Health Information Science and Systems

RESEARCH

# EAPR: explainable and augmented patient representation learning for disease prediction

Jiancheng Zhang[1,2*], Yonghui Xu[1,2*], Bicui Ye[3,4], Yibowen Zhao[1,2], Xiaofang Sun[1,2], Qi Meng[5], Yang Zhang[5*] and Lizhen Cui[1,2]

**Abstract**

Patient representation learning aims to encode meaningful information about the patient's Electronic Health Records (EHR) in the form of a mathematical representation. Recent advances in deep learning have empowered Patient representation learning methods with greater representational power, allowing the learned representations to significantly improve the performance of disease prediction models. However, the inherent shortcomings of deep learning models, such as the need for massive amounts of labeled data and inexplicability, limit the performance of deep learning-based Patient representation learning methods to further improvements. In particular, learning robust patient representations is challenging when patient data is missing or insufficient. Although data augmentation techniques can tackle this deficiency, the complex data processing further weakens the inexplicability of patient representation learning models. To address the above challenges, this paper proposes an Explainable and Augmented Patient Representation Learning for disease prediction (EAPR). EAPR utilizes data augmentation controlled by confidence interval to enhance patient representation in the presence of limited patient data. Moreover, EAPR proposes to use two-stage gradient backpropagation to address the problem of unexplainable patient representation learning models due to the complex data enhancement process. The experimental results on real clinical data validate the effectiveness and explainability of the proposed approach.

**Keywords:** Patient representation, Data augmentation, Disease prediction, Explanation method

## Introduction

Patient representations are statistical characteristics of patient clinical indicators and other multidimensional information such as name, gender, age, contact person, address, occupation, family history of the disease, etc., obtained through analysis of the patient's Electronic Health Record. Patient representation learning aims to accurately learn patient characteristics using machine learning, deep learning, and other methods to help improve the performance of disease diagnosis models [1], length of stay prediction models [2], and drug recommendation models [3], thus enabling accurate personalized treatment.

To improve patient representations' quality, researchers utilize deep learning models to encode more patient information into patient representation as Fig. 1. Then learned patient representations are used to solve various tasks such as disease prediction [4], medical image recognition [5], and so on. Because these Deep Learning-based Patient Representation learning (DLPR) models [6] have a large number of parameters, training a high-quality patient representation learning model usually requires a large amount of labeled data. Due to the privacy of patient information and cost considerations, most DLPR methods cannot obtain enough EHR datasets for model training. Inadequate training data result in DLPR models failing to learn good patient representations, which leads to poor performance on disease predictions tasks [7]. Therefore, it is important to exploit high-quality patient representations with limited EHR data. Especially, when patient representations are expected to be used for multiple disease prediction tasks, it is particularly important

*Correspondence: zhangjianchengyes@outlook.com; Xu.yonghui@hotmail.com; drzhy001@163.com
[2] School of Software, Shandong University, Jinan, China
[5] Department of Radiology, Qilu Hospital of Shandong University, Jinan, China
Full list of author information is available at the end of the article

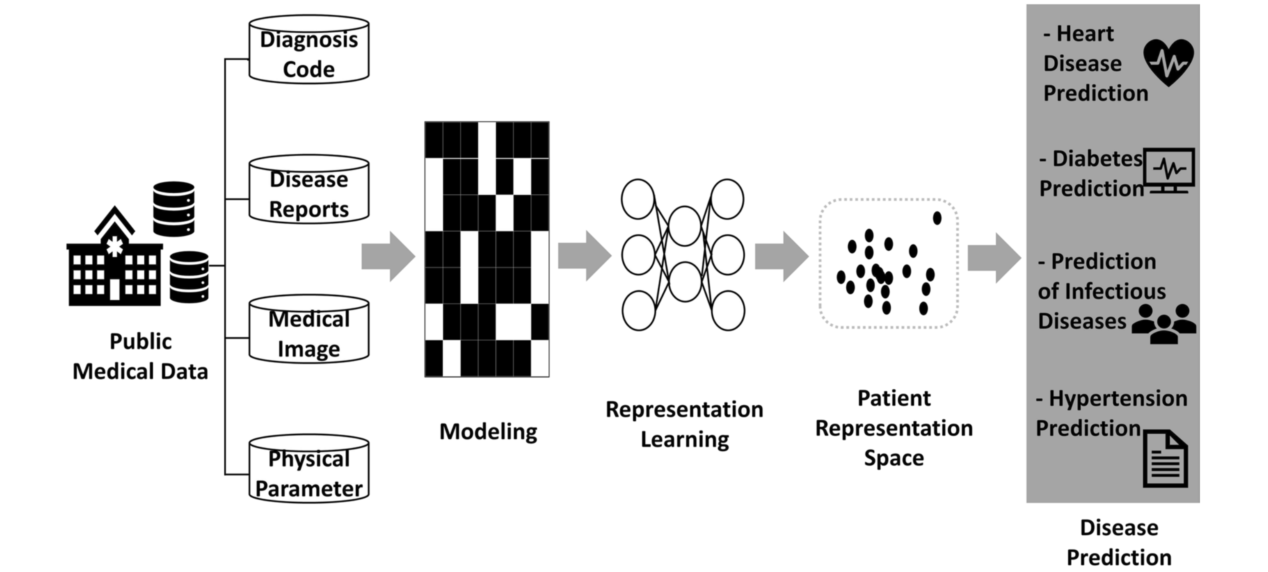


**Fig. 1** Illustration of patient representation for disease prediction

to learn robust general patient representations from EHR data.

Another challenge in DLPR is the lack of explainability. Because the learning process of most DLPR models is a black box, disease prediction results obtained are facing the challenge of lacking explainability. The researchers or doctors using the DLPR models cannot understand the different effects of patient characteristics or representations on the disease. Although there are gradient-based explanation methods, they can only explain how the learned patient representations affect the results of disease prediction, and cannot explain how the original patient representations (EHR data) affect the results of disease prediction. Because, according to the chain rule of gradient back propagation, the gradient needs to be transmitted to the patient's representation augmented (learned representations), and then to the original EHR data. And when data augmentation is introduced, randomness is introduced, so that there will be many uncertainties in the process of transferring intermediate gradient to the original EHR data. However, the methods mentioned above have not addressed this issue well. Therefore it is necessary to fully consider the explainability of patient representation learning models while enhancing patient representations.

This paper proposes an Explainable and Augmented Patient Representation learning method (EAPR) for the above challenges. EAPR augments clinical medical data with a data augmentation strategy via confidence interval control, improving capabilities such as generality and accuracy of patient representation. At the same time, to make the learning process and disease prediction process of obtained patient representations explainable, our method also provides a strategy based on a two-stage gradient backward propagation, which can help people understand the impact of different original representations of patients on diseases. The main contributions of the study are summarized as follows,

- We propose a general patient representation learning method by using the data augmentation technique, aiming at the problem of the poor generalization ability of patient representations caused by insufficient clinical data.
- We propose a two-stage gradient backward propagation method to explain disease prediction results with learned general patient representation. By this method, we ensure the explainability of disease prediction models while enhancing patient representation learning.

## Related works

Related research are in three aspects: *patient representation learning*, *data augmentation*, and *explainable methods*.

### Patient representation learning

Diversified research has been proposed to learn patient representation, [8] proposed a novel deep learning framework for the inter-patient electrocardiogram (ECG) heartbeat classification. In the approach, the symbolic representation of the heartbeat was used by a multi-perspective convolutional neural network (MPCNN) to learn automatically, which can be seen as patient representations. And [9] also proposed a method based on adversarial feature encoding with the concept of a Rateless Autoencoder (RAE). Their goal was to exploit disentangled, nuisance-robust, and universal representations used for their tasks and got effective performance. Since their model did not take into account the temporal characteristic of patient representations, after their work, [10] proposed an EHR representation method called temporal tree considering the temporal relation of data, which was based on temporal hierarchical representation of temporal co-occurrence and used doc2vec embedding technology to enhance the representation. Not only did they, but there was also a lot of work focused on this. For example, [11] developed a temporal deep learning model that can perform bidirectional representation learning on EHR sequences using a transformer model to predict future diagnosis of depression. And [12] also took time into consideration. We could find more relevant works from [13]. In addition, [14] presented a self-supervised spatiotemporal learning framework for remote physiological signal representation learning. With increasing research on graph neural networks, many researchers are also learning patient representations based on graph, such as [15, 16] and [17]. However, most of the current work learns patient representations based on supervised information like [18]. In reality, due to privacy, it is difficult to obtain rich supervised information.

### Data augmentation

Data augmentation has a significant impact on the EHR data when only a small amount of patient data is available. Therefore, [19] proposed a new text data augmentation method to generate artificial clinical notes in patients' EHR data that can be used as training data to better predict patient outcomes. Later, [20] used transfer learning and data augmentation to enhance EHR data. They systematically studied three neural network architectures, different loss function, four transfer learning strategies and four data enhancement technologies, including mixup and generative models, which were taken together to achieve data augmentation. However, the enhanced data obtained by the above work is still difficult to have higher accuracy in specific tasks. Subsequently, [21] proposed the Data Fusion using the Improved Context-aware Data Fusion algorithm, which can achieve data augmentation to some extent to solve the problem of data scarcity, but their method still suffers from inaccuracy. In conclusion, current methods have certain problems in both accuracy and robustness.

### Explainable disease prediction

The explainability of disease prediction is important for both doctors and patients. To enhance the explainability of disease prediction, [22] proposed a novel interpretable tool to explain the prediction factors in the model, which provided prediction-interpretation through a high-resolution visualization module and prediction-based creation and retrieval module. To better fuse explainability with data augmentation, [23] proposed a new interpretable random pooling neural network, in which Grad-CAM was used to interpret the results. Later, [24] proposed an interpretable network based on an attention mechanism, which applied multi-channel data augmentation and used Grad-CAM to interpret the results. And there were many works about the explanation methods [25–27]. However, the existing work like Grad-CAM can only explain how patient representations augmented(learned patient representations) affect the disease prediction results, and cannot explain how the original patient representations(i.e. before patient representations are enhanced) affect the disease prediction results. Because after the data augmentation, the patient representation space has changed.

## Methods

This section describes in detail how to augment patient representations in the presence of insufficient patient data, and how to use the augmented representations for multi-disease prediction tasks, and finally presents explainable methods for disease prediction results.

### Definition and problem statement

Suppose we have an original EHR dataset $\mathbf{E} = \{E_1, E_2, \ldots, E_i, \ldots, E_n\}$ for $n$ patients, where $E_i = \{e_{ij}^t \| t = 1, 2, \ldots, T; j = 1, 2, \ldots, M\}$ indicates EHR data of $i$-th patient. $e_{ij}^t$ is $j$-th health record item of $i$-th patient at time $t$. For different health record items, the range of values and data types of $e_{ij}^t$ vary widely. Therefore the raw data in $\mathbf{E}$ without processing cannot be directly used for disease prediction tasks. And the data distribution patterns and statistical properties in $\mathbf{E}$ cannot be represented in $E_i$ either. To solve this problem, we need to learn a new representation $P_i \in \mathcal{R}^{d\times 1}$ of patient $i$ instead of $E_i$. The process of learning $P_i$ can be formalized as learning a mapping function $f(\theta)$ from $E_i$ to $P_i$, where $\theta$ denotes the parameters in $f$. Compared to existing patient representation learning methods, we aim to use data augmentation controlled by confidence interval to further enhance the expressive power of $P_i$ to suit diverse downstream disease prediction tasks. Furthermore, we expect the learned patient representation model to remain well explainable in the disease prediction task. In other words, our method can provide the importance of different features in the original EHR in influencing a patient's development of a certain disease.

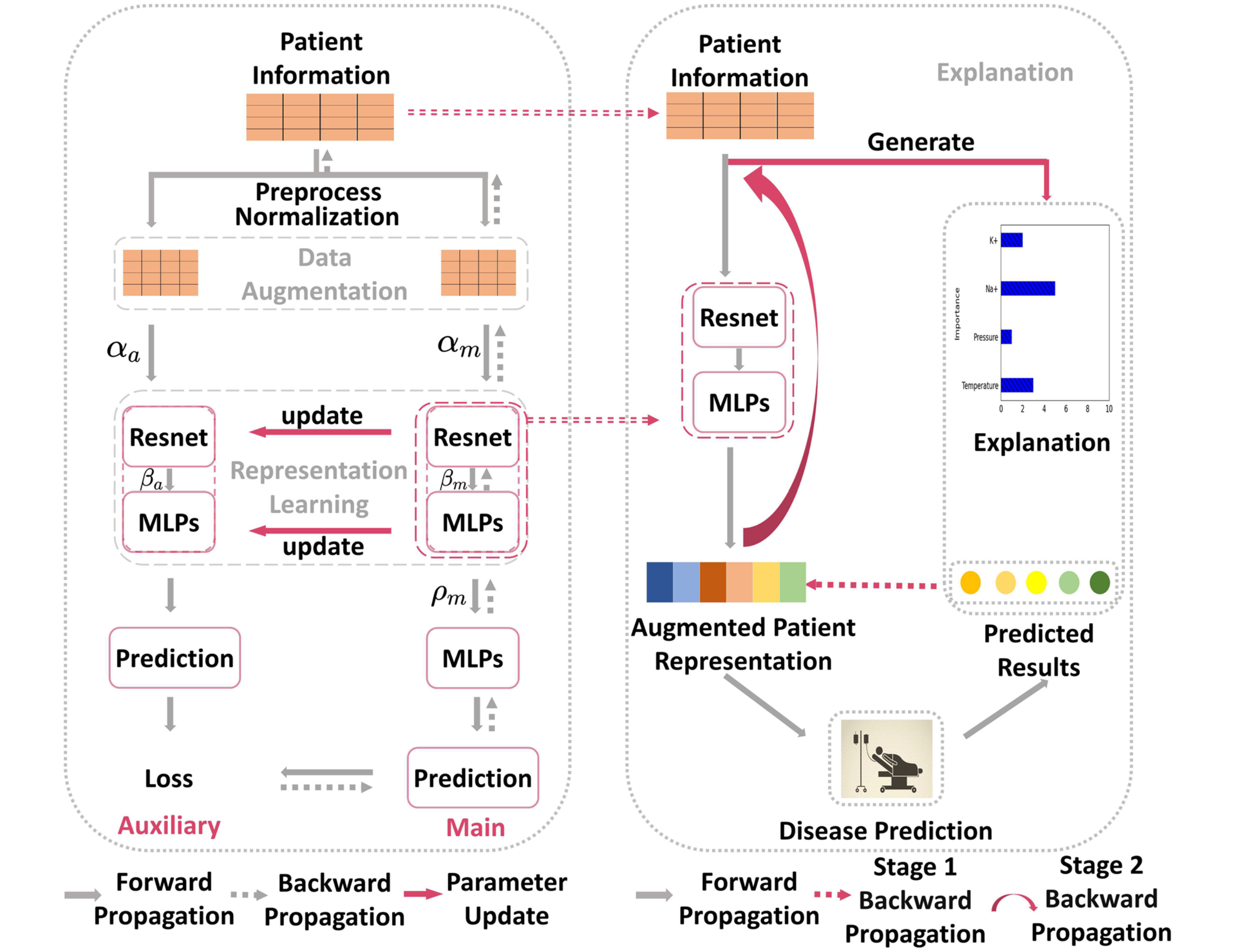


**Fig. 2** Workflow of the proposed explainable and augmented patient representation learning for disease prediction method

### Framework overview

The EAPR model comprises four key components: data augmentation controlled by confidence interval, general patient representation learning(PRL), explainable disease prediction tasks, and a two-stage gradient backward propagation approach to explain prediction results. Starting from the top left corner of the figure of workflow (Fig. 2), we begin by augmenting patient data via confidence interval control to enhance the robustness of patient representations across varying tasks. We then use a combination of ResNET and MLP to learn these patient representations. With this learned representation, we develop a disease prediction model. Finally, it can be seen from the right side of Fig. 2 where we use a two-stage gradient back-propagation strategy to calculate the gradient of disease prediction results for each feature in the original EHR data, which is the importance mentioned above, interpreting the outputs of the disease prediction model. Refer to Fig. 2 for a detailed overview of the steps involved in EAPR.

### Data augmentation controlled by confidence interval

We improve the quality of patient data by incorporating random information, which is controlled by confidence interval. To achieve this, we replicate the patient's original representation matrix, denoted as $E_i$, twice within the model to create $E_m$ and $E_a$. Both $E_m$ and $E_a$ undergo the same transformation process to introduce random information. The formula for the transformation is in Eq. (1):

$$\tilde{g}_{ij}^{t} = g_{ij}^{t} \cdot (1 - \pi) + \hat{g}_{ij}^{t} \cdot \pi \tag{1}$$

where $\hat{g}_{ij}^{t}$ refers to the random expansion or reduction of the original value $g_{ij}^{t}$ within a certain range, $\pi$ refers to the mixing rate, which is a decimal between 0 and 1, and the mixing rate imposed by $E_m$ and $E_a$ cannot be same.

However, how to value the $\pi$ above is an important issue. Because if the $\pi$ is too large, it can cause excessive loss of original useful EHR information and then reduce the accuracy of the data. And if the $\pi$ is too small, it will reduce the effectiveness of data augmentation. Therefore, we introduce a confidence interval control mechanism. The method calculates the confidence interval of the distribution of one representation($j$-th health record) at all time points for one patient($i$-th patient) in the original dataset according to Eq. (2).

$$CI_{ij} = [\frac{1}{T}\sum_{t=1}^{T} e_{ij}^{t} - 2.58 * \sqrt{\frac{1}{T}\sum_{i=1}^{T}(e_{ij}^{t} - \frac{1}{T}\sum_{t=1}^{T} e_{ij}^{t})^2}, \frac{1}{T}\sum_{t=1}^{T} e_{ij}^{t} + 2.58 * \sqrt{\frac{1}{T}\sum_{i=1}^{T}(e_{ij}^{t} - \frac{1}{T}\sum_{t=1}^{T} e_{ij}^{t})^2}] \quad (2)$$

where $CI_{ij}$ is the confidence interval When the confidence level is 99%. Using this formula, the confidence intervals for each patient's representation at all time points can be obtained. The model takes the value of $\pi$ based on the union of all confidence intervals. Based on the experimental results, the method finally decided that $\pi$ equals 0.3.

Considering the robustness of model, the patient characteristic matrices with added random information also need to be flipped horizontally to get $\acute{E}_m$ and $\acute{E}_a$, and a convolution calculation is performed separately. The convolution calculation is as Eq. (3):

$$\check{E}_m = \sum_{i,j=1}^{M} \frac{1}{2\pi\sigma^2} exp(-\frac{i^2+j^2}{2\sigma^2}) \cdot \acute{E}_m(x-i, y-j) \quad (3)$$

among them, $\sigma$ is the standard deviation of the Gaussian convolution kernel. The above operations are also applied to the $\acute{E}_a$ matrix. To retain most of the original information, this paper applies the above convolution calculation to $\acute{E}_m$ and $\acute{E}_a$ with a certain probability. Finally, we normalize the two matrices to obtain the final enhanced patient data: $\alpha_m$ and $\alpha_a$. Overall, random factors are introduced and appropriate transformations are applied to original EHR during data augmentation. As a result, the EHR data can be seen as a mixture of the original data part and the newly generated data part. As for the accuracy of EHR data, the remaining accuracy from original data part of the augmented data will be retained and enhanced in the subsequent patient representation learning.

By incorporating enhanced patient data, our patient representation learning process is significantly improved. The matrices $\alpha_m$ and $\alpha_a$ are integrated into the main network and auxiliary network depicted in Fig. 2, respectively. During the network training phase, these two matrices are treated as self-supervised information that facilitates the learning of general patient representations, and per-patient temporal dimension of data is handled that is ignored in the ResNet/MLP blocks. As a result of data enhancement, the network can better capture the effect of variations in patient data, leading to richer and more robust patient representations that can be leveraged for downstream tasks.

### Enhanced patient representation learning

To generate universal patient representations using enhanced patient data, we employ data-augmented matrices $\alpha_m$ and $\alpha_a$ as inputs in two networks, a main network and an auxiliary network. Both networks consist of a Resnet [28] and two multilayer perceptron(MLP). The Resnet and one MLP form the patient representation encoder. While $\alpha_m$ passes through the main network, $\alpha_a$ goes through a similar structure with different parameters in the auxiliary network. This approach enables us to capture more variation in patient data and produce more robust patient representations that can be utilized in downstream tasks.

We propose a self-supervised contrastive learning approach to train the main and auxiliary networks in our patient representation learning process. This method of self-supervised learning is inspired by [29], which is also a self-supervised learning model with good performance for image processing. Specifically, we formulate the weight parameters $\nu$ and $\epsilon$ for the main and auxiliary networks in Eq. (4) as follows:

$$\epsilon = \tau\epsilon + (1-\tau)\nu \quad (4)$$

where $\tau$ is the decay rate of the auxiliary network update. To enhance the patient data, we construct two factors of self-supervised contrast. We first encode the augmented matrices $\alpha_m$ and $\alpha_a$ to obtain representations $\rho_m$ and $o_a$ , respectively. Then, we use $\rho_m$ to produce a prediction $o_m$ through a multilayer perceptron, and compute the loss function as follows,

$$L_{m,a} = 2 - 2 \cdot \frac{\langle o_m, o_a \rangle}{\|o_m\|_2 \cdot \|o_a\|_2} \quad (5)$$

The main network is trained using the loss function derived from Eq. (5). Since our patient representation learning is based on self-supervised learning, it does not require labeled data.

During training, the enhanced patient data passes through the parallel network to obtain two predicted outputs $o_m$ and $o_a$. Both predictions are then used to calculate the loss function using Eq. (5), and the resulting gradient of the loss is only sent back to the main network for updating. Meanwhile, the auxiliary network is updated according to Eq. (4). Through this self-supervised contrastive learning approach, our model learns general patient representations, which are presented as high-dimensional vectors. After data augmentation and patient representation learning, the original EHR feature space changes, and the new feature space loses original physical meaning, which will bring some difficulties to explanation. Therefore, our method considers using a two-stage calculation to obtain corresponding explanation results.

### Disease prediction via PRL

Using the representations produced in previous step, we work perform multiple sets of disease prediction tasks, as shown in Fig. 2. In this section, the medical prediction tasks $D$, such as sepsis prediction, based on the general patient representation will be discussed. Through experiments, this work selects learner $L$ with the best prediction performance on $D$ to learn and predict diseases. For each task $D^{(i)} \in D$, the work defines the loss function of the learner $L$ as Eq. (6):

$$Loss = \sum_{k=0}^{K} -y_k logp_k(\rho_n|D^{(i)}, L(\xi)) \tag{6}$$

where $y_k$ refers to whether the patient is in the $k$ state, $p_k$ refers to the probability that the learner predicts the patient to be in the $k$ state under the $D^{(i)}$ task, and there are a total of $(K+1)$ states, such as illness, discharge or recovery, etc., depending on the specific task. $\xi$ refers to the parameters of the learner $L$. $L$ through learning, according to the Eq. (7) get the optimal parameters,

$$\tilde{\xi} = \arg\min_{\xi} \sum_{k=0}^{K} -y_k logp_k(\rho_n|D^{(i)}, L(\xi)) \tag{7}$$

The trained learner accepts the general patient representation and can get the probability of disease prediction:

$$p_j(\rho_n|D^{(i)}, L) = \frac{exp(L(\rho_n, \tilde{\xi})_j)}{\sum_{k=0}^{K} exp(L(\rho_n, \tilde{\xi})_k)} \tag{8}$$

In the Eq. (8), $L(\rho_n, \tilde{\xi})$ represents the disease probability of the patient given by $L$ after inputting the general patient representation of the nth patient to the learner $L$. The whole disease prediction process is shown in the disease prediction in Fig. 2. Since our model is universal, the learner $L$ here can have multiple choices, and the explanation of disease prediction will be discussed below.

### Explanation via two-stage gradient backward

This section will expound that our proposed explanation method, which is based on two-stage gradient backward propagation. As disease prediction alone is insufficient to support doctors in making informed decisions, it is crucial for the model to show how the patient's representation influences the prediction outcome.

**Drawbacks of traditional explanation methods:** Traditional approaches achieve explanation is by backpropagating the gradient of the prediction result step by step to the input of the prediction model. However, this method only provides information on the influence of the general patient representation, which is a high-dimensional vector with no practical significance for doctors. Additionally, existing methods are unable to explain the impact of the pre-data augmentation patient representation on the prediction results. Therefore, this paper proposes a two-stage gradient backward propagation explanation method for disease prediction results.

**First-stage gradient backward:** To explain the prediction result, we need to start by performing the initial step of gradient backward propagation. This step involves backpropagating the gradient from the disease prediction model's output to its input. To accomplish this, we need to calculate the significance of the overall patient representation in disease prediction, which can be done using Eq. (9).

$$W_n = \frac{\partial p(\rho_n|D^{(i)}, L)}{\partial \rho_n} \tag{9}$$

$W_n$ is the corresponding importance of the generic patient representation for the $n$-th patient. If the first stage is not deep learning, then an explanation method based on a specific model is required to get our $W_n$. For example, random forest, first needs to be calculated the Gini index of each node:

$$G(\rho) = 1 - \sum_{k=0}^{K} \rho_{wk}^2 \tag{10}$$

where $\rho_{wk}$ denotes the weight of samples of class $k$. According to the Eq. (11), model can get the importance of a feature in the generic patient representation:

$$W'_{\rho^{(j)}} = \sum_{i \in n'} \sum_{m \in M'} \Delta G_{mi}^{\rho^{(j)}} \tag{11}$$

The result calculated by the Eq. (11) is the importance of the $j$-th feature in the general patient representation, and $\Delta G_{mi}^{\rho^{(j)}}$ represents the difference between the Gini index before and after the $j$-th feature branch on the $m$ node of the $i$ tree. Finally, after normalization, model get:

$$W_{\rho^{(j)}} = \frac{W'_{\rho^{(j)}}}{\sum_{i=1}^{Z'} W'_{\rho^{(i)}}} \tag{12}$$

The model finally returns the importance corresponding to the general representation of all patients, namely $W_n$, to the input of the model for downstream tasks.

**Second-stage gradient backward:** The second stage of gradient backward propagation is to feed the gradient of the importance of the generic patient representation back to the input of the patient representation encoder. In summary, the input of the gradient feedback in the second stage is the gradient of the importance of the learned patient representation for the disease, and the output is the gradient of the original patient representation that is the importance of the original patient representation for the disease. The model first passes the gradient of $W_n$ back along the patient representation encoder, and the returned gradient first passes through the MLP of the patient representation encoder. The calculation is as Eq. (13):

$$\frac{\partial W_n}{\partial z^{(i)}} = \frac{\partial W_n}{\partial h^{(i)}} \cdot \frac{\partial h^{(i)}}{\partial z^{(i)}} \tag{13}$$

In the Eq. (13), $h^{(i)}$ represents the output of each layer in the MLP, $z^{(i)}$ is the result of applying the activation function to $h^{(i)}$, and the Eq. (13) is used for each layer of the MLP to obtain the gradient of $W_n$ with respect to the intermediate result $\beta_m$ of patient representation encoder. The gradient is passed back along the Resnet of the patient representation encoder, and the corresponding importance of $E_n$ is calculated:

$$w_n = \frac{\partial W_n}{\partial \beta_m} \cdot \frac{\partial \beta_m}{\partial E_n} = \frac{\partial W_n}{\partial E_l}(1 + \frac{\partial}{\partial E_n}\sum_{k=n}^{l-1} G(E_k, \theta'_k)) \tag{14}$$

where $G$ is the residual function and $\theta'_k$ is the weight parameter of the corresponding layer. Through the Eq. (14), the model can get the gradient of $W_n$ to the corresponding element of the patient's original information, and assigns the gradient value to $w_n$. Finally, model will get a matrix containing $w_n$ as the explanation result, representing the changing relationship between the data of the original information and the prediction results of the downstream tasks. Hence, Doctors can know which features in the original EHR data have a greater impact.

**Discussion on explainability:** The proposed explainable method in this paper provides significant advantages over existing explanation methods, such as Shapley values, which are commonly used to explain feature importance. Shapley-based methods require considering multiple missing features during the importance calculation process, resulting in a large search space and prolonged computation time. In contrast, our explanation method only needs two backward propagations for accurate results, making it highly efficient.

EAPR possesses another advantage compared to interpretation methods that have smaller search spaces. These methods typically assume that the same input will consistently yield the same output explanation. However, due to the utilization of data augmentation techniques, random information is incorporated, resulting in different outputs even with the same inputs. Our explanation method can circumvent the influence of these randomly generated factors and directly provide the gradient to the original patient representation, guaranteeing utmost accuracy. Consequently, our approach presents a superior option for elucidating disease prediction outcomes.

## Results

This section conducts extensive experiments to verify the validity and generalization of generic patient representations.

### Datasets

We use three datasets extracted publicly, which anyone can access, from MIMIC-[1]: **sepsis dataset**, **acute hypotension dataset** and **cancer dataset**[2], to carry out contrast experiments to verify our algorithm.

**The sepsis dataset** includes 2164 sepsis patients and 4383 non-septic patients. The dataset set 20-time windows to record information for each patient, and each time window was four hours. Patient data is limited to 6 physical parameters, 33 laboratory parameters, and 7 other personal information. Specific information on sepsis patients is provided in Table 1.

**The acute hypotension dataset** includes information on 3910 acute hypotension patients and 2637 non-acute hypotension patients as shown in Table 2. This dataset aims to predict whither patient has acute hypotension. Each patient has 48 one-hour time windows, each of which records a patient's data. The dataset is limited to three physical parameters, seventeen laboratory parameters, and two personal information.

**The cancer dataset** includes 6547 patient information and records phenotypic information of cancer patients. The label of this data set is whether it is a cancer patient, which is used for cancer prediction task. This dataset has 15 features in total. The specific phenotypic information is shown in Table 3.

[1] https://physionet.org/content/mimiciii/1.4/

[2] https://physionet.org/content/synthetic-mimic-iii-health-gym/1.0.0/

**Table 1 Parameter range summary in sepsis datasets**

| Name | Unit | (25th/50th/75th percentile) |
|---|---|---|
| Age | year | 58.3/65.4/73.0 |
| Heart Rate (HR) | bpm | 78.5/89.1/ 99.8 |
| Systolic BP | mmHg | 114.4/123.7/133.0 |
| Mean BP | mmHg | 75.2/81.0/86.9 |
| Diastolic BP | mmHg | 50.4/58.9/67.0 |
| RR | bpm | 18.7/21.5/24.3 |
| K+ | meq/L | 3.8/4.1/4.5 |
| Na+ | meq/L | 136.6/140.0/143.6 |
| Cl- | meq/L | 102.1/105.2/108.0 |
| Ca++ | mg/dL | 7.4/8.0/8.7 |
| Ionised Ca++ | mg/dL | 1.0/1.1/1.2 |
| CO2 | meq/L | 23.4/25.3/27.3 |
| Albumin | g/dL | 2.7/3.0/3.3 |
| Hb | g/dL | 9.2/10.2/11.2 |
| pH | – | 7.3/7.4/7.4 |
| BE | meq/L | − 2.0/0.2/2.5 |
| HCO3 | meq/L | 22.6/24.4/26.1 |
| FiO2 | fraction | 0.4/0.5/0.5 |
| Glucose | mg/dL | 108.2/134.1/167.1 |
| BUN | mg/dL | 19.9/25.4/31.9 |
| Creatinine | mg/dL | 0.9/1.1/1.4 |
| Mg++ | mg/dL | 1.8/2.0/2.3 |
| SGOT | u/L | 31.5/50.8/89.0 |
| SGPT | u/L | 26.2/40.0/65.7 |
| Total Bili | mg/dL | 0.6/1.2/2.3 |
| WBC | E9/L | 8.0/10.6/13.9 |
| Platelets | E9/L | 142.0/184.4/239.4 |
| PaO2 | mmHg | 84.2/109.1/139.6 |
| PaCO2 | mmHg | 34.9/ 39.3/ 45.0 |
| Lactate | mmol/L | 1.4/1.8/2.4 |
| Input Total | mL | 1887.8/4867.5/11155.8 |
| Input 4H | mL | 13.8/58.7/229.0 |
| Max Vaso | mcg/kg/min | 7.9E-06/ 0.0/0.0 |
| Output Total | mL | 585.5/2505.5/6733.7 |
| Output 4H | mL | 44.7/159.3/361.7 |
| Gender | 0=male, 1=female | 0/0/1 |
| Readmission | – | 0/0/1 |
| Mech | – | 0/0/1 |
| GCS | – | 10/14/15 |
| SpO2 | % | 2/ 5/ 7 |
| Temperature | Celcius | 2/5/8 |
| PTT | s | 3/5/8 |
| PT | s | 2/5/8 |
| INR | – | 2/5/7 |
| IDs | – | 540.8/ 1081.5/1622.3 |
| Timepoints | – | 4.8/9.5/14.3 |

**Table 2 Parameter range summary in acute hypotension datasets**

| Name | Unit | (25th/50th/75th percentile) |
|---|---|---|
| MAP | mmHg | 59.3/65.3/71.2 |
| Diastolic BP | mmHg | 48.4/54.3/60.3 |
| Systolic BP | mmHg | 104.2/113.2/121.6 |
| Urine | mL | 68.9/106.2/164.2 |
| ALT | IU/L | 24.6/32.6/46.1 |
| AST | IU/L | 35.8/46.8/67.8 |
| PaO2 | mmHg | 91.3/103.0/114.7 |
| Lactate | mmol/L | 1.3/1.5/1.8 |
| Serum Creatinine | mg/dL | 0.8/1.1/1.6 |
| Fluid Boluses | mL | 0/0/0 |
| Vasopressors | mcg/kg/min | 0/0/0 |
| FiO2 | fraction | 0.5/0.5/0.5 |
| GCS | – | 11/15/15 |
| Urine (M) | – | 0/0/1 |
| ALT/AST (M) | – | 0/0/0 |
| FiO2 (M) | – | 0/0/0 |
| GCS (M) | – | 0/0/0 |
| PaO2 (M) | – | 0/0/0 |
| Lactic Acid (M) | – | 0/0/0 |
| Serum Creatinine (M) | – | 0/0/0 |
| IDs | – | 977/1954/2932 |
| Timepoints | – | 11.8/23.5/35.3 |

### Comparison baselines

We used 6 baseline models in three disease prediction tasks, and compared the performance of our model and baseline models in disease prediction tasks to verify the validity and generalization of general patient representation learned by our model. Firstly, considering that our model is a kind of multi-classification disease prediction model, we compare with Decision Trees and SVM respectively, which are classic machine learning algorithms of binary or multi-classification. Secondly, considering that our data are complex patient data, we include CNN [30] with strong expression ability to fit complex data as a baseline for comparison. Thirdly, considering that the data we use is time series data, we need to compare with the baseline model that is good at processing time series data. For this purpose, we first compare with RNN [31], which can predict disease by mining temporal information. Then, since GRU [32] is a classical deep learning model with stronger ability to process time series data than RNN, we include it as a baseline, which combines the memory of patients' historical information to predict diseases. Finally, considering the shortcomings of GRU in some tasks, we add BiLSTM [33] as the baseline model, which can predict diseases by memorizing

**Table 3 Parameter summary in cancer datasets**

| Phenotype |
|---|
| Advanced Cancer |
| Advanced Heart Disease |
| Advanced Lung Disease |
| Alcohol Abuse |
| Chronic Neurological Dystrophies |
| Chronic Pain Fibromyalgia |
| Dementia |
| Depression |
| Developmental Delay |
| Non Adherence |
| None |
| Obesity |
| Other Substance Abuse |
| Schizophrenia and other Psychiatric Disorders |
| Unsure |

time series data and has a strong ability to process historical information of patients.

To evaluate the effectiveness of EAPR, our use 4 evaluation metrics: F1-score, accuracy, precision, and recall. These four metrics are used to evaluate the performance of our model and the baseline models on the disease prediction tasks. The evaluation results are shown below. To better reflect the effectiveness of the patient representation our model learned, we also choose four other baseline models [34], [35], [36], [37] to validate EHR representation learning.

### Implementation details

The basic network for patient representation learning in Section III-D is based on a pre-trained Resnet50, the projection and prediction layers of the network are multilayer perceptrons, and the linear layers and common activation functions are used to construct them. The EHR data is transformed into an image because adjacent rows' numbers in the EHR data matrix have correlations due to timepoint and this is similar to image. When learning the general patient representation, the network is trained with a learning rate of 3e-4. Since our model and all baseline models are based on supervised learning in disease prediction, the same training sets were used in the training phase of all models and the same testing sets were used in testing phase. Specifically, in each disease prediction task, the corresponding datasets were randomly divided into 70% as the training sets, with the remaining being the testing sets.

### Comparisons on the sepsis dataset

This work compares and analyzes the performance of our model and the baseline models on the sepsis prediction task. The experimental results are shown in Table 4. According to the results, the decision tree(DT) performs well on the recall rate(around 93.6%), but the other performance is poor. The BiLSTM and SVM perform well on other performance than the recall rate. For example, the BiLSTM has 96.0% on the F1-score and 97.3% on the accuracy, which is better than others. The performance of our model surpasses all the baseline models, with 97.7%, 98.3%, 99.7%, 95.7% on F1-score, accuracy, precision and recall respectively. From the analysis results, in the process of constructing the DT, it made decisions by selecting the best features. However, its expression ability is poor, resulting in the inability to fit complex patient data information. The results of SVM show that the model finds the key data very accurately, but the performance is also limited by its expressive ability. BiLSTM benefits from its memory of the patient's past information, but the recall rate is not high.

Analyzing our model, due to data augmentation, EAPR not only captures the temporal relationship of patient data well but also achieves much better prediction recall(around 95.7%) than all baseline models through the encoded high-dimensional patient representation vector. This also verifies that the learned general patient representation contains a lot of valid and rich patient information, which enables the model to perform well on the task of predicting sepsis patients.

### Comparisons on the acute hypotension dataset

This section compares and analyzes the performance of our model and the baseline models on the acute hypotension prediction task. The experimental results are shown in Table 5. Looking at the results, the RNN has a good recall rate(around 96.2%), but does not perform well on other metrics like 88.9% on the precision. Although the recall rate of CNN(around 95.2%) is not as high as that of RNN, other performance is the best, especially on the

**Table 4 Comparison results on sepsis prediction task**

| Sepsis | F1-score (%) | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| SVM | 96.2 | 97.3 | 99.6 | 93.0 |
| DT | 92.4 | 94.5 | 91.2 | 93.6 |
| CNN | 95.7 | 97.1 | 99.6 | 92.1 |
| GRU | 95.8 | 97.2 | 99.6 | 92.3 |
| RNN | 95.9 | 97.3 | 99.6 | 92.5 |
| BiLSTM | 96.0 | 97.3 | 99.5 | 92.7 |
| EAPR | **97.7** | **98.3** | **99.7** | **95.7** |

The bold represents the performance of our method/model (EAPR)

**Table 5 Comparison results on acute hypotension prediction task**

| AH | F1-score (%) | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| SVM | 97.2 | 96.7 | 99.6 | 94.9 |
| DT | 95.5 | 94.4 | 95.2 | 95.8 |
| CNN | 97.4 | 96.8 | 99.7 | 95.2 |
| GRU | 92.8 | 90.6 | 89.9 | 95.9 |
| RNN | 92.4 | 90.1 | 88.9 | 96.2 |
| BiLSTM | 97.4 | 96.7 | 99.4 | 95.5 |
| EAPR | **98.1** | **97.7** | **99.7** | **96.5** |

The bold represents the performance of our method/model (EAPR)

**Table 6 Comparison results on cancer prediction task**

| Cancer | F1-score (%) | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| SVM | 96.0 | 92.5 | 92.5 | 99.8 |
| DT | 91.8 | 85.1 | 92.8 | 90.8 |
| CNN | 95.8 | 92.1 | 92.1 | 99.8 |
| GRU | 96.1 | 92.8 | 92.8 | 99.6 |
| RNN | 95.9 | 92.3 | 92.3 | 99.8 |
| BiLSTM | 95.8 | 92.1 | 92.1 | 99.8 |
| EAPR | **96.3** | **92.9** | **93.1** | **99.8** |

The bold represents the performance of our method/model (EAPR)

precision(99.7%). Our model also has better performance than the baseline models like 98.1% on the F1-score. From the analysis results, compared with CNN, RNN can have a good recall rate because it can handle the time series relationship in patient data. However, RNN may not perform well on other metrics due to the difficulty of obtaining remote patient data information. When CNN processes patient data, through step-by-step convolution, it can finally see the complete data of the patient, so it can achieve good performance on multiple indicators, but unfortunately, the recall rate(around 95.2%) is not high, which may be limited due to its small number of parameters and it is not enough to describe all patient characteristics.

Since EAPR uses the patient representation encoding a large amount of patient historical information after representation learning, so the model can see the temporal relationship of patient data. Furthermore, our patient representations are more dimensional and therefore more conducive to depicting complex relationships between patient data, ultimately yielding results that are higher than the highest recall in the baseline models. This still validates the effectiveness of general patient representations on medical tasks.

### Comparisons on the cancer dataset

In this section, the work compares and analyzes the performance of our model and the baseline models on the cancer prediction task. The experimental results are shown in Table 6. Looking at Table 6, the DT has the highest precision(around 92.8%), but other indicators are not high. The precision of GRU(around 92.8%) is close to the DT, but its other indicators except recall are the highest like 96.1% on the F1-score. Our model outperforms all baseline models on most metrics like 92.9% on the accuracy. Analyzing the results, the recall rate(around 90.8%) of the DT is low, but the precision is higher than all the baseline models, and this work attributes this result to the feature dimension of the dataset. Because the feature dimension of this data set is very small compared with the previous data set, this leads to the DT may not be able to identify the best classification method when selecting features for decision-making. Therefore, although the precision is high, the recall rate is very low. The average performance of GRU is so good, mainly due to its ability to remember the historical data of patients. GRU can predict cancer well by capturing the time series relationship of patient data. The reason why its performance is not as good as our model is that our model is more able to fully mine and fit the time series relationship existing in the patient data, and mine the important features of the prediction due to the comparison learning between the main network and the auxiliary network when the patient representation is learned.

### Comparisons on EHR representation learning

This section compares and analyzes the performance of our model and the four baseline models on the three diseases prediction task. The experimental results are shown in Table 7, Tables 8 and 9. Looking at the results, the first baseline model generally has lower four indicators on the three prediction tasks compared to other models, and the remaining three baseline models have relatively good performance in all four indicators of three tasks. However, the performance of our model surpasses all the baseline models on all prediction tasks. From the analysis results, [34](ES) embeds medical codes and temporal features of patients, and then uses GRU to learn patient representations. However, there is currently no medical code in our data, which results in the representations learned by this method not performing well in disease prediction. Compared to this, the advantage of EAPR is reflected. Even if there is a lack of medical code information, it can still compensate for the negative impact of data scarcity through its own data augmentation. Whether it is the [35] (EZ) that requires two-stage learning of patient representation, the [36] (J-S) based on Word2Vec, or the [37] (YW) based on nonnegative matrix factorization, all have learned meaningful patient representation that

**Table 7 EHR representation learning for predicting sepsis**

| Sepsis | F1-score (%) | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| ES | 56.6 | 74.7 | 69.5 | 47.8 |
| EZ | 95.1 | 96.7 | 99.5 | 91.1 |
| J-S | 96.6 | 97.7 | 99.7 | 93.7 |
| YW | 95.7 | 97.1 | 98.8 | 92.8 |
| EAPR | **97.7** | **98.3** | **99.7** | **95.7** |

The bold represents the performance of our method/model (EAPR)

**Table 8 EHR representation learning for predicting AH**

| AH | F1-score (%) | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| ES | 82.4 | 76.1 | 76.2 | 89.7 |
| EZ | 97.8 | 97.3 | 99.5 | 96.1 |
| J-S | 97.5 | 97.0 | 99.6 | 95.6 |
| YW | 97.5 | 96.9 | 99.1 | 95.8 |
| EAPR | **98.1** | **97.7** | **99.7** | **96.5** |

The bold represents the performance of our method/model (EAPR)

**Table 9 EHR representation learning for predicting cancer**

| Cancer | F1-score (%) | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| ES | 95.6 | 91.5 | 91.7 | 99.8 |
| EZ | 95.9 | 92.2 | 92.4 | 99.8 |
| J-S | 95.6 | 91.5 | 92.2 | 99.2 |
| YW | 95.8 | 91.9 | 92.9 | 98.8 |
| EAPR | **96.3** | **92.9** | **93.1** | **99.8** |

The bold represents the performance of our method/model (EAPR)

can perform well in all disease prediction tasks, which also validates the achievements of these three works. Compared to the [35] (EZ), EAPR demonstrates higher efficiency as it does not require a two-stage complex process, and can only learn data augmented to acquire patient representations. And compared to Word2Vec and nonnegative matrix factorization, EAPR also shows the advantages of combining our data augmentation and our proposed patient representation learning.

From the analysis of our model, after EAPR introduces a certain degree of randomness through data augmentation, the network structure of patient representation learning can better extract the patterns existing in the original EHR data. Then our model can code more effective patient information. Finally, the representation learnt also reflects its own effectiveness and generalization in our disease prediction tasks.

## Explanation and analysis

To verify the explainability of the proposed method, we analyzed the explainability of the method on the sepsis dataset and acute hypotension dataset. The results of top10 are displayed in Figs. 3 and 4. The parameter relevance can be normalized to a range between 0 and 1.The paper presents the explanation results for sepsis prediction in Fig. 3. From the observation results, longitudinally, the WBC characteristic, Creatinine characteristic, and BE characteristic of patients had the greatest influence on the model prediction of sepsis, and PaO2 and other characteristics had a greater influence. Horizontally, when the prediction time is closer, important features such as WBC affect the prediction of sepsis to a greater extent, which reflects the correlation between parameters and time. Analyzing the results, the explanation can describe in the horizontal direction that the degree to which the feature affects the prediction result varies with the distance from the prediction time point. In addition, our explainable method is for all features at all time points, and the factors considered are comprehensive enough, so that the importance of all features affecting the prediction results can be analyzed longitudinally.

The explanation results for acute hypotension prediction are presented in Fig. 4. Observing the results, AST and FiO2(M) had a great influence on the prediction of the disease, and characteristics such as PaO2 and ALT/AST(M) had a less greater influence. The correlation between features and time can be observed from this figure. As time goes by, the influence of features tends to increase. Analyzing the results, due to the computational characteristics of our explainable means, after the prediction task, the degree of influence of all the features on the prediction results can be given. Moreover, through visualization, our explainable approach can provide clinicians with an intuitive explanation and a reference for physicians, that is, which features are more important and which are less sensitive to the patient's disease, and how these features affect the disease changing over time.

## Ablation study

This section validates the significance of important structures in the model by eliminating them. The experimental results depicted in Fig. 5 demonstrate that the performance of EAPR on the three disease prediction tasks noticeably deteriorated after excluding the data enhancement module, predictor and deactivating the pre-trained Resnet. The reason behind this decline lies that the acquired generalized patient representations fail to encode comprehensive patient information in the absence of data augmentation, ultimately resulting in a considerable performance drop. Additionally, eliminating the second MLP from the main network deprives the network of an opportunity to reintegrate the previously extracted features, consequently diminishing the performance of the disease prediction tasks. Not

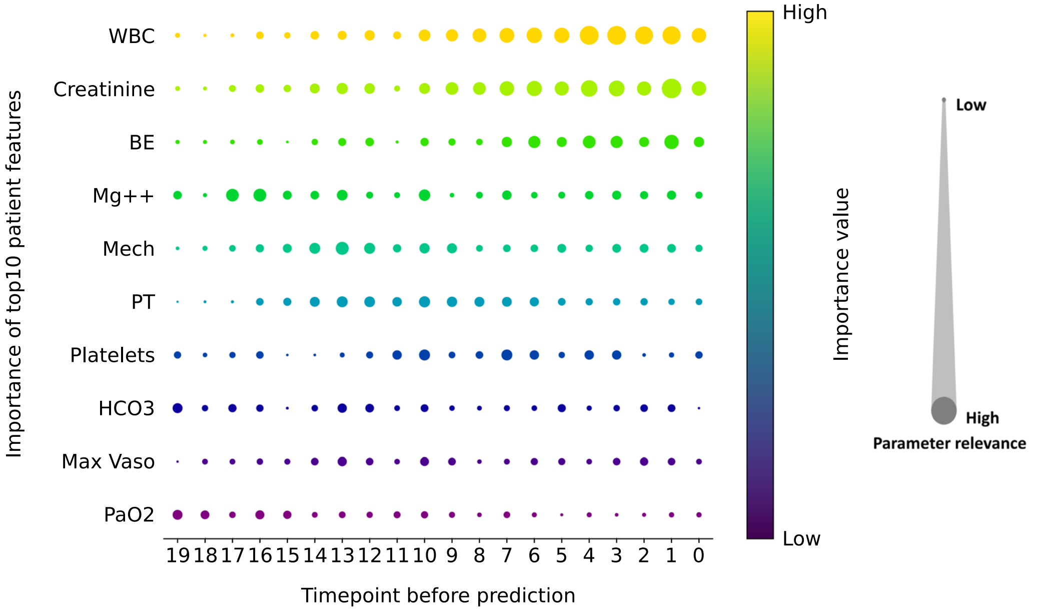


**Fig. 3** Explanation of prediction results of sepsis

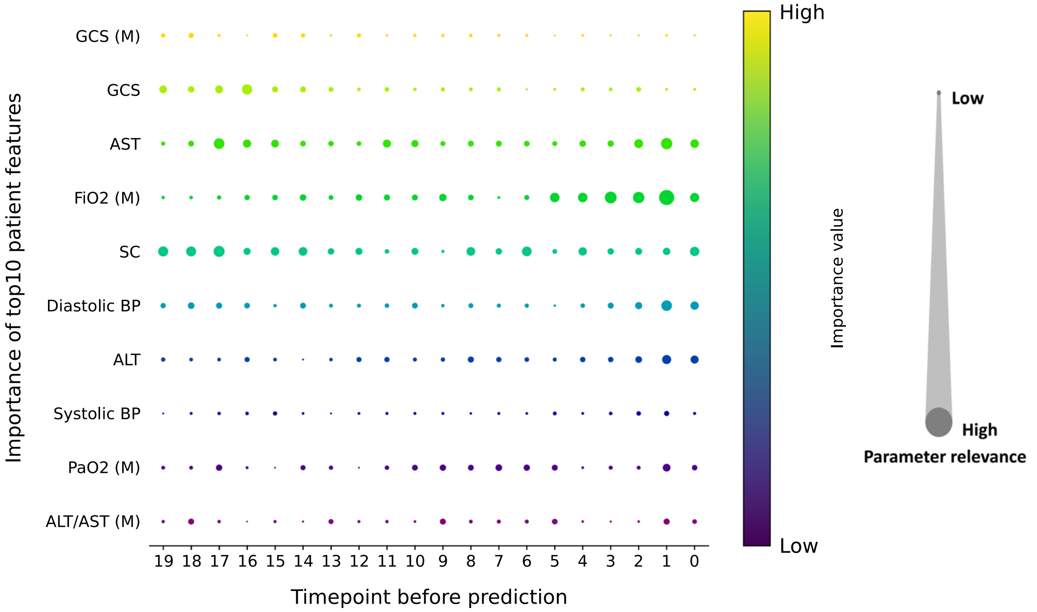
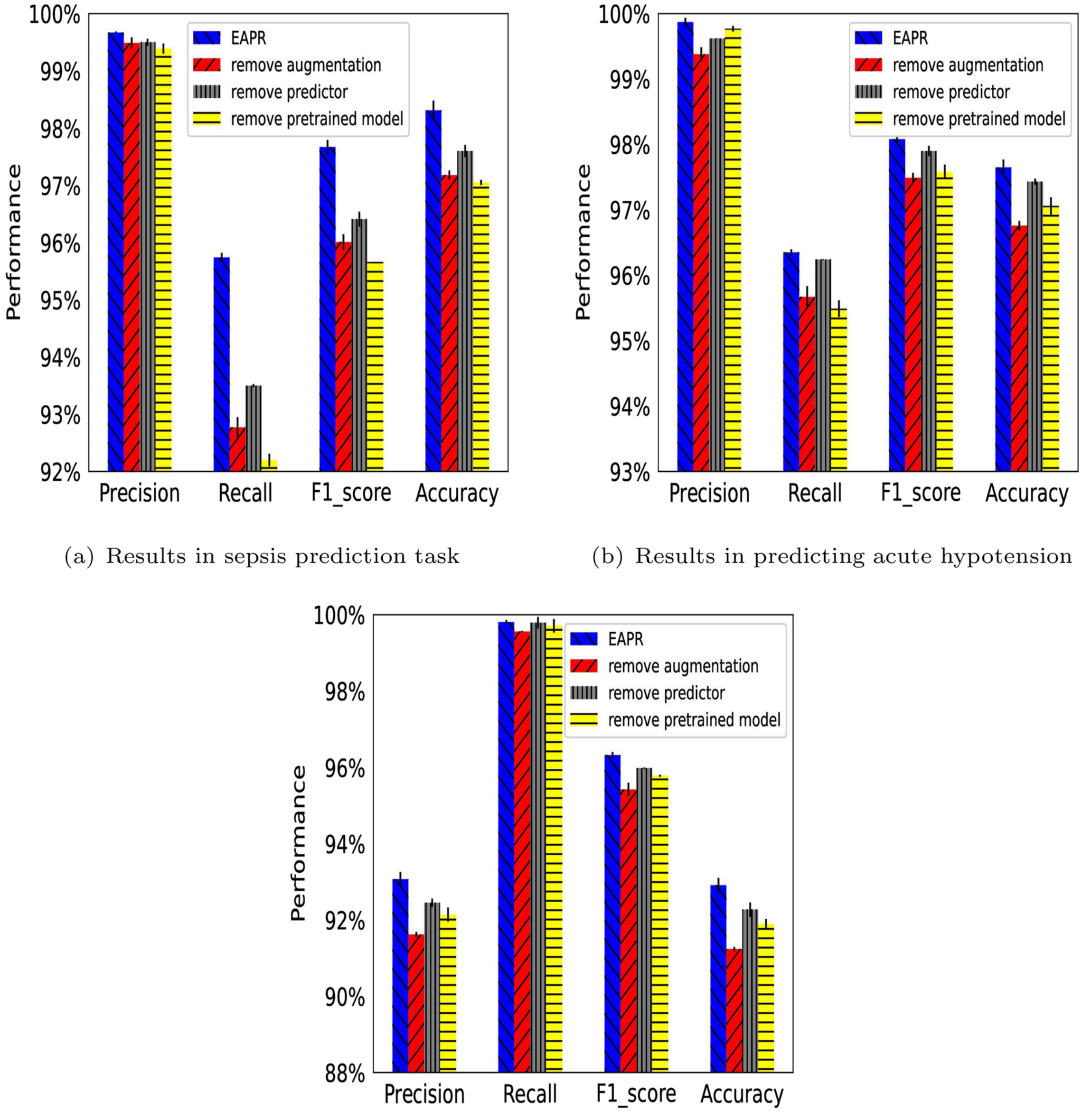


**Fig. 4** Explanation of prediction results of acute hypotension

employing a pre-trained Resnet renders it challenging for our patient representation learning to converge, thereby preventing the acquisition of valuable patient representations. This point can validate a fact that it is pre-trained somehow useful. Therefore, our model can solely produce optimal results when all three components are incorporated.

## Case study

In this section, we applied EAPR to a specific real-world example to demonstrate the effectiveness of the model. We randomly selected a person on the testing sets of sepsis data to predict the sepsis and obtain corresponding explanation results showed in (a) of Fig. 6. Without knowing whether the person will suffer from sepsis, we



(a) Results in sepsis prediction task

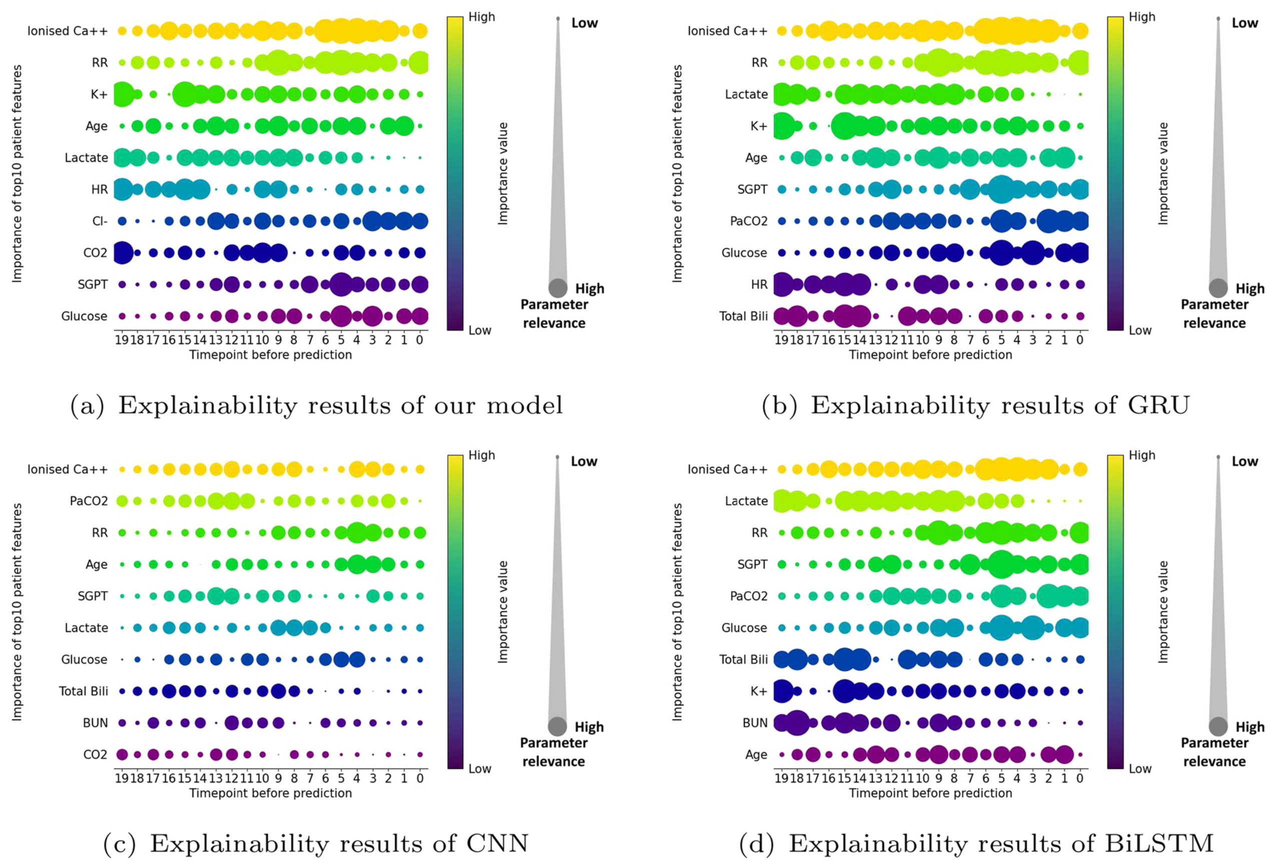
(b) Results in predicting acute hypotension

(c) Results in cancer prediction task

**Fig. 5** Ablations of EAPR with performance results

can calculate the person's representations through EAPR to obtain the probability of developing the disease. And we can also know which of the person's original representations have the greatest impact on suffering from sepsis. As showed in (a) of Fig. 6, Ca++ may have the greatest impact on suffering from sepsis for the person.

To reflect how different explanations could be when we use different disease prediction model, we randomly selected a patient(mentioned above) on the testing sets of sepsis data to predict the sepsis and obtain corresponding explanation results. Experimental results in Fig. 6 show that the three most significant influencing factors on sepsis prediction in this patient are Ionized Ca++, RR, and K+, with only the third largest influencing factor being different from GRU's results. CNN's results show that the three most significant influencing factors are Ionized Ca++, PaCO2, and RR. And regarding the three most influential factors, BiLSTM's results are very close to the results of GRU. Overall, the explanation results of these three selected baseline models clearly do not perform as



(a) Explainability results of our model

(b) Explainability results of GRU

(c) Explainability results of CNN

(d) Explainability results of BiLSTM

**Fig. 6** Comparison of results on explainability

well as our method in the influential factors having relatively lower ranking, as results reflect a greater degree of impact at a longer time point, which goes against common sense. From the analysis, firstly, due to the fact that this comparison only involves one patient, the accuracy of explanation results may not be as high as when testing more patients. Secondly because of the similarity in structure between BiLSTM and GRU, their explanation results also exhibit some similarity. Most importantly, due to the best performance of our model in predicting sepsis, it has also led to a more accurate explanation results.

## Conclusions

This study introduces a novel and general patient representation learning model designed to learn generic patient representations for disease prediction. Specifically, the proposed approach employs data augmentation based on confidence interval control to perform representation learning and obtain comprehensive patient representations. To enhance interpretability, we introduce a two-stage gradient backhaul explanation method to explain the disease prediction results using these patient representations. Experimental results on benchmark datasets reveals that EAPR outperforms many advanced models in disease prediction while ensuring highly explainable results. These findings suggest promising applications of our model in clinical practice.

### Acknowledgements

This research is partially supported by the National Key R &D Program of China 2021YFF0900800, NSFC No.62202279, the Shandong Provincial Natural Science Foundation (No.ZR2022QF018), Shandong Provincial Outstanding Youth Science Foundation (No.2023HWYQ-039), the Fundamental Research Funds of Shandong University.

## Declarations

### Conflict of interest

The authors declare that they have no competing interests or potential conflict.

### Author details

[1] Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Jinan, China. [2] School of Software, Shandong University, Jinan, China. [3] Wuzhou Red Cross Hospital, Wuzhou, China. [4] Jinan University, Jinan, China. [5] Department of Radiology, Qilu Hospital of Shandong University, Jinan, China.

Received: 12 September 2023 Accepted: 17 October 2023
Published: 14 November 2023

## References

1. Wang T, Bendayan R, Msosa Y, Pritchard M, Roberts A, Stewart R, Dobson R. Patient-centric characterization of multimorbidity trajectories in patients with severe mental illnesses. J Biomed Inform. 2022;127:104010.
2. Ma F, Yu L, Ye L, Yao DD, Zhuang W. Length-of-stay prediction for pediatric patients with respiratory diseases using decision tree methods. IEEE J Biomed Health Inform. 2020;24(9):2651–62. https://doi.org/10.1109/JBHI.2020.2973285.
3. Zheng Z, Wang C, Xu T, Shen D, Chen E. Drug package recommendation via interaction-aware graph induction. 2021.
4. Lu H, Uddin S, Hajati F, Moni MA, Khushi M. A patient network-based machine learning model for disease prediction: the case of type 2 diabetes mellitus. Appl Intell. 2022;52(3):2411–22.
5. Fan Y, Tao Z, Lin J, Chen H. An encoder-decoder network for automatic clinical target volume target segmentation of cervical cancer in CT images. Int J Crowd Sci. 2022;6(3):111–6.
6. Yu F, Cui L, Chen H, Cao Y, Liu N, Huang W, Xu Y, Lu H. Healthnet: a health progression network via heterogeneous medical information fusion. IEEE Trans Neural Netw Learn Syst. 2022.
7. Yu F, Cui L, Cao Y, Liu N, Huang W, Xu Y. Similarity-aware collaborative learning for patient outcome prediction. In: International conference on database systems for advanced applications. Springer, Berlin; 2022; p. 407–422.
8. Niu J, Tang Y, Sun Z, Zhang W. Inter-patient ECG classification with symbolic representations and multi-perspective convolutional neural networks. IEEE J Biomed Health Inform. 2020;24(5):1321–32.
9. Han M, Özdenizci O, Koike-Akino T, Wang Y, Erdoğmuş D. Universal physiological representation learning with soft-disentangled rateless autoencoders. IEEE J Biomed Health Inform. 2021;25(8):2928–37. https://doi.org/10.1109/JBHI.2021.3062335.
10. Pokharel S, Zuccon G, Li X, Utomo CP, Li Y. Temporal tree representation for similarity computation between medical patients. Artif Intell Med. 2020;108:101900.
11. Meng Y, Speier W, Ong MK, Arnold CW. Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. IEEE J Biomed Health Inform. 2021;25(8):3121–9. https://doi.org/10.1109/JBHI.2021.3063721.
12. Darabi S, Kachuee M, Fazeli S, Sarrafzadeh M. Taper: time-aware patient EHR representation. IEEE J Biomed Health Inform. 2020;24(11):3268–75. https://doi.org/10.1109/JBHI.2020.2984931.
13. Xie F, Yuan H, Ning Y, Ong MEH, Feng M, Hsu W, Chakraborty B, Liu N. Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies. J Biomed Inform. 2022;126:103980. https://doi.org/10.1016/j.jbi.2021.103980.
14. Wang H, Ahn E, Kim J. Self-supervised representation learning framework for remote physiological measurement using spatiotemporal augmentation loss. 2021.
15. Kim N, Piao Y, Kim S. Clinical note owns its hierarchy: multi-level hypergraph neural networks for patient-level representation learning. In: Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers), p. 5559–5573. Association for computational linguistics, Toronto, Canada. 2023. https://doi.org/10.18653/v1/2023.acl-long.305. https://aclanthology.org/2023.acl-long.305.
16. Daniali M, Galer PD, Lewis-Smith D, Parthasarathy S, Kim E, Salvucci DD, Miller JM, Haag S, Helbig I. Enriching representation learning using 53 million patient notes through human phenotype ontology embedding. Artif Intell Med. 2023;139:102523. https://doi.org/10.1016/j.artmed.2023.102523.
17. Huang Y, Luo F, Wang X, Di Z, Li B, Luo B. A one-size-fits-three representation learning framework for patient similarity search. Data Sci Eng. 2023; p. 1–12.
18. Zhang C, Gao X, Ma L, Wang Y, Wang J, Tang W. Grasp: generic framework for health status representation learning based on incorporating knowledge from similar patients. In: Proceedings of the AAAI conference on artificial intelligence. 2021; vol. 35, p. 715–723.
19. Lu Q, Dou D, Nguyen TH. Textual data augmentation for patient outcomes prediction. In: 2021 IEEE international conference on bioinformatics and biomedicine (BIBM). 2021; p. 2817–2821. https://doi.org/10.1109/BIBM52615.2021.9669861.
20. Deng Y, Lu L, Aponte L, Angelidi AM, Mantzoros CS. Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. NPJ Digit Med.
21. Saranya SS, Fatima NS. IoT-based patient health data using improved context-aware data fusion and enhanced recursive feature elimination model. IEEE Access. 2022;10:128318–35. https://doi.org/10.1109/ACCESS.2022.3226583.
22. Yu L, Xiang W, Fang J, Phoebe Chen Y-P, Zhu R. A novel explainable neural network for Alzheimer's disease diagnosis. Pattern Recogn. 2022;131:108876.
23. Wang S-H, Zhang Y, Cheng X, Zhang X, Zhang Y-D: Psspnn: Patchshuffle stochastic pooling neural network for an explainable diagnosis of covid-19 with multiple-way data augmentation. Comput Math Methods Med 2021.
24. Zhang Y, Zhang X, Zhu W. Anc: Attention network for covid-19 explainable diagnosis based on convolutional block attention module. Comput Model Eng Sci. 2021; p. 1037–1058.
25. Payrovnaziri SN, Chen Z, Rengifo-Moreno P, Miller T, Bian J, Chen JH, Liu X, He Z. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. J Am Med Inform Assoc. 2020;27(7):1173–85. https://doi.org/10.1093/jamia/ocaa053
26. Zhang J, Yu H. Eid: facilitating explainable ai design discussions in team-based settings. Int J Crowd Sci. 2023;7(2):47–54. https://doi.org/10.26599/IJCS.2022.9100034.
27. Shang Z, Meng H, Zhao Y, Xu R, Xu Y, Cui L. Cross-domain credit default prediction via interpretable ensemble transfer. Int J Crowd Sci. 2023;7(3):106–12. https://doi.org/10.26599/IJCS.2023.9100011.
28. Shah R, Kumar V. Rrl: Resnet as representation for reinforcement learning. arXiv preprint arXiv:2107.03380. 2021.
29. Grill JB, Strub F, Altché F, Tallec C, Richemond PH, Buchatskaya E, Doersch C, Pires BA, Guo ZD, Azar MG. Bootstrap your own latent: a new approach to self-supervised learning. 2020.
30. Ding X, Zhang X, Han J, Ding G. Scaling up your kernels to 31×31: revisiting large kernel design in CNNS. In: 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2022; p. 11953–11965. https://doi.org/10.1109/CVPR52688.2022.01166.
31. Aitken K, Ramasesh VV, Garg A, Cao Y, Sussillo D, Maheswaranathan N. The geometry of integration in text classification RNNS. In: International conference on learning representations. 2021.
32. De Brouwer E, Simm J, Arany A, Moreau Y. Gru-ode-bayes: continuous modeling of sporadically-observed time series. Advances in neural information processing systems **32**. 2019.
33. Abdul W, Alsulaiman M, Amin SU, Faisal M, Ghaleb H. Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM. Comput Electr Eng. 2021;95(6):107395.
34. Steinberg E, Jung K, Fries JA, Corbin CK, Pfohl SR, Shah NH. Language models are an effective representation learning technique for electronic health record data. J Biomed Inform. 2021;113:103637.
35. Zhang E, Robinson R, Pfahringer B. Deep holistic representation learning from ehr. In: 2018 12th international symposium on medical information and communication technology (ISMICT). 2018. https://doi.org/10.1109/ISMICT.2018.8573698.
36. Jaume-Santero F, Zhang B, Proios D, Yazdani A, Gouareb R, Bjelogrlic M, Teodoro D. Cluster analysis of low-dimensional medical concept representations from electronic health records. In: International conference on health information science. 2022.
37. Wang Y, Wu T, Wang Y, Wang G. Enhancing model interpretability and accuracy for disease progression prediction via phenotype-based patient similarity learning. In: Pacific symposium on biocomputing 2020. World Scientific. 2019; p. 511–522.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.