2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 1-4, 2023, Oahu, Hawaii, USA

FGRL-Net: Fine-Grained Personalized Patient Representation Learning for Clinical Risk Prediction Based on EHRs

KaKit Chio[1], Wenhao Zhu[2], Lihua He[1], Dian Zhang[3], Xu Yang[1], Wuman Luo[1,*]

***Abstract*— Personalized patient representation learning (PPRL) is a critical element in clinical risk prediction. It aims to obtain a complete portrait of each patient based on Electronic Health Records (EHR). Although existing works have achieved remarkable progress in healthcare prediction, there are still three major issues. First, feature correlation is crucial for risk prediction, but it has not yet been fully exploited by existing works. Second, variation pattern of dynamic feature contains useful information about patient's physical status, but adaptive pattern recognition is still a challenge. Third, existing works usually adopt a two-stage embedding process to process each dimension of the EHR data. However, some useful low-level information for PPRL will be lost. To address these issues, in this paper, we propose a fine-grained PPRL architecture named FGRL-Net for clinical risk prediction based on EHR. Specifically, we propose a Medical Feature Correlation Detection Module (FCM) to effectively learn the feature correlations for each patient and a Temporal Variation Pattern Recognition Module (TVM) to effectively detect the variation patterns of each dynamic feature. Moreover, we design a Fine-Grained Representation Mechanism (FGRM) to preserve the low-level information (from both feature and visit dimensions) useful for risk prediction. In addition, in the stage of data preprocessing, We utilize generic medical classification knowledge to classify numerical dynamic data. We conduct the in-hospital mortality experiment and the decompensation experiment on a real-world dataset. The experiment results show that the FGRL-Net outperforms state-of-the-art approaches. The source code is provided in github https://github.com/JackyChio/FGRL-Net.**

I. INTRODUCTION

Nowadays, the widespread use of Electronic Health Records (EHR) has enabled the popularity of personalized clinical risk prediction. Typically, EHR data is composed of a time series of patient visit records. Each visit record consists of a large number of medical features, which can be divided into three groups: demographic information (e.g., gender and age), primary diseases (e.g., diabetes and heart failure), and dynamic features (e.g., temperature and albumin). A critical problem of personalized clinical risk prediction is how to effectively detect and represent each patient's portrait, which is called personalized patient representation learning (PPRL) [1] [2]. In this paper, we focus on PPRL for clinical risk prediction based on EHR.

So far, various PPRL methods based on EHR have been proposed for clinical risk prediction. However, these works still face three major issues:

1) Correlation detection of medical features: Medical features are usually highly correlated [3]. For example, patients with diabetes tend to have eye diseases [4]. Some existing works [5] [6] [7] [8] adopted attention-based mechanism to study the importance of each feature, which is not suitable for feature correlation detection. Several recently proposed work [9] [10] used convolutional neural networks (CNN) to study feature correlations. Particularly, StageNet adopted the average method [11] to calculate the stage-weight of features. AEFNet used 2D CNN networks with maxpolling for feature recalibration. The major problem of these two methods is that, after averaging or 2D maxpolling, some low-level medical information that is important for risk prediction is lost.

2) Variation pattern recognition of dynamic features: The dynamic feature of EHR usually exhibits regular temporal variation patterns, which is very helpful for risk prediction [12] [13]. AdaCare [8] used three single-layer dilated one-dimensional (1D) CNN to capture the changing patterns of the dynamic features. AEFNet [10] improved AdaCare by using 2D maxpooling for downsampling and upsampling for each feature. However, The variation pattern detection of AEFNet suffered from the same problem caused by 2D maxpooling as discussed previously.

3) Two-stage embedding for representation learning: Due to the inherent 2D structure of EHR, patient representation learning has to process the data in both visit and feature dimensions. We observe that existing methods of this area typically employs two types of two-stage embedding processes, namely visit-feature and feature-visit. Visit-feature process first embeds all features of each visit seperately and then deals with the embedded temporal visits [6] [5] [8] [10]. Feature-visit process first embeds the time series of each feature separately and then extracts the relationships between the embedded features [7] [9].

The main issue of the two-stage embedding is that some low-level information important for risk prediction will be lost after the first embedding stage. For visit-feature process, temporal variation information of each feature cannot be fully retained after visit embedding, which could potentially affect the performance of risk prediction. For example, assume that a patient has been diagnosed with diabetes and respiratory arrest. Respiratory arrest is acute and will be fatal if not

* denotes the corresponding author
[1]KaKit Chio, Lihua He, Xu Yang and Wuman Luo are with Faculty of Applied Sciences, Macao Polytechnic University, Macao S.A.R, China kakit.chio, lihua.he, xuyang, luowuman@mpu.edu.mo
[2]Wenhao Zhu is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Sichuan Province, China zhu_wenhao@std.uestc.edu.cn
[3]Dian Zhang is with Department of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China zhangd@szu.edu.cn

2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 979-8-3503-3702-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/SMC53992.2023.10393910

treated immediately, while diabetes is chronic and its importance increases over time. However, after visit embedding, the temporal dynamic information of both these features will be weakened. For feature-visit process, it is difficult to detect the exact time when a feature becomes abnormal, which may lead to failed or delayed risk prediction. For example, research has demonstrated that patients who got acute kidney injury (AKI) and elevated troponin levels simultaneously have a significantly higher risk of mortality in comparison to those with either one or none of these two abnormalities [14]. However, after feature embedding, the co-occurrence of these features will be difficult to capture.

To address these issues, in this paper, we propose a PPRL architecture named FGRL-Net for clinical risk prediction based on EHR. FGRL-Net can adaptively capture the medical feature correlations in the visit dimension and the variation patterns of features in the feature dimension. To overcome the problem of two-stage embedding, FGRL-Net adopts Fine-Grained Representation Mechanism (FGRM) to maintain the useful low-level information detected in both feature and visit dimensions. To better represent patient's medial status, we introduce generic classification knowledge [15]–[21] to classify the numerical dynamic features in the data preprocessing stage. The contributions of this paper are as follows:

- We propose an PPRL architecture named FGRL-Net for clinical risk predication based on EHR. FGRL-Net uses the FGRM to retain the low-level information (from both feature and visit dimensions) useful for risk prediction.
- We propose a Medical Feature Correlation Detection Module (FCM) to effectively learn the feature correlations for each patient and a Temporal Variation Pattern Recognition Module (TVM) to effectively detect the variation patterns of each dynamic feature.
- We introduce the generic medical classification knowledge to classify the numerical dynamic features in the preprocessing stage. The classified results will be used as part of input for patient representation learning.
- We conduct extensive experiments on the real-world MIMIC III dataset for two key patient clinical prediction tasks: in-hospital mortality and decompensation. We use AUROC, AUPRC, and min(Se, P+) as evaluation metrics. The experimental results demonstrate that our proposed model outperforms the baseline models.

The remainder of this paper is organized as follows. We review the related works of our research topic in Section II. In Section III, we provide a detailed description of the FGRL-Net model. We evaluate the performances of FGRL-Net in Section IV, and conclude our work in Section V.

II. RELATED WORK

In this section, we discuss the methods closely related to PPRL for risk prediction based on EHR.

Correlation detection of medical features: To achieve a better representation of patients, many methods have been proposed to study the relationship between medical features, including attention-based, CNN-based and RNN-based methods. Typical examples are RETAIN [5], Dipole [22], SAnD [6], ConCare [7] and COVIDCare [23]. RETAIN was proposed to discover correlations between medical features by using a two-level RNN to learn temporal attention and feature attention. Dipole was an end-to-end, simple and robust model for patients' future health prediction. It combined bidirectional RNN architecture with the attention mechanism to study the relationships of patients' past visits. SAnD used CNN to embed visits and then used a self-attention mechanism, a dense interpolation strategy, and positional encoding to make prediction. ConCare obtained the temporal embedding by time-aware attention and then study the interdependencies between dynamic features and static baseline information by cross-head decorrelation. COVIDCare fed the time-series medical data to GRU to generate the feature embedding and then employs multi-head self-attention and transfer learning mechanisms to obtain patient representation. Note that we do not compare with Dipole, SAnD and COCIDCare in the experiments. Dipole and SAnD are early attention-based methods that have been adequately compared with ConCare. COVIDCare focuses on embedding COVID-19-related medical features to handle transfer learning, but our study does not target these medical knowledge.

Variation pattern recognition of dynamic features: Several methods such as T-LSTM [24], Health-ATM [25], StageNet [9] have been proposed to deal with the the time series of EHR. T-LSTM addressed the issue of visit-level irregular time intervals by introducing time decay. Health-ATM used attentive and Time-Aware Module (ATM) and a hybrid CNN&RNN network to extract time series information. StageNet calculated the average progression theme within an observation time window and then utilized a channel-wise attention mechanism like to obtain patient representation. In 2016, Cheng et al. [26] proposed 1D CNN to learn the variation patterns of the dynamic features of EHR. AdaCare [8] improved Cheng et al. [26] by using three dilated 1D CNNs to capture the long-term and the short-term changing patterns of the dynamic features. Besides, AdaCare used a fully connected network to learn the correlations of the medical features. AEFNet [10] improved AdaCare by using 2D Maxpooling for downsampling and upsampling for each feature. Note that T-LSTM, Health-ATM and the method proposed by Cheng et al are very early methods, and their problems have been discussed in detail in the works [8] [9].

Integrated with prior medical knowledge: Several researchers have attempted to enhance the prediction performance of clinical risk prediction tasks by integrating external medical knowledge, such as medical knowledge graphs or prior medical knowledge. A medical knowledge graph contains relationships between diseases or disease and its ancestors, including [27]–[31]. GRAM [27] uses co-occurrence matrix to study diseases and their ancestors and generate a graph-based attention model for prediction. Mime [28] proposed a Multilevel Medical Embedding method to improve GRAM. KAME [30] improves GRAM by proposing a knowledge attention mechanism to obtain

TABLE I
RESPIRATORY RATE CLASSIFICATION TABLE

Reference Range	Label	One-hot vector
>20	High	001
>= 12 and <= 20	Norma	010
<12	Low	100

better prediction accuracy. Additionally, prior knowledge can be provided by specific prediction tasks or authoritative sources [32], [33]. MetaPred [32] aims to employ meta-learning to integrate domain knowledge from related medical prediction tasks. None of the works mentioned above are suitable for classifying dynamic features, as they are specific to a particular task. Therefore, we do not use them as our baseline model.

III. METHODOLOGY

In this section, we first formulate the problem of clinical risk prediction. Then, we present the proposed FGRL-Net in detail.

A. Problem Formulation

In our work, we use EHR data as input for PPRL.

EHR: Let $X = \{x_1, x_2, ..., x_i\}$ be the visit records of patients, where x_i is a patient's hospitalization record. Specifically, $x_i = \{v_1^{(i)}, v_2^{(i)}, \ldots, v_t^{(i)}\}$, where $v_j^{(i)}$ is patient i's j-th test records. that Let $v_w^{(i)} = \{D^{(i)}, P^{(i)}, r_{(w,1)}^{(i)}, \ldots, r_{(w,n)}^{(i)}\}$, where $D^{(i)}$ is the demographic information of patient i and $P^{(i)}$ is the diagnosed diseases of patient i. $P^{(i)} \in \{0,1\}^{|c|}$ is a multi-hot binary vector, where $|C|$ is the number of diseases. $D^{(i)}$ and $P^{(i)}$ are static features and do not change during a visit. $\{r_{(w,1)}^{(i)}, \ldots, r_{(w,n)}^{(i)}\}$ are dynamic medical test features. The number of dynamic test features is n, and the total test times of a patient is t in one visit. Thus, x_i can be denoted as follow:

$$x_i = \begin{pmatrix} v_1^{(i)} \\ v_2^{(i)} \\ \vdots \\ v_t^{(i)} \end{pmatrix} = \begin{pmatrix} D^{(i)} & P^{(i)} & r_{1,1}^{(i)} & \ldots & r_{1,n}^{(i)} \\ D^{(i)} & P^{(i)} & r_{2,1}^{(i)} & \ldots & r_{2,n}^{(i)} \\ \vdots & & & & \\ D^{(i)} & P^{(i)} & r_{t,1}^{(i)} & \ldots & r_{t,n}^{(i)} \end{pmatrix} \quad (1)$$

Problem Statement: We can define the clinical risk prediction problem, given $X = \{x_1, x_2, ..., x_i\}$, aim to predicting whether the patient will suffer from the target health risk after a certain time (e.g., 48 hours) in the future. This is a binary classification problem, namely $\hat{y} = \mathbb{F}(X)$, where $\hat{y} \in \{0, 1\}$.

B. FGRL-Net

The framework of FGRL-Net is shown in Fig. 1. FGRL-Net consists of three stages, namely (a) Medical Data Preprocessing, (b) Patient Representation Learning, and (c) Risk Prediction. In stage (a), we preprocess each patient's hospitalization record x_i, which contains demographic information, primary diseases and dynamic features, and organize the preprocessing results into a matrix shown in equation 1. At the end of stage (a), $X = \{x_1, x_2, ..., x_i\}$ is created. In stage (b), X is fed into two parallel modules, i.e., FCM and TVM, to obtain feature representation and temporal representation, respectively. In stage (c), the feature representation and the temporal representation are fused together to obtain the patient representation matrix. After that, this matrix is fed to the GRU module for clinical risk prediction.

1) Medical Data Preprocessing: As shown in Fig. 1(a), in dataset MIMIC-III, for each patient's hospitalization information x_i, we first normalize its three components, i.e, demographic information, primary diseases and dynamic features, into vector $D^{(i)}$, vector $P^{(i)}$, and matrix $R^{(i)}$. The entry $r_{(w,v)}^{(i)}$ of $R^{(i)}$ denotes the w-th test results of the v-th dynamic feature, $1 \leq w \leq t$, $1 \leq v \leq n$.

Then we attach the medical classification information to $R^{(i)}$. The dynamic features can be divided into two types, i.e, numerical features and categorical features. We classify the numerical data in term of medical knowledge, and encode them using one-hot encoding. In particular, we classify 7 numerical dynamic features in dataset MIMIC-III, i.e., glucose [18], heart rate [16], oxygen saturation [17], respiratory rate [15], systolic blood pressure [19], temperature [20] and pH [21]. As an example, Table I shows the classification knowledge of respiratory rate. The classification results is a h-dimensional vector, which is attached to the end of each row of $R^{(i)}$. As a result, $R^{(i)}$ is transformed into a $t \times (n+h)$ matrix $\tilde{R}^{(i)}$.

At last, we organize $D^{(i)}$, $P^{(i)}$ and $\tilde{R}^{(i)}$ into the matrix $\tilde{x}_i$ with size $t \times (n+h)$. In fact, $\tilde{x}_i$ is the x_i (in equation 1) augmented with classification encoding.

2) Patient Representation Learning: In this stage, the preprocessed data $\tilde{X} = \{\tilde{x}_1, ..., \tilde{x}_i\}$ will be input into two modules, i.e., FCM and TVM. Before introducing FCM and TVM in detail, we will first discuss the proposed fine grained representation mechanism for the fusing of the data from the visit and the feature dimensions.

FGRM: As discussed in Section I, one of the tasks of the FGRL-Net is to avoid the problem caused by the two-stage embedding. To this end, we design an approach called FGRM to maintain the low-level useful information detected from both visit and feature dimensions, as shown in Fig. 2. The basic idea is to use a 1D CNN network to learn a scalar pattern matrix and then multiply it with the original data to obtain an enhanced patient representation. The design philosophy behind this is: 1) use 1D CNNs instead of 2D CNNs. Although the dimensionality of EHR data is two (i.e., n features and t visits), which resembles a 2D image, it is essentially a time-series 1D signal. Therefore, processing EHR data with 1D CNNs is more suitable than using 2D CNNs [34]. 2) To maintain low-level useful information, we utilize multi-channel 1D CNN and preserve the output of each convolution stride to compute a pattern for each patient individually. This Fine-Grained scalar pattern allows our model to generate more individualized representations, thereby improving the model's ability to capture patient-specific characteristics. 3) For learning feature correlations, some methods use two fully connected (FC) layers. To reduce

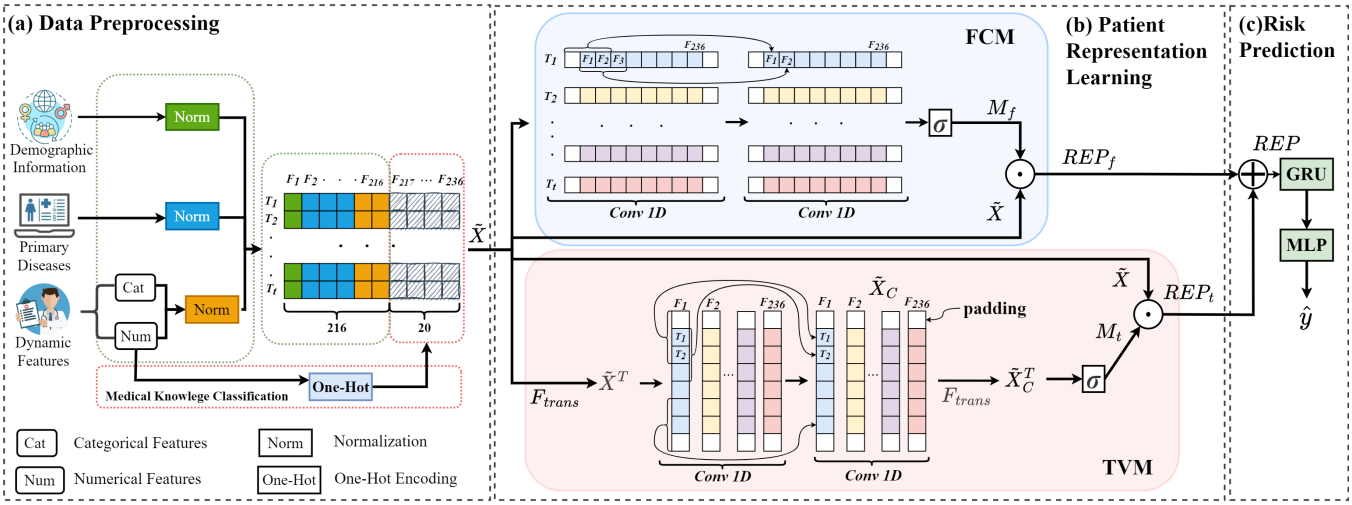


Fig. 1. The Framework of FGRL-Net

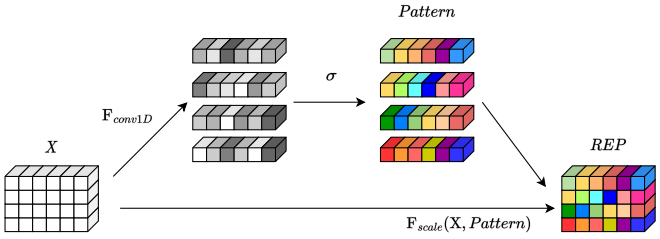


Fig. 2. Fine-Grained Representation Mechanism (FGRM): X denotes the input EHR data. F_{conv1D} denotes the two layer multi-channel 1D CNN networks. σ denotes the sigmoid activation function. After the convolution operation of F_{conv1D} and the activation function σ, the original EHR data are transformed to the fine-grained scale matrix called $Pattern$. $\mathrm{F}_{scale}(.,.)$ denotes the dot product function. After $\mathrm{F}_{scale}(X, Pattern)$, we get the fine-grained patient representation REP. Note that the FGRM can be used in both visit and feature dimensions of EHR data. For the visit dimension, each row of $Pattern$ is the feature scale of each visit. For the feature dimension, each row of $Pattern$ is the variation pattern of each feature.

the number of model parameters and enable more efficient feature correlation learning, we replace these FC layers with two 1D CNN layers.

Base on the FGRM, we design FCM and TVM module to extract the visit-level enhanced representation and the feature-level enhanced representation, respectively. Finally, we fuse these two representations to obtain the patient representation. By FGRM, we can preserve the low-level useful information and solve the problem of information loss caused by two-stage embedding.

FCM: The purpose of this module is to learn the correlation between medical features. As discussed previously, the preprocessed patient's hospitalization information $\tilde{x}_i$ is a two dimensional matrix, where each row denotes the medical data of one visit. Therefore, we input each row of $\tilde{x}_i$ into a multi-layer 1D CNN to learn the correlation of the features at a certain time. We observe that the feature correlations for the same visit can better represent the patient's health status. The output is stored in a feature correlation matrix M_f shown below.

$$M_f = \sigma(\mathbb{F}_{conv1D}(\tilde{X})) \tag{2}$$

where $\mathbb{F}_{conv1D}(\cdot)$ denotes a two layer multi-channel 1D CNN network, and σ is the sigmoid activation function. Specifically, the different timestamps of the temporal dimension are used as the convolution channels. Besides, we set the filter kernel size to 3, the stride to 1 and the padding to 1 for the first layer of 1D CNN, and the filter kernel size to 5, the stride to 1 and the padding to 2 for the second layer. The purpose is to make sure that the output dimension of each layer remains the same as the input dimension. As shown below, the feature representation REP_f can therefore be calculated as the dot product of M_f and $\tilde{X}$. The purpose is to increase the weights of important features, while reduce the weights of unimportant features.

$$REP_f = M_f \odot \tilde{X} \tag{3}$$

TVM: The purpose of TVM is to learn the variation patterns of each dynamic feature. We perform 1D convolution on each feature. So we transpose each matrix in $\tilde{X}$, and use $\tilde{X}^T$ to denote the transpose results.

$$\tilde{X}^T = \mathbb{F}_{transpose}(\tilde{X}) \tag{4}$$

We regard different features as the convolution channels. The settings of the multi-layer 1D CNN are the same as in FCM. As discussed previously, pooling is not recommended after each convolution layer. The output of the multi-layer 1D CNN is $\tilde{X}_C$.

$$\tilde{X}_C = \mathbb{F}_{conv1D}(\tilde{X}^T) \tag{5}$$

Then, we transpose $\tilde{X}_C$ again to obtain temporal correlation matrix M_t.

$$M_t = \sigma(\mathbb{F}_{transpose}(\tilde{X}_C) \tag{6}$$

Finally, the temporal representation REP_t can therefore be calculated as the dot product of M_t and $\tilde{X}$. The purpose is the same as in FCM.

$$REP_t = M_t \odot \tilde{X} \tag{7}$$

3) Risk Prediction: At this stage, we merge REP_f and REP_t to obtain the patient representation REP.

$$REP = REP_f + REP_t \tag{8}$$

Then, we feed REP into the GRU layer to obtain the hidden representations h_t:

$$h_t = GRU(h_{t-1}, REP) \tag{9}$$

where h_{t-1} is the hidden representation of time step $t-1$.

Finally, the risk prediction result $\hat{y}$ is calculated as follows:

$$\hat{y} = \sigma(W_y h_t + b) \tag{10}$$

where W_y and b are the weight matrix and bias, respectively.

We choose the Binary Cross Entropy function to calculate the loss as follows:

$$loss = -(ylog(\hat{y}) + (1-y)log(1-\hat{y})) \tag{11}$$

where y is the label of dataset, and $\hat{y}$ is the predicted label.

IV. EXPERIMENT

In this section, we conduct the in-hospital mortality prediction experiment and decompensation prediction experiment on the dataset MIMIC-III [35]. The source code will be published as soon as the paper is accepted.

A. Experiment Settings

1) Dataset: The MIMIC-III is available at [35]. It contains the clinical records of approximately 50,000 patients from the Intensive Care Unit of Beth Israel Deaconess Medical Centre between 2001 and 2012. Our experiments are based on the preprocessed data of the MIMIC-III provided by the benchmark study [36]. Specifically, the preprocessed data contains three categories, i.e, patient demographics (12 dimensions), primary diagnosis codes (128 dimensions), and dynamic features (76 dimensions).

2) Two Tasks: **In-hospital mortality task** aims to predict whether a patient will die after 48 hours given 48 hours of monitoring data. The preprocessed data for this task contain 21,138 patients, of whom 2,797 died (i.e., positive rate is 13.23%). And each patient has 48 records with only one label indicating die or not. **Decompensation task** aims to predict whether a patient will die after 24 hours given a fixed time window of monitoring data. The preprocessed data for this task contain 33,678 patients. Unlike the in-hospital mortality task, the number of test records for each patient in this task varies, and each record of a patient has a label indicating whether the patient is dead or not (2,202,114 records in total with positive rate 4.2%). We note that the number of records for each patient is no greater than 400. For those whose number of records less than 400, we pad 0's to make sure the total number of records of each patient is 400. After convolution, we delete the redundant rows of 0's.

3) Evaluation Metrics: Both tasks are binary classification tasks, and exhibit significant label imbalance. This makes the evaluation metrics such as accuracy, precision and recall no longer suitable [37]. Therefore, we adopt Area Under Receiver Operating Characteristic curve (AUROC) [38], Area Under Precision-recall curve (AUPRC) [39], and Minimum of precision and sensitivity (min(Se, P+)) to evaluate the performances of the methods.

4) Baselines: To compare FGRL-Net against other EHR analysis methods, we choose five mostly related approaches as baselines, namely RETAIN, ConCare, StageNet, AdaCare, and AEFNet.

- **RETAIN** [6] employed two RNNs to learn time attention and feature attention, and sumed up the weighted visit embeddings to perform the prediction.
- **ConCare** [7] used a time-aware attention to detect the important values of the dynamic features generated at past times, and used Cross-Head Decorrelation to capture the relationship between the medical features.
- **StageNet** [9] designed a stage-aware neural network model to extract disease progression stage information.
- **AdaCare** [8] used multiple scale dilated CNNs to capture variation patterns of the dynamic features, and then used feature recalibration to extract clinical representations.
- **AEFNet** [10] adopted the techniques of upsampling, downsampling and 2D CNN to learn patient feature representations.

5) Implementation Details: We implement the proposed FGRL-Net and the 6 baselines on PyTorch 1.12.1 and CUDA Version 11.4. The training work is done on Intel(R) Xeon(R) Gold 6138 CPU @ 2.00GHz(80 cores), Nvidia Tesla P40 with 22929MB Ram. We use the Adam optimizer with learning rate set to 1e-4. Batch size is set to 256 for the in-hospital mortality task and 128 for the decompensation task.

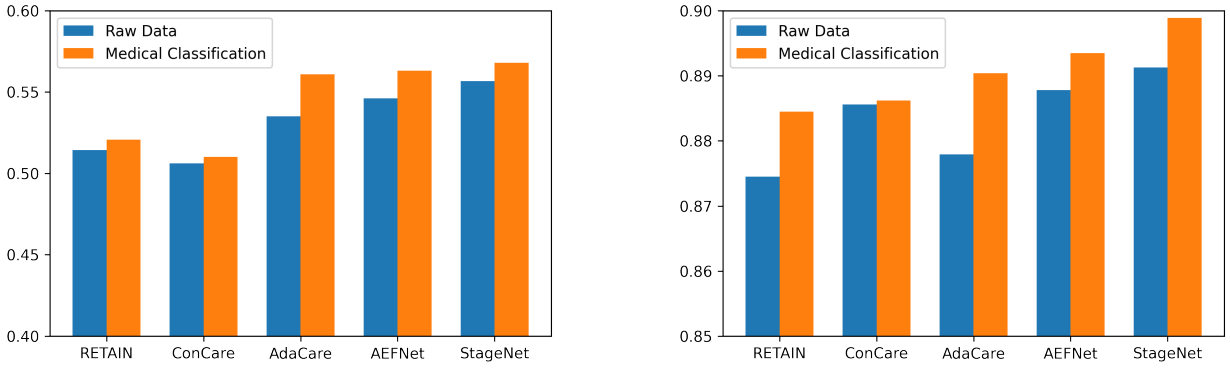
B. Results and Analysis

Table II shows the experiment results of our proposed FGRL-Net and the 5 baseline models in terms of AUROC, AUPRC and min(Se, P+) based on the MIMIC-III. The Table II consists of two parts. The upper part is the results of baselines, and the lower part is the results of ablation. We can see that our proposed FGRL-Net outperforms all the baselines and the variants of FGRL-Net, i.e., FGRL-Net$_{F,T,\text{-}}$, FGRL-Net$_{F,\text{-},C}$ and FGRL-Net$_{\text{-},T,C}$.

TABLE II

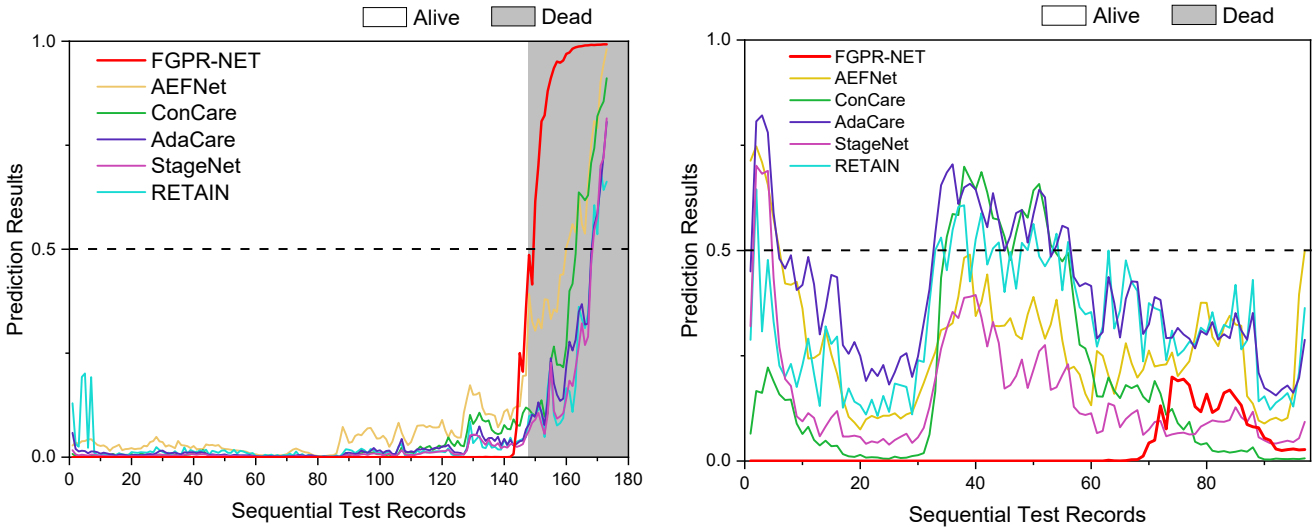
RESULTS OF HEALTH RISK PREDICTION.

Model	In-hospital mortality(IHM)			Decompensation		
	AUROC	AUPRC	min(+P, Se)	AUROC	AUPRC	min(+P, Se)
RETAIN [6]	0.8745	0.5143	0.5267	0.8936	0.2656	0.3187
ConCare [7]	0.8856	0.5062	0.4882	0.9065	0.3318	0.3626
AdaCare [8]	0.8779	0.5351	0.5241	0.9129	0.3271	0.3583
AEFNet [10]	0.8878	0.5462	0.5427	0.9121	0.3335	0.3774
StageNet [9]	0.8913	0.5567	0.5425	0.9111	0.3145	0.3585
FGRL-Net$_{F,T,\text{-}}$	0.8852	0.5584	0.5481	0.9834	0.7020	0.6886
FGRL-Net$_{F,\text{-},C}$	0.8893	0.5664	0.5464	0.9828	0.7107	0.6735
FGRL-Net$_{\text{-},T,C}$	0.8930	0.5614	0.5340	0.9231	0.4046	0.4375
FGRL-Net	**0.8934**	**0.5739**	**0.5587**	**0.9886**	**0.7953**	**0.7323**



(a) AUPRC (b) AUROC

Fig. 3. AUPRC and AUROC of different baselines on in-hospital mortality task



(a) Case 1 (b) Case 2

Fig. 4. Case Study

1) Results of Baselines: Table II shows that FGRL-Net outperforms the baselines in terms of all three metrics in both prediction tasks. In in-hospital mortality task, we observed that StageNet get better performances than all the other baselines in terms of AUROC (0.8913) and AUPRC (0.5567). The reason is that StageNet can automatically learn the stages of disease progression. Compared with StageNet, FGRL-Net improves AUROC by 0.24% and AUPRC by 3.09%, respectively. This verifies that FGRL-Net can better capture the variation patterns of the dynamic features and thus build

more complete portrait of each patient. Besides, the AEFNet get best result in terms of min(Se, P+) (0.5425). This is because the 2D CNN can capture more complete temporal information of each clinical record. However, 1D multi-layer CNNs by regarding features as convolution channels are more suitable for EHR data compared with 2D CNN. As shown in Table II, FGRL-Net improves the min(Se, P+) by 2.99% compared with that of AEFNet.

In decompensation task, as shown in Table II, FGRL-Net achieves a significant improvement in AUROC, AUPRC and min(Se, P+) by 8.39%, 138.47%, and 94.04%, respectively. The reason is that the number of labeled records of the preprocessed data for decompensation task (2,202,114 labeled records in total with positive rate 4.2%) is significantly greater than that for in-hospital mortality task (21,138 labeled records in total with positive rate 13.23%). Thus, the 1D multi-layer CNN of FCM is able to capture more useful information for feature correlation detection. On the contrary, the baselines cannot make good use of the large number of the labeled records. They either used fully connected layer or 2D CNN, or even ignore the correlations of medical features.

To further evaluate the rationality of the prediction results of FGRL-Net, we calculate the confusion matrix for decompensation task using FGRL-Net. The results of the confusion matrix are as follow: True Negative is 464800, False Negative is 2955, False Positive is 1679, and True Positive is 6048. The precision class 1 (positive value) is therefore 0.7827, and recall class 1 (positive value) is 0.6717. This shows that there is no extreme conditions in the test data which would lead to unreasonable results.

2) Results of Ablation: To evaluate the effectiveness of each component of FGRL-Net, we propose three variants of FGRL-Net. As shown in Table II, FGRL-Net$_{F,T,_}$ denotes the variant removing the classified numerical dynamic data, FGRL-Net$_{F,_,C}$ denotes the variant removing the TVM, FGRL-Net$_{_,T,C}$ denotes the variant removing the FCM, and FGRL-Net denotes the model with all components. We can see that FGRL-Net outperforms all the variants, and the above table shows the positive contribution of each component in the model. Even if one component is removed, the performance is still better than all the baseline models. In the task of in-hospital mortality, the results of FGRL-Net$_{F,T,_}$ show that the classification of numerical dynamic data using medical classification knowledge is of most importance. In the task of decompensation, the results of FGRL-Net$_{_,T,C}$ show that the FCM is of most importance.

3) Generality of medical knowledge classification: To verify the importance of the numerical dynamic data classification using medical knowledge, we apply the preprossing stage of FGRL-Net to each of the baselines, and compare each pair of the original baseline and its variant with the preprocessing stage. As shown in Fig. 3, after numerical data classification, the performance of each baseline has been improved. This verifies the effectiveness and the generality of the numerical dynamic data classification.

4) Case Study: To further evaluate the effectiveness of our proposed model, we conduct case study of two types of patients in the decompensation task and visualize the prediction results. Case 1: After admission, the patient was first alive, then the patient's condition suddenly got worse and eventually died. Case 2: The patient was stable and alive throughout the monitoring phase. The purpose of the case study is to evaluate whether the models can predict the patient's different status both accurate and in time. As shown in Fig. 4, the x-axis is the patient's sequential monitoring records. The decompensation task will predict whether the patient is alive (=0) or dead (=1) for each newly arrived monitoring record. The y-axis is the predicted result (between 0 and 1) corresponding to each newly arrived test record. Besides, we set the threshold to 0.5. Thus, the predicted value greater than 0.5 indicates the patient "dead", otherwise "alive". In both cases, we compare the proposed FGRL-Net with all the baselines.

Fig. 4 (a) shows the prediction results of Case 1 (dead). Specifically, the patient (testset number 4927_episode1) was initially alive, and all the models made correct predictions. However, we can see from the figure that the prediction results of the FGRL-Net are much closer to the true value (=0) of the label than all the baseline models. On the arrival of the 150th record, the patient died, and the FGRL-Net made the correct prediction immediately. However, AEFNet, which predicted patient death the earliest among all the baselines, did not make a correct prediction until the arrival of the 160th record. Although Fig. 4 (a) shows great increases in the prediction values of both FGRL-Net and AEFNet after the 140th record, the results of FGRL-Net grow much faster and greater than those of AEFNet. To explain this observation, we checked the testset of this patient. We found that the respiratory rate value of this patient changed to 0 at the arrival time of the 140th record, which could be a transient respiratory arrest with a high fatality rate. Apart from the respiratory rate, we also found that some other risk indicator values (like Glascow coma scale eye opening) [40] increase after the 140th record.

Fig. 4 (b) shows the prediction results of Case 2 (survival). The patient was alive during the entire surveillance phase. We can see that the predictions of FGRL-Net are correct throughout the surveillance phase, but all the baseline models are quite unstable. During the monitoring period, all the baseline models made more or less incorrect predictions (i.e., the patent is predicted to be dead), mainly concentrated in the range of the first 8 records and the 12th to 58th records. In contrast, FGRL-Net made the correct predictions throughout the experiment. Moreover, the predicted values of FGRL-Net are close to 0 (i.e., the true value of the label) for most of the test records, and show a little upward fluctuation for only a small number of the test records. In particular, all predictions of FGRL-Net were lower than those of AEFNet, AdaCare and RETAIN, and most of the predictions were lower than those of Concare and StageNet. Note that, in this case, the smaller the predicted value the better.

In summary, FGRL-Net performs better than all the baselines in both cases. This verifies the effectiveness of the patient representation of FGRL-Net. Note that the value of AUPRC is calculated from the actual predicted numerical

value instead of the binary value of 'correct' or 'wrong'. Therefore, this also explains why FGRL-Net achieves a significant improvement in terms of the three metrics in the task of decompensation.

V. CONCLUSION

In this paper, we propose a fine-grained PPRL architecture named FGRL-Net for clinical risk predication based on EHR. we propose a FGRM to preserve the low-level useful information in both visit and feature dimensions. Besides, We propose FCM to learn feature representation and TVM to learn temporal representation. In addition, we adopt the medical classification knowledge to preprocess the numerical dynamic features. We conduct the in-hospital mortality prediction and decompensation prediction on the real-world dataset MIMIC III. The experiment results show that our proposed model outperforms the baselines.

ACKNOWLEDGMENT

This work was supported in part by the Macao Polytechnic University (RP/FCA-11/2022) and Shenzhen University (JCYJ20220531091407016, NSFC 61872247).

REFERENCES

[1] Y. An, N. Huang, X. Chen, F. Wu, and J. Wang, "High-risk prediction of cardiovascular diseases via attention-based deep neural networks," *IEEE/ACM transactions on computational biology and bioinformatics*, 2019.

[2] Z. Yu, W. Luo, R. Tse, and G. Pau, "Dmnet: A personalized risk assessment framework for elderly people with type 2 diabetes," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–11, 2023.

[3] L.-Y. Chang, T.-Y. Lin, K.-H. Hsu, Y.-C. Huang, K.-L. Lin, C. Hsueh, S.-R. Shih, H.-C. Ning, M.-S. Hwang, H.-S. Wang *et al.*, "Clinical features and risk factors of pulmonary oedema after enterovirus-71-related hand, foot, and mouth disease," *The Lancet*, 1999.

[4] I. Pearce, R. Simó, M. Lövestam-Adrian, D. T. Wong, and M. Evans, "Association between diabetic eye disease and other complications of diabetes: implications for care. a systematic review," *Diabetes, obesity and metabolism*, vol. 21, no. 3, pp. 467–478, 2019.

[5] H. Song, D. Rajan, J. Thiagarajan, and A. Spanias, "Attend and diagnose: Clinical time series analysis using attention models," in *AAAI*, vol. 32, no. 1, 2018.

[6] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," *Advances in neural information processing systems*, vol. 29, 2016.

[7] L. Ma, C. Zhang, Y. Wang, W. Ruan, J. Wang, W. Tang, X. Ma, X. Gao, and J. Gao, "Concare: Personalized clinical feature embedding via capturing the healthcare context," in *AAAI*, 2020.

[8] L. Ma, J. Gao, Y. Wang, C. Zhang, J. Wang, W. Ruan, W. Tang, X. Gao, and X. Ma, "Adacare: Explainable clinical health status representation learning via scale-adaptive feature extraction and recalibration," in *AAAI*, 2020.

[9] J. Gao, C. Xiao, Y. Wang, W. Tang, L. M. Glass, and J. Sun, "Stagenet: Stage-aware neural networks for health risk prediction," in *IW3C2*, 2020.

[10] Y. Liu, J. Wu, Y. Wei, B. Mao, C. Li, and T. Gong, "Aefnet: Adaptive scale feature based on elastic-and-funnel neural network for healthcare representation," in *IEEE BIBM*, 2021.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE CVPR*, 2018, pp. 7132–7141.

[12] V. K. Somers, D. P. White, R. Amin, W. T. Abraham, F. Costa, A. Culebras, S. Daniels, J. S. Floras, C. E. Hunt, L. J. Olson *et al.*, "Sleep apnea and cardiovascular disease: An american heart association/american college of cardiology foundation scientific statement," *Journal of the American College of Cardiology*, 2008.

[13] S. Bharadwaj, S. Ginoya, P. Tandon, T. D. Gohel, J. Guirguis, H. Vallabh, A. Jevenn, and I. Hanouneh, "Malnutrition: laboratory markers vs nutritional assessment," *Gastroenterology report*, 2016.

[14] D. Joseph, J. A. Ker, P. Rheeder, and V. Ueckermann, "The association between elevated troponin levels and all cause 30 day mortality in critically ill patients seen at an academic hospital–a prospective cohort study," 2017.

[15] (2021) What is a normal respiratory rate for adults and children? [Online]. Available: https://www.healthline.com/health/normal-respiratory-rate

[16] (2022) What's a normal resting heart rate? [Online]. Available: https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979

[17] Oxygen levels, pulse oximeters, and covid-19. [Online]. Available: https://www.health.state.mn.us/diseases/coronavirus/pulseoximeter.html

[18] What are normal blood glucose levels? [Online]. Available: https://www.singlecare.com/blog/normal-blood-glucose-levels/

[19] Understanding blood pressure readings. [Online]. Available: https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings

[20] (2016) How is body temperature regulated and what is fever? [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279457/

[21] (2019) What's a normal blood ph and what makes it change? [Online]. Available: https://www.healthline.com/health/ph-of-blood

[22] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," in *ACM SIGKDD*, 2017.

[23] L. Ma, X. Ma, J. Gao, C. Zhang, Z. Yu, X. Jiao, W. Ruan, Y. Wang, W. Tang, and J. Wang, "Covidcare: Transferring knowledge from existing emr to emerging epidemic for interpretable prognosis," *arXiv preprint arXiv:2007.08848*, 2020.

[24] I. M. Baytas, C. Xiao, X. Zhang, F. Wang, A. K. Jain, and J. Zhou, "Patient subtyping via time-aware lstm networks," in *Proceedings of the 23rd ACM SIGKDD*, 2017, pp. 65–74.

[25] T. Ma, C. Xiao, and F. Wang, "Health-atm: A deep architecture for multifaceted patient health record representation and risk prediction," in *SDM*. SIAM, 2018.

[26] Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," in *SDM*, 2016.

[27] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, "Gram: graph-based attention model for healthcare representation learning," in *Proceedings of the 23rd ACM SIGKDD*, 2017, pp. 787–795.

[28] E. Choi, C. Xiao, W. Stewart, and J. Sun, "Mime: Multilevel medical embedding of electronic health records for predictive healthcare," *Advances in neural information processing systems*, vol. 31, 2018.

[29] F. Ma, Y. Wang, H. Xiao, Y. Yuan, R. Chitta, J. Zhou, and J. Gao, "A general framework for diagnosis prediction via incorporating medical code descriptions," in *BIBM*. IEEE, 2018, pp. 1070–1075.

[30] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao, "Kame: Knowledge-based attention model for diagnosis prediction in healthcare," in *Proceedings of the 27th ACM CIKM*, 2018.

[31] J. Gao, C. Yang, J. Heintz, S. Barrows, E. Albers, M. Stapel, S. Warfield, A. Cross, J. Sun *et al.*, "Medml: Fusing medical knowledge and machine learning models for early pediatric covid-19 hospitalization and severity prediction," *Iscience*, 2022.

[32] X. S. Zhang, F. Tang, H. H. Dodge, J. Zhou, and F. Wang, "Metapred: Meta-learning for clinical risk prediction with limited patient electronic health records," in *ACM SIGKDD*, 2019.

[33] D. Zhao and C. Weng, "Combining pubmed knowledge and ehr data to develop a weighted bayesian network for pancreatic cancer prediction," *Journal of biomedical informatics*, vol. 44, no. 5, pp. 859–868, 2011.

[34] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, p. 107398, 2021.

[35] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[36] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, and A. Galstyan, "Multitask learning and benchmarking with clinical time series data," *Scientific data*, vol. 6, no. 1, pp. 1–18, 2019.

[37] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *ICML*, 2006.

[38] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, 1982.

[39] J. Keilwagen, I. Grosse, and J. Grau, "Area under precision-recall curves for weighted and unweighted data," *PloS one*, 2014.

[40] G. L. Sternbach, "The glasgow coma scale," *The Journal of emergency medicine*, vol. 19, no. 1, pp. 67–71, 2000.