prediction tasks for various diseases, e.g., metabolic syndrome, oncology, etc., by exploring the huge potential of EHR data. In future work, we plan to evaluate our approach with a larger scale of EHR data from a wide spectrum of patient samples suffering from different diseases, and exploit the potential of the proposed models in the further clinical applications.

Moreover, we plan to deploy the proposed model in our cooperative hospitals and utilize a larger scale of data to evaluate its performance. Such a service may help clinicians to better estimate the prognosis of HF patients and adjust the treatments accordingly to reduce morbidity and mortality.

5. Conclusion

In this paper, we propose a novel deep adversarial learning model and a multi-task learning model to address the task of HF endpoint prediction. With such strategies, a suite of patient features observed at the prediction time, as the multiple auxiliary information, can be sufficiently utilized in the model training process to boost the prediction performance. The proposed approach was evaluated on a real clinical dataset collected from the cardiology department of a Chinese hospital, and the experimental results show significant improvements with considerable speedup over state-of-the-art models. In addition, our model has great potential to identify informative risk factors of HF endpoints. Some of these extracted factors are not only consistent with existing medical domain knowledge, but also contain suggestive hypotheses that could be validated by further investigations in the medical domain.

CRediT authorship contribution statement

Jiebin Chu: Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Wei Dong:** Resources, Data curation, Supervision, Funding acquisition. **Zhengxing Huang:** Conceptualization, Methodology, Writing - review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the National Key Research and Development Program of China under Grant No. 2018YFC2001204, and National Nature Science Foundation of China under Grant No. 61672450. The authors would like to extend special thanks to all experts who cooperated in the evaluation of the proposed method. The authors are particularly thankful for the positive support received from the cooperative hospitals as well as to all medical staff involved.

Appendix A. Supplementary material

In supplementary material, detailed frequency distribution of mentioned factors in error analysis and normal range of features with continues values are provided. Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2020.103518.

References

[1] P. Ponikowski, S.D. Anker, K.F. AlHabib, et al., Heart failure: preventing disease and death worldwide, ESC Heart Fail. 1 (1) (2014) 4–25.
[2] D. Gu, G. Huang, J. He, Investigation of prevalence and distributing feature of chronic heart failure in Chinese adult population, Zhonghua Xin Xue Guan Bing Za Zhi [Chinese J. Cardiol]. 31 (1) (2003) 3–6.
[3] Y. Zhang, J. Zhang, J. Butler, et al., Contemporary Epidemiology, Management, and Outcomes of Patients Hospitalized for Heart Failure in China: Results From the China Heart Failure (China-HF) Registry, J. Cardiac Fail. 23 (12) (2017) 868–875.
[4] S. Stewart, A.J. Vandenbroek, S. Pearson, et al., Prolonged beneficial effects of a home-based intervention on unplanned readmissions and mortality among patients with congestive heart failure, Arch. Intern. Med. 159 (3) (1999) 257–261.
[5] J.G.F. Cleland, M. Tendera, J. Adamus, et al., The perindopril in elderly people with chronic heart failure (PEP-CHF) study, Eur. Heart J. 27 (19) (2006) 2338–2345.
[6] Y. Guo, G.Y.H. Lip, A. Banerjee, Heart Failure in East Asia, Curr. Cardiol. Rev. 9 (2) (2013) 112–122.
[7] R. Amarasingham, B.J. Moore, Y.P. Tabak, et al., An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data, Med. Care 48 (11) (2010) 981–988.
[8] D. Kansagara, H. Englander, A. Salanitro, et al., Risk prediction models for hospital readmission: a systematic review, JAMA 306 (15) (2011) 1688–1698.
[9] B. Jin, C. Che, Z. Liu, et al., Predicting the Risk of Heart Failure With EHR Sequential Data Modeling, IEEE Access 6 (2018) 9056–9061.
[10] A.N. Jagannatha, Y. Hong, Structured prediction models for RNN based sequence labeling in clinical text, Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing, 2016, pp. 856–865.
[11] F. Ma, R. Chitta, J. Zhou, et al., Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks, Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017, pp. 1903–1911.

[12] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[13] K. Cho, B. Van Merrienboer, C. Gulcehre, et al., Learning phrase representations using RNN encoder-decoder for statistical machine translation, Empirical Methods in Natural Language Processing (EMNLP) (2014) 1724–1734.
[14] E. Choi, A. Schuetz, W.F. Stewart, et al., Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inform. Assoc. 24 (2) (2017) 361–370.
[15] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative Adversarial Nets, Advances in neural information processing systems, 2014, pp. 2672–2680.
[16] Loshchilov I, and Frank H. Fixing weight decay regularization in adam. arXiv preprint arXiv:1711.05101, 2017.
[17] M.L. Bouvy, E.R. Heerdink, H.G.M. Leufkens, et al., Predicting mortality in patients with heart failure: a pragmatic approach, Heart 89 (6) (2003) 605–609.
[18] S. Yu, F. Farooq, A.Van Esbroeck, et al., Predicting readmission risk with institution-specific prediction models, Artif. Intell. Med. 65 (2) (2015) 89–96.
[19] R.K. Cheng, M. Cox, M.L. Neely, et al., Outcomes in patients with heart failure with preserved, borderline, and reduced ejection fraction in the Medicare population, Am. Heart J. 168 (5) (2014) 721–730.e3.
[20] A.M. Richards, M.G. Nicholls, E.A. Espiner, et al., B-type natriuretic peptides and ejection fraction for prognosis after myocardial infarction, Circulation 107 (22) (2003) 2786–2792.
[21] R.S. Vasan, M.G. Larson, E.J. Benjamin, et al., Congestive heart failure in subjects with normal versus reduced left ventricular ejection fraction: prevalence and mortality in a population-based cohort, J. Am. Coll. Cardiol. 33 (7) (1999) 1948–1955.
[22] V.M. van Deursen, K. Damman, H.L. Hillege, et al., Abnormal Liver Function in Relation to Hemodynamic Profile in Heart Failure Patients, J. Cardiac Fail. 16 (1) (2010) 84–90.
[23] L.A. Allen, G.M. Felker, S. Pocock, et al., Liver function abnormalities and outcome in patients with chronic heart failure: data from the Candesartan in Heart Failure: Assessment of Reduction in Mortality and Morbidity (CHARM) program, Eur. J. Heart Fail. 11 (2) (2009) 170–177.
[24] C. Hall, NT-ProBNP: the mechanism behind the marker, J. Cardiac Fail. 11 (5) (2005) S81–S83.
[25] M. Weber, H. Christian, Role of B-type natriuretic peptide (BNP) and NT-proBNP in clinical routine, Heart 92 (6) (2006) 843–849.
[26] Y. Seino, A. Ogawa, T. Yamashita, et al., Application of NT-proBNP and BNP measurements in cardiac care: a more discerning marker for the detection and evaluation of heart failure, Eur. J. Heart Fail. 6 (3) (2004) 295–300.
[27] A. Palazzuoli, M. Gallotta, I. Quatrini, et al., Natriuretic peptides (BNP and NT-proBNP): measurement and relevance in heart failure, Vascular Health and Risk Manage. 6 (2010) 411–418.
[28] M. Richards, R.W. Troughton, NT-proBNP in heart failure: therapy decisions and monitoring, Eur. J. Heart Fail. 6 (3) (2004) 351–354.
[29] U. Alehagen, U. Dahlström, T.L. Lindahl, Elevated D-dimer level is an independent risk factor for cardiovascular death in out-patients with symptoms compatible with heart failure, Thromb. Haemost. 92 (12) (2004) 1250–1258.
[30] J.A. Ezekowitz, F.A. McAlister, P.W. Armstrong, Anemia is common in heart failure and is associated with poor outcomes: insights from a cohort of 12 065 patients with new-onset heart failure, Circulation 107 (2) (2003) 223–225.
[31] I. Anand, J.J. McMurray, J. Whitmore, et al., Anemia and its relationship to clinical outcome in heart failure, Circulation 110 (2) (2004) 149–154.
[32] G.M. Felker, W.A. Gattis, J.D. Leimberger, et al., Usefulness of anemia as a predictor of death and rehospitalization in patients with decompensated heart failure, Am. J. Cardiol. 92 (5) (2003) 625–628.
[33] J.L. Alonso-Martinez, B. Llorente-Diez, M. Echegaray-Agara, et al., C-reactive protein as a predictor of improvement and readmission in heart failure, Eur. J. Heart Fail. 4 (3) (2002) 331–336.
[34] W.H. Yin, J.W. Chen, H.L. Jen, et al., Independent prognostic value of elevated high-sensitivity C-reactive protein in chronic heart failure, Am. Heart J. 147 (5) (2004) 931–938.
[35] P.A. Kavsak, A.R. MacRae, A.M. Newman, et al., Elevated C-reactive protein in acute coronary syndrome presentation is an independent predictor of long-term mortality and heart failure, Clin. Biochem. 40 (5–6) (2007) 326–329.
[36] E. Choi, M.T. Bahadori, E. Searles, et al., Multi-layer representation learning for medical concepts, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1495–1504.
[37] C. Xiao, T. Ma, A.B. Dieng, et al., Readmission prediction via deep contextual embedding of clinical concepts, PLoS ONE 13 (4) (2018).
[38] A. Ashfaq, A. Sant'Anna, M. Lingman, et al., Readmission prediction using deep learning on electronic health records, J. Biomed. Inform. 97 (2019) 103256.
[39] X. Min, B. Yu, F. Wang, Predictive modeling of the hospital readmission risk from patients' claims data using machine learning: a case study on COPD, Sci. Rep. 9 (1) (2019) 1–10.
[40] E. Choi, M.T. Bahadori, J. Sun, et al., Retain: An interpretable predictive model for healthcare using reverse time attention mechanism, Adv. Neural Inform. Process. Syst.. (2016:) 3504–3512.
[41] T. Bai, S. Zhang, B.L. Egleston, S. Vucetic, Interpretable representation learning for healthcare via capturing disease progression through time, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 43–51.
[42] H. Duan, Z. Sun, W. Dong, et al., On Clinical Event Prediction in Patient Treatment Trajectory Using Longitudinal Electronic Health Records, IEEE J. Biomed. Health. Inf. (2019).
[43] H. Wang, Z. Cui, Y. Chen, et al., Predicting hospital readmission via cost-sensitive deep learning, IEEE/ACM Trans. Comput. Biol. Bioinf. 15 (6) (2018) 1968–1978.
[44] Z. Huang, T. Chan, Wei Dong, MACE prediction of acute coronary syndrome via boosted resampling classification using electronic medical records, J. Biomed. Inform. 66 (2017) 161–170.