

ELDA: Learning Explicit Dual-Interactions for Healthcare Analytics

Qingpeng Cai*, Kaiping Zheng*, Beng Chin Ooi*, Wei Wang†, Chang Yao‡

*School of Computing, National University of Singapore

†ByteDance, Singapore ‡School of Computing, Zhejiang University

Email: {qingpeng, kaiping, ooibc}@comp.nus.edu.sg, wangwei.cs@gmail.com, chang.yao@zju.edu.cn

Abstract—Interaction learning plays an essential role in learning patients’ comprehensive representations that contribute to improved performance in many analytical tasks. In healthcare, interactions among medical features (i.e., feature-level interactions) can exhibit different abnormal patterns in detail, while interactions among time steps (i.e., time-level interactions) can indicate the dynamic changes in patients’ health conditions. Therefore, it is necessary to capture and analyze both types of interactions when conducting healthcare analytics.

In this paper, we propose a general framework ELDA that is supported by the novel model ELDA-Net to learn dual-interactions for healthcare analytics in an explicit manner. Specifically, we devise a Feature-level Interaction Learning Module that can enrich a separately processed medical feature by learned interactions among medical features, and a Time-level Interaction Learning Module that can enhance the representations of the patients’ health conditions by learned interactions among time steps. In both levels, ELDA can provide explicit and intuitive interpretations via explaining through the designed attention mechanism. Further, to facilitate the feature-level interaction learning, we propose a novel Bi-directional Embedding Module in ELDA-Net which can efficiently embed the medical features recorded in numerical values. We evaluate the effectiveness and interpretability of ELDA over two public real-world clinical datasets. The experimental results confirm that ELDA consistently outperforms existing state-of-the-art methods with a significant margin, and supports fine-grained interpretability in both the feature level and the time level with medical insights.

I. INTRODUCTION

Over the years, a tremendous amount of data has been collected by companies and organizations for record and analysis purposes. To analyze such valuable data, there is an increasing demand for using advanced deep learning libraries [1], [2], [3] to design powerful models as part of database management systems (DBMS) [4], [5], [6], which can facilitate complex analytics [7], [8]. Among complex analytic applications, healthcare analytics [9], [10] is a highly critical one, as it involves life-and-death clinical decision-making. Healthcare analytics aims to analyze various sources of healthcare data via data-driven approaches to facilitate critical clinician decisions for improved patient management.

In healthcare analytics, one of the most essential data sources is the electronic medical records (EMRs) that include various types of patients’ time-series medical features (e.g., lab tests, medications, etc). However, it is non-trivial to utilize the EMR data for analysis, and many existing studies propose advanced models to address representative issues such as irregularity [11], [12], [13], [14], [15], bias [16], heterogeneity [17],

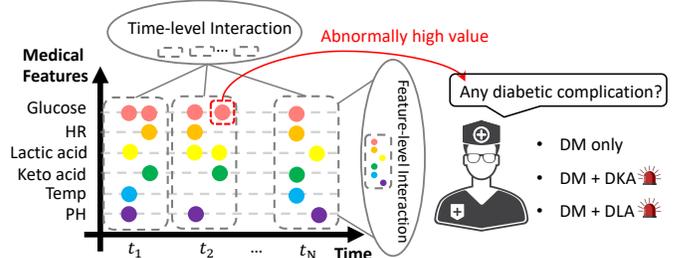


Fig. 1. An example diabetic patient’s time-series EMR data. We aim to explicitly explore the interactions in both the feature level and the time level to provide meaningful medical insights for clinicians.

sparsity [14], [15], [18], interpretability [19], [20], [21], [22], etc, and to facilitate human-in-the-loop data analysis [23], [24].

In the design of analytic models for time-series EMR data, interactions are essential for reflecting patients’ conditions. To illustrate these crucial interactions, we take diabetes mellitus (DM) as an example, which has become one of the most serious metabolic diseases around the world [25]. We show the time-series EMR data of an example diabetic patient in Figure 1. As illustrated, there exist two types of interactions in time-series EMR data: (i) **feature-level interactions** that denote the interactions among medical features, e.g., the interaction between Glucose and Lactate at t_1 ; (ii) **time-level interactions** that denote the interactions among time steps, e.g., the interactions between t_1 and t_N . Both types of interactions are of vital significance to healthcare analytics. Specifically, the feature-level interactions can depict abnormal patterns among medical features, and we take the example patient’s abnormal glucose as an example to further elaborate on the feature-level interactions. As shown in Figure 1, glucose suddenly increases at t_2 , and clinicians immediately examine several other medical features. Through analyzing the interactions between glucose and other features, clinicians can evaluate the severity of patients’ conditions and arrange necessary treatments to avoid the occurrence of DM complications. Different types of DM complications exhibit different symptoms:

- **DM only**: In general, DM patients only suffer from the problem of high blood glucose [26];
- **DM + Diabetic ketoacidosis (DKA)** [26]: DM patients with DKA always develop high keto acid, and a low Potential of hydrogen (PH) along with high blood glucose;
- **DM + Diabetic lactic acidosis (DLA)** [27]: DM patients with DLA tend to exhibit a high lactic acid and a low PH

value along with high blood glucose.

According to these two DM complications, the same abnormal value of blood glucose may indicate different health conditions that can be reflected by other closely related medical features, such as PH, keto acid, and lactic acid. Therefore, learning feature-level interactions can depict different abnormal patterns among medical features and provide insightful interpretation results for clinicians when analyzing these features. In healthcare analytics, some existing studies attempt to model the feature-level interactions in an implicit manner. For example, some models [13], [14], [28] simply project the raw medical features into a latent space via a deep neural network, while some others [20], [29] employ attention-based networks to summarize feature representations via learned attention scores as interaction representations. However, none of these models can learn separate representations to denote the interactions between each pair of medical features, that is to model the feature-level interactions in an explicit manner. It is advantageous to explicitly learn inherent feature-level interactions, as the explicit modeling provides more intuitive interaction representations that can contribute to a better assessment of interrelationships among features.

From another perspective, the time-level interactions can describe the dynamic change of a patient’s health conditions and unveil several crucial time steps. As illustrated in Figure 1, the pathogenetic process of a patient’s health conditions is dynamically changing over time. To model such changes in conditions, most deep learning models employ recurrent neural networks (RNN) [30] based models (e.g., LSTM [31], GRU [32]) and use the representation at the last time step to denote a patient’s final health condition [13], [14], [29] as it summarizes the information of all time steps and reflects the latest condition. Due to the RNN-based mechanism, these models cannot take the time-level interactions into account. It is of vital importance to model the time-level interactions, as it helps both fully leverage the historical medical records and reflect the changes in patients’ health conditions. However, most existing studies only learn such time-level interactions in an implicit manner. For instance, some studies (e.g., [19], [33]) learn the importance in the visit level, which cannot analyze the interactions between pairs of time steps. Other studies (e.g., [20], [21], [34]) take the last time step as a key to learn the attention weights of all the previous time steps and summarize them with these weights as the time-level interaction representations. Unfortunately, none of them can explicitly learn the separate time-level interaction representations between pairs of time steps. In practice, such explicit time-level interactions are crucial as they contribute to improving models’ interpretability and thus, facilitating the trust of clinicians.

Another line of research is Factorization Machine (FM) [35] based models that provide a feasible solution for explicit interaction representation. FM-based models can explicitly model pair-wise feature interactions, but it is non-trivial to apply the FM-based models in the time-series EMR data, due to the following three reasons. First, FM-based models

simply separate the information into two parts, i.e., the original feature part and the feature interaction part. Such a coarse-grained information separation cannot build a tight relationship between a feature and its corresponding interactions. Second, some FM-based models attempt to take time-series data into consideration when modeling feature interactions, e.g., TransFM [36] and SeqFM [37]. However, they cannot effectively learn the interactions among multiple time-series medical features across time in the EMR data. Third, while the FM-based embedding mechanism can work well for categorical medical concepts (e.g., diagnoses), it is ineffective for embedding numerical medical features (e.g., lab tests).

To tackle the aforementioned challenges, we propose a general framework ELDA to learn **ExpLicit Dual**-interaction for healthcare analytics with time-series EMR data as input. Specifically, we devise the model ELDA-Net as the core component of ELDA, which consists of four modules, namely Bi-directional Embedding Module, Feature-level Interaction Learning Module, Time-level Interaction Learning Module, and Prediction Module. The Feature-level Interaction Learning Module processes medical features separately to preserve the feature individuality and enriches them with their corresponding interactions with other medical features. Therefore, it is capable of depicting different abnormal patterns among medical features. The Time-level Interaction Learning Module models the interactions among time steps, thereby learning the dynamic changes of patients’ conditions and unveiling crucial time steps that vary among patients. In both modules, we propose a designed attention network to differentiate the importance of the interactions, and hence, can provide fine-grained interpretation results in both levels. Further, to facilitate feature interaction learning, we devise the novel Bi-directional Embedding Module to produce more informative embedding vectors for numerical medical features. With ELDA-Net, ELDA is capable of facilitating accurate predictive analytics for time-series EMR data, and depicting diverse abnormal patterns based on both feature-level and time-level interactions. We have released our code in: <https://github.com/KimballCai/ELDA>.

The main contributions are summarized as follows:

- We develop a general framework ELDA to learn explicit dual-interactions for healthcare analytics. At the core of ELDA, we propose a novel end-to-end model ELDA-Net that is the first proposal to explicitly learn the feature-level interactions at each time step and the time-level interactions simultaneously, to the best of our knowledge.
- We devise a novel Bi-directional Embedding Module for numerical medical features, which can generate more informative embedding vectors with superior performance compared with the FM-based embedding mechanism.
- We evaluate the effectiveness of ELDA for healthcare analytics in two public real-world clinical datasets. The extensive experimental results confirm that ELDA-Net consistently achieves more accurate analytics than existing state-of-the-art methods.
- By visualizing the interpretation results in detail, we demon-

strate that ELDA can identify meaningful abnormal patterns in both the feature level and the time level, which provides informative medical insights on patient management and reveals the potential for advancing medical research.

The remainder of this paper is organized as follows: Section II reviews the related work. Section III demonstrates the ELDA framework for healthcare analytics. We elaborate on the detailed design of ELDA-Net and its modules in Section IV. We then evaluate the performance and the interpretability of ELDA through extensive experiments in Section V, and conclude in Section VI.

II. RELATED WORK

Factorization Machine based Models. Factorization machine (FM) [35] is a classic method, which can explicitly represent the feature interaction, which first embeds each feature into an embedding vector and then measures the pair-wise interaction via the inner-product of their embedding vectors. Various FM-based models [38], [39], [40], [41], [42], [43] have been successfully applied in different information retrieval tasks such as the recommendation system, the click-through rate (CTR) prediction, etc. With interpretability as a focus, we concentrate on one extension of FM, i.e., Attention Factorization Machine (AFM) [38] that adopts an attention neural network to discriminate the importance of different feature interactions. Although such FM-based models can effectively capture the interactions in static features (e.g., ID, gender, etc), most of them cannot take time-series data into consideration. To bridge this gap, TransFM [36] attempts to incorporate FM into translation operation for sequential recommendations, but it mainly focuses on the impact of a sequence’s last item, which is insufficient to model the dependencies of the whole sequence [37]. Further, SeqFM [37] is the first study to employ FM-based models to systematically learn sequential dependencies, and it can learn the interactions between static features and a sequence of dynamic features (i.e., the purchase sequence) by employing multi-view learning mechanisms. However, it is non-trivial to employ SeqFM to the scenarios with multiple sequences of dynamic features as in time-series EMR data. In a nutshell, it is still challenging to apply FM-based methods for time-series EMR data analytics.

Medical Concept Embedding. Inspired by word embedding in natural language processing, medical embedding has been proposed and attracting a great deal of attention recently. Most of them [19], [44], [45], [46] focus on handling categorical medical concepts and directly assign each feature with an embedding vector. However, we cannot directly assign embedding to numerical medical features with infinite possible values.

Healthcare Analytics. There exist numerous studies investigating how to improve the performance of healthcare analytics. For example, RETAIN [19] devises a two-level attention model to capture visit-level and variable-level feature importances. SANd [28] borrows the idea of the transformer and employs self-attention mechanism [47] to learn patients’ representations. Besides, Dipole [21] proposes three attention mecha-

nisms to capture the essential relationships among patients’ visits in the time-series EMR data. Concare [20] processes each feature separately via different GRUs and finally applies the self-attention mechanism to integrate the information of all features to explore cross-feature interdependencies. However, all of these attention-based models cannot model the interactions in an explicit manner that is to learn separate representations to denote the interactions between each pair of medical features or time steps. Besides the studies above, several solutions are proposed to tackle other challenges. For example, GRU-D [14] devises an exponential decay mechanism to model time difference and impute missing data with either the empirical mean value or the last observation. StageNet [48] proposes a stage-aware LSTM module to capture the stage of the disease progression and applies convolution operation to learn progression patterns.

Different from these studies, our proposed ELDA-Net focuses on modeling two types of interactions in an explicit and fine-grained manner which can depict abnormal patterns for medical features and time steps, and hence, contribute to improved performance and significant interpretability.

III. ELDA FRAMEWORK

In this section, we elaborate on how our proposed ELDA works for healthcare analytics. Benefiting from the novel design of ELDA, as illustrated in Figure 2, ELDA manages to facilitate accurate predictive analytics and depict abnormal patterns in both feature-level and time-level simultaneously.

Data. ELDA first extracts the time-series EMR data from the database system of hospitals for analysis. Based on these history EMR data, ELDA is able to train ELDA-Net with satisfactory performance superior to other existing state-of-the-art models. Further, when new EMR data is generated (e.g., new patients are admitted to ICU in the hospital), ELDA can employ the trained ELDA-Net to process such newly generated EMR data with detailed predictive analysis and interpretations.

ELDA-Net. As the core component of the framework ELDA, ELDA-Net consists of four modules (in order): (a) Bi-directional Embedding Module to augment numerical medical features into informative representations, which is fundamental to the following module; (b) Feature-level Interaction Learning Module to explicitly capture the interactions between pairs of medical features in each time step and demonstrate the change of the importance (i.e., attention weights) of the interactions over time; (c) Time-level Interaction Learning Module to explicitly capture the interactions between the last time step and all previous time steps; (d) Prediction Module to summarize the information and derive the final prediction. Detailed descriptions of each module are in Section IV.

Functionality. Interpretations of the prediction results are one of the most essential concerns for clinicians to understand how a model works and therefore, trust the model. Benefiting from the novel design of the ELDA-Net, ELDA can facilitate healthcare analytics with three advanced functionalities. We take in-hospital mortality prediction for ICU patients as an example scenario to elaborate on each functionality as follows:

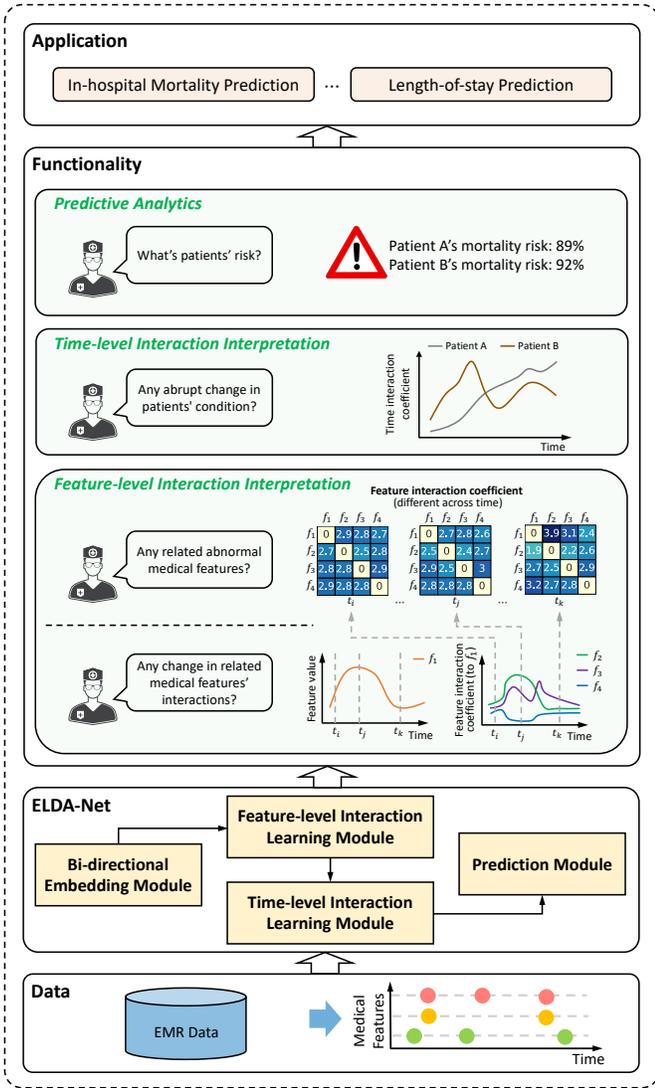


Fig. 2. Overview of the ELDA framework for healthcare analytics.

- **Predictive Analytics.** For effective patient management and efficient medical resource utilization, it is essential to analyze patients' EMR data to monitor their dynamic health conditions. During the hospitalization in ICU, ELDA is able to predict the risk of passing away for patients regularly via Prediction Module, and if the prediction exceeds a predefined threshold, ELDA can trigger timely alerts to inform clinicians to pay close attention.
- **Time-level Interaction Interpretation.** As patients tend to exhibit personalized disease development over time, it is crucial for clinicians to be aware of their health conditions in a time-aware manner. For example, the conditions of the patients with severe complications always fluctuate and may cause disease recurrence, whereas the conditions of the patients with mild severity are generally more stable during hospitalization. With Time-level Interaction Learning Module, ELDA is able to depict the abnormal pattern on the dynamic change of patients' conditions by identifying crucial time steps that vary among patients. With such

time-level interaction interpretations, ELDA is able to assist clinicians in investigating the dynamic development of each patient's conditions and therefore benefit patient monitoring.

- **Feature-level Interaction Interpretation.** For a DM patient, if the value of Glucose exhibits a sudden increase at a certain time point, clinicians will generally investigate other DM-related features to verify whether this patient develops any DM complications. In practice, ELDA supports this via illustrating the importance of feature-level interactions, i.e., the feature interaction coefficients learned explicitly in Feature-level Interaction Learning Module. As illustrated, the related medical features (e.g., f_2 with an abnormal value and a close relationship with f_1) tend to attach a high coefficient value. Such interpretation results from ELDA can not only unveil informative relationships among medical features and help identify the underlying disease development, but also provide medically meaningful insights for potential medical research advancement.

Further, during the ICU hospitalization, patients' conditions are continuously changing, which also implies the change in feature interactions. In general, abnormal medical features usually return to a normal level after careful treatment and the attention paid to these medical features should also change accordingly. To support this, ELDA can demonstrate the change of the coefficient for each medical feature over time. An example is illustrated via the line chart in Figure 2, in which medical features with tight relationships to the abnormal f_1 (e.g., f_2, f_3) will always have high coefficients during the time when they become abnormal, while others do not (e.g., f_4). In addition, the variation trends of these coefficients are different among medical features.

Application. Our proposed ELDA is able to support diverse applications in healthcare analytics. In this paper, we investigate how ELDA supports the analytics of two crucial healthcare applications, i.e., in-hospital mortality prediction task and length-of-stay prediction task. More details can be found in Section V.

IV. METHODOLOGY

In this section, we first introduce the notations used in this paper and then elaborate on the detailed design of ELDA-Net.

A. Basic Notations

In this paper, we denote the multivariate time-series EMR data with $|C|$ medical features of length T as $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T) \in \mathbb{R}^{T \times |C|}$, where $t \in \{1, 2, \dots, T\}$ is the index of the time step (e.g., an hour) and $x_t \in \mathbb{R}^{|C|}$ denotes $|C|$ medical features (e.g., lab tests) observed in the time step t . Without loss of generality, we formulate the classification problem with the binary classification case, in which each sample \mathbf{X} has a corresponding binary label $y \in \{0, 1\}$. The goal is to learn a mapping function \mathbf{f} from the input data \mathbf{X} to the label y . For simplicity, in the rest of this paper, we drop the subscript t when discussing a specific time step.

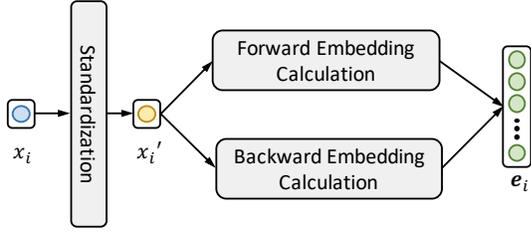


Fig. 3. Bi-directional Embedding Module of ELDA-Net.

B. ELDA-Net

We illustrate the overview of our proposed ELDA-Net in Figure 2 and explain ELDA-Net’s each module in order.

Bi-directional Embedding Module. Medical feature embedding has been widely investigated in healthcare analytics, such as [19], [44], [46], [49], [50]. Typically, these existing studies simply look up an embedding vector to embed a categorical medical feature. However, such an embedding mechanism is not suitable to learn an informative representation for medical features that are recorded in numerical values. One reason is that such numerical medical features can have infinite possible values and thus, it is impossible to directly assign each possible value with a certain embedding vector.

To tackle this challenge, FM-based models adopt a linear embedding mechanism that merely multiplies an embedding vector by its corresponding numerical feature, i.e., the embedding vector for the i -th feature x_i is represented as $\mathbf{v}_i x_i$, and the formula of FM is as follows:

$$\hat{y}_{FM}(\mathbf{x}) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \quad (1)$$

where w_0 is the global bias, w_i is the weight of the i -th feature, and $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors.

This FM-based embedding approach can handle numerical medical features to some extent, but it still has non-negligible limitations. To start with, the embedding vector is strongly related to the scale of the feature value, and this problem cannot be solved by standardizing all medical features. In practice, a feature with a smaller value will be mapped into an embedding vector with a relatively smaller scale (i.e., L2-norm). More specifically, zero value will be mapped to a zero embedding vector that has no influence on the latter computation. In healthcare analytics, a standardized zero value in lab test always denotes that this lab test is close to normal, which is essential to analyze a patient’s conditions. Further, a feature with opposite values will be embedded into opposite vectors (i.e., vectors with the same magnitude but opposite directions), which only shows a minor difference and tends to be not effective enough in capturing various abnormal patterns.

To tackle these issues, we design the Bi-directional Embedding Module for the numerical features as illustrated in Figure 3. We first standardize medical features from x_i to x'_i and then apply our embedding module to map x'_i into an embedding vector with a predefined lower bound a and upper bound b :

$$\mathbf{e}_i = \frac{1}{b-a} (\mathbf{V}_i^a (x'_i - a) + \mathbf{V}_i^b (b - x'_i)) \quad (2)$$

where $\mathbf{V}^a, \mathbf{V}^b \in \mathbb{R}^{|C| \times e}$ are two embedding matrices that convert x'_i into a lower-dimensional feature representation, and e denotes the dimension of the embedding vector. The design of our Bi-directional Embedding Module has following advantages. First, this module can preserve the advantages of the linear embedding mechanism: (i) the embedding vectors of the same feature with different values vary from each other; (ii) for continuous values, it can learn consecutive embedding vectors, and two close values for the same medical feature will be mapped into similar embedding vectors. Second, compared with the FM-based embedding mechanism, this module can control the scale of embedding vectors.

ELDA-Net also embeds missing values that are quite common in EMR data. The reasons for such missing data can be generally classified into three categories: (i) medical features are unconcerned before the first observation; (ii) medical features are stable and will not change frequently; (iii) medical features are not essential to be observed as they are not so much about the patient’s disease. We treat the first type of missing data as the normal case and use the global average value for imputation, whereas for the second type of missing data, we impute them with the last observation. As for the last type of missing data, we argue that such unobserved medical features are also informative as they indicate that these medical features are not necessary for these patients. Hence, we transform medical features that are unobserved in the patient’s EMR data into a separate embedding vector \mathbf{V}_i^m , where $\mathbf{V}^m \in \mathbb{R}^{|C| \times e}$ and m denotes that the feature is missing.

Feature-level Interaction Learning Module. With Bi-directional Embedding Module, we manage to enrich each medical feature with an informative embedding vector. Based on the enriched representations, we then aim to explicitly model the feature-level interactions between pairs of medical features at each time step, and FM-based models provide a feasible solution that has achieved great success in the existing work [38], [39], [40], [41], [42]. Therefore, we adopt the element-wise product approach to devise the Feature-level Interaction Learning Module in ELDA-Net, which measures the interaction between feature i and feature j as $\mathbf{r}_{i,j} \in \mathbb{R}^e$:

$$\mathbf{r}_{i,j} = \mathbf{e}_i \odot \mathbf{e}_j \quad (3)$$

where \odot denotes the element-wise product of two embedding vectors. Different from original FM-based models that simply partition the information into two parts: the original feature part and the feature interaction part, ELDA-Net models the feature interactions as complementary information of the feature itself. For each medical feature, ELDA-Net processes it separately to preserve its individuality and incorporates its interactions with all the other features to increase its ability to distinguish various abnormal patterns. Hence, compared with FM-based models that incorporate the interactions in a coarse-grained manner, in ELDA-Net, each feature and its interactions

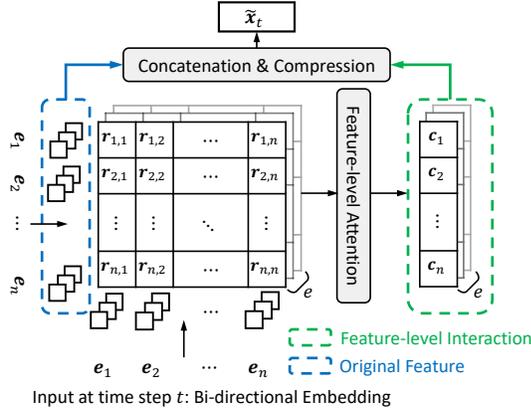


Fig. 4. Feature-level Interaction Learning Module of ELDA-Net.

can be more tightly connected in our proposed Feature-level Interaction Learning Module.

We note that different feature interactions should have discriminative importance and hence, should be assigned different attention weights. To achieve this goal, we devise an attention neural network in this module to learn the importance weights of interactions between each pair of medical features, which contributes to more fine-grained analytics. Specifically, our proposed attention neural network calculates the attention weights α as follows:

$$\alpha'_{i,j} = (\mathbf{W}_i^\alpha)^T \mathbf{r}_{i,j} + b_i^\alpha \quad (4)$$

$$\alpha_{i,j} = \frac{\exp(\alpha'_{i,j})}{\sum_{j=1, j \neq i}^{|C|} \exp(\alpha'_{i,j})} \quad (5)$$

where $\mathbf{W}^\alpha \in \mathbb{R}^{|C| \times e}$, $\mathbf{b}^\alpha \in \mathbb{R}^{|C|}$ are parameters, and $\alpha_{i,j}$ denotes the attention scores of the interaction between i -th feature and j -th feature when processing the i -th feature.

With the attention weights $\alpha_{i,j}$, ELDA-Net aggregates i -th feature's interactions with all other features as the complementary information and derives an overall representation of the feature-level interaction $\mathbf{c}_i = \sum_{j=1, j \neq i}^{|C|} \alpha_{i,j} \mathbf{r}_{i,j}$. We then combine the original feature embedding vector \mathbf{e}_i and this learned feature interaction \mathbf{c}_i as the new comprehensive representation of the i -th feature. Next, we transform this new feature representation into a lower-dimensional space \mathbf{f}_i via linear regression, which aims to speed up the processing and reduce parameters in the temporal modeling stage:

$$\mathbf{f}_i = \mathbf{p}^T \text{Relu}([\mathbf{e}_i; \mathbf{c}_i]) \quad (6)$$

where $\mathbf{p} \in \mathbb{R}^{2e \times d}$ is a trainable vector and d is a predefined hyper-parameter that denotes the size of the final feature representation, defined as *compression factor*. With a larger d , more information can be maintained, but the parameter size of the model will be increased in the meanwhile.

Through integrating \mathbf{f}_i for all features, we can derive a patient's new representation at time step t , which is $\tilde{\mathbf{x}}_t = [\mathbf{f}_1; \mathbf{f}_2; \dots; \mathbf{f}_{|C|}]$. With this devised Feature-level Interaction Learning Module, ELDA can provide informative feature-level interaction interpretation results as depicted in Figure 2.

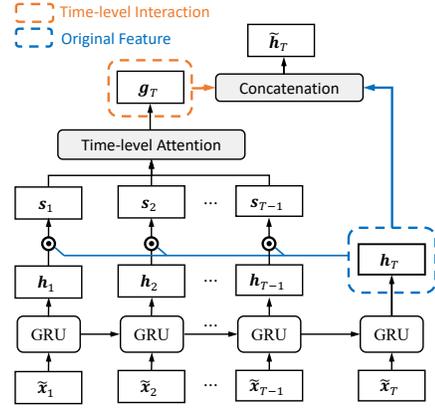


Fig. 5. Time-level Interaction Learning Module of ELDA-Net.

Time-level Interaction Learning Module. In Time-level Interaction Learning Module of ELDA-Net, we first model the dynamics of the processed time-series EMR data (i.e., $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_T$ learned in the Feature-level Interaction Learning Module) via a standard Gated Recurrent Unit (GRU) model [32]:

$$\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T = \text{GRU}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_T) \quad (7)$$

where $\mathbf{h}_i \in \mathbb{R}^l$ and l is a hyper-parameter that denotes the size of the hidden state of GRU.

In GRU, the last time step's output that summarizes all previous time steps' information is usually used as the overall representation of the time-series data for downstream applications, such as [30], [31], [32]. However, [21] demonstrates that the last time step's output cannot always adequately represent a patient's health condition because the influence of earlier time steps tends to be decayed along with time and relationships among time steps are ignored by simple RNN-based approaches. Therefore, in ELDA-Net, we also model the interactions in the time level to represent the dynamic change of the patient's health condition in a more comprehensive manner. Specifically, we focus on integrating the conditions at the last time step with its interactions with conditions at all earlier time steps because the conditions at the last time step are more essential compared with earlier ones due to the chronological order. Similar to the Feature-level Interaction Learning Module, we also adopt the element-wise product in this module to explicitly model the interactions $\mathbf{s}_{i,T}$:

$$\mathbf{s}_{i,T} = \mathbf{h}_i \odot \mathbf{h}_T, i = 1, 2, \dots, T-1 \quad (8)$$

where $\mathbf{s}_{i,T} \in \mathbb{R}^l$ denotes the interactions between the i -th time step and the last time step.

Further, to concentrate more on several critical time steps for the patient, an attention neural network is devised to differentiate the importance of the interactions between each earlier time step and the last time step:

$$\beta'_{i,T} = (\mathbf{w}^\beta)^T \mathbf{s}_{i,T} + b^\beta \quad (9)$$

$$\beta_{i,T} = \frac{\exp(\beta'_{i,T})}{\sum_{i=1}^{T-1} \exp(\beta'_{i,T})} \quad (10)$$

where $\mathbf{w}^\beta \in \mathbb{R}^l$ and $b^\beta \in \mathbb{R}$ are parameters of the attention network. $\beta_{i,T}$ denotes the attention scores of the interactions between the i -th time step and the last time step. With these learned attention weights β , we can aggregate time-level interactions and then generate an overall representation \mathbf{g}_T to depict the patient’s dynamic health condition:

$$\mathbf{g}_T = \sum_{i=1}^{T-1} \beta_{i,T} \mathbf{s}_{i,T} \quad (11)$$

Finally, we obtain a comprehensive representation of the patient’s health condition $\tilde{\mathbf{h}}_T = [\mathbf{h}_T; \mathbf{g}_T]$ that enriches the information of the last time step’s condition \mathbf{h}_T with the interaction information modeled in the Time-level Interaction Learning Module (i.e., \mathbf{g}_T): With this module, ELDA is capable of providing meaningful time-level interaction interpretation results as illustrated in Figure 2 to assist clinicians in patients’ condition assessment.

Prediction Module. Based on $\tilde{\mathbf{h}}_T$ derived in Time-level Interaction Learning Module, we can conduct different downstream prediction tasks. Take binary classification prediction as an example, we derive the prediction as \tilde{y} with the sigmoid activation function σ :

$$\tilde{y} = \sigma(\mathbf{w}_{pred}^T \tilde{\mathbf{h}}_T + b_{pred}) \quad (12)$$

where \mathbf{w}_{pred} and b_{pred} are parameters to learn. After deriving the prediction via this module, ELDA is able to support accurate predictive analytics as shown in Figure 2.

C. Optimization

We use the binary cross-entropy as the objective function to calculate the loss between the true label y and the predicted label \tilde{y} in the Prediction Module for all the patients:

$$Loss(\tilde{y}, y) = - \sum_{i=1}^N (y_i \log \tilde{y}_i + (1 - y_i) \log(1 - \tilde{y}_i)) \quad (13)$$

V. EXPERIMENTS

A. Experimental Set-up

Datasets. We evaluate the effectiveness of our proposed framework ELDA in two real-world public clinical datasets below.

PhysioNet2012 [51] is a public dataset from PhysioNet Challenge 2012 with 12,000 intensive care unit (ICU) admissions. In each admission, time-series EMR data such as chart events, lab tests, and output events within 48 hours are recorded. For each admission, 37 common medical features are selected [51], including Albumin, pH, Serum sodium, etc.

MIMIC-III [52] consists of more than 58,000 hospital admissions collected at Beth Israel Deaconess Medical Center spanning from 2001 to 2012. In the MIMIC-III dataset, we follow the cohort selection in [53] to sample patients. As in the PhysioNet2012 dataset, we extract the same 37 features of 21,139 admissions over 48 hours from the MIMIC-III dataset as samples in our analysis.

Each sample in the dataset denotes the medical records of a patient’s admission. In both datasets, we evaluate ELDA

TABLE I
STATISTICS OF THE PHYSIONET2012 AND THE MIMIC-III DATASETS.

	PhysioNet2012	MIMIC-III
# of admissions	12000	21139
survivor : non-survivor	10293 : 1707	18342 : 2797
LOS \leq 7 : LOS $>$ 7	4095 : 7738	9134 : 12005
avg. # of records per patient	359.19	346.05
# of medical features	37	37
missing rate (without imputation)	79.78%	80.52%

and baseline methods on two healthcare applications, i.e., in-hospital mortality prediction and Length-of-Stay (LOS) prediction. The first task is to predict whether a patient will pass away in the hospital with the EMR data collected within 48 hours after admission. Based on the same input, the second task is to predict whether a patient will be discharged from the hospital in 7 days after admission to the ICU, i.e., have a LOS over 7 days. Detailed statistics of both datasets are summarized in Table I. The survivor denotes the patients who pass away in the hospital and vice versa. Before conducting these two tasks, we apply a mean-std standardization on all medical features for both datasets. Besides, we also clean some noisy values that are erroneous in the real-world medical records, such as negative values, as suggested in [51].

Evaluation. We divide all samples into 80%:10%:10% as the train set, the validation set, and the test set respectively. After learning the best-performing model in the validation set, we report its performance in the test set, in terms of the binary cross-entropy loss (BCE loss), the area under the receiver operator characteristic curve (AUC-ROC), and the area under the precision recall curve (AUC-PR). The AUC-ROC score and the AUC-PR score are commonly used to measure the performance of a classifier for binary classification which is usually conducted on a highly imbalanced dataset [54]. Further, we run the experiments five times for each model per application to report the experimental results.

Model Configurations. As for the ELDA-Net and its variants, we set the hidden dimension of GRU as 64, the embedding dimension as 24, and the compression factor as 4. The lower bound and the upper bound in Bi-directional Embedding Module are set to -3 and 3 in the experiments. We train all models using an initial learning rate of 0.001 and a batch size of 64. As the numbers of features are the same in both datasets, we adopt this setting consistently for all tasks.

Implementation Details. We implement the ELDA-Net in Keras 2.3.1 with Tensorflow 1.14.0 as the backend. All the experiments are conducted with one Intel(R) Xeon(R) W-2133 CPU @ 3.60GHz and one GeForce RTX 2080 Ti GPU.

Baseline Methods. We compare ELDA-Net with the following baseline methods.

- **Logistic Regression (LR)** [55] takes the mean of the time-series values for each feature as input, and such input is also used for FM and AFM as well.
- **FM** [35] captures the pair-wise feature interactions based on a linear embedding mechanism.
- **AFM** [38] adds attention weights to differentiate the importance of each feature interaction.

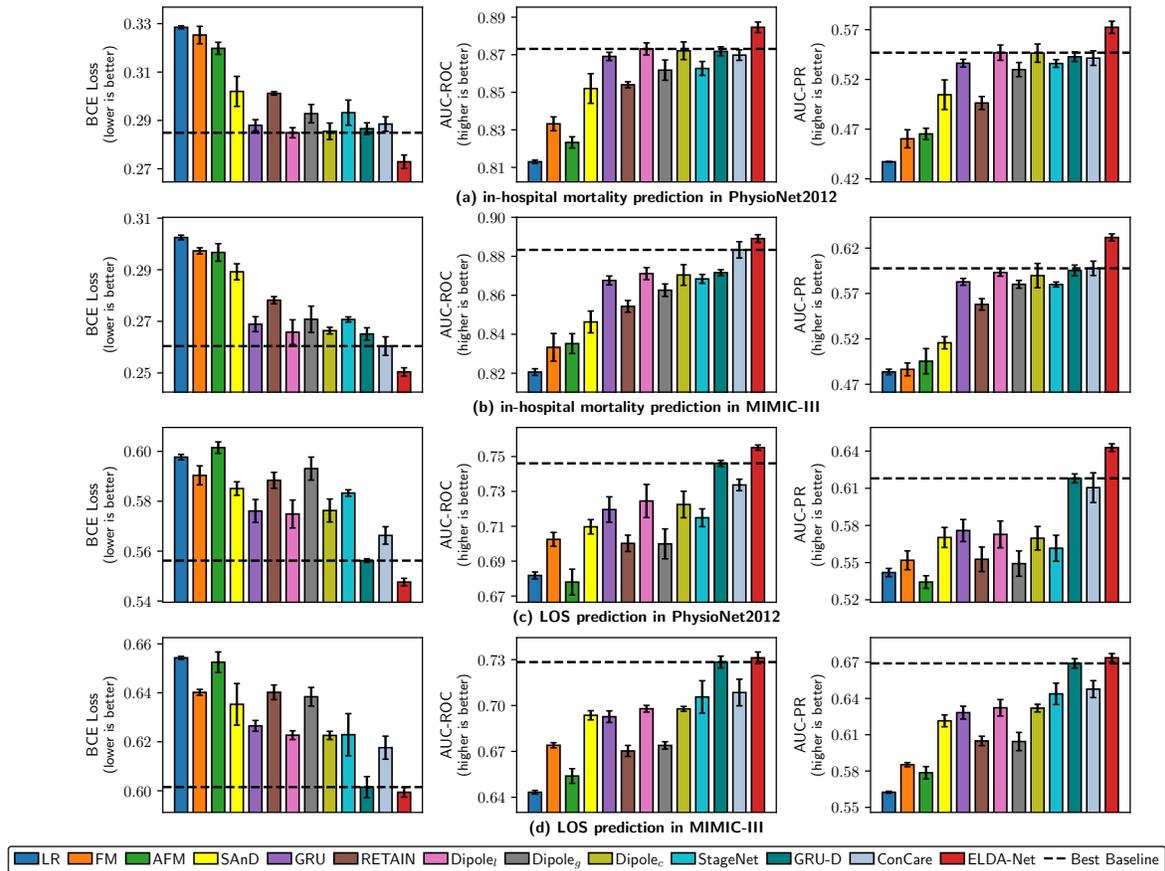


Fig. 6. Experimental results of ELDA-Net and baselines.

- **SAnD** [28] borrows the idea of transformer and employs a masked self-attention mechanism to model time-series data.
- **GRU** [32] is a widely adopted RNN-based model to process time-series data.
- **RETAIN** [19] applies two levels of GRU to learn attention weights in the visit level and the variable level.
- **Dipole** [21] takes the bidirectional GRU as the backbone model and devises three attention mechanisms, i.e., Dipole_l, Dipole_g, Dipole_c.
- **StageNet** [48] devises a stage-aware LSTM module to capture the stage of the disease progression and applies convolution operation to learn progression patterns.
- **GRU-D** [14] utilizes an exponential time decay mechanism to model time intervals and impute missing data, and hence, can handle irregular time-series data.
- **ConCare** [20] employs separate GRUs to process each medical feature across time and combines static data and time-series data via a self-attention mechanism.

Further, we conduct an ablation study to investigate the effectiveness of each module in ELDA-Net (in Section V-C), and the involved variants of ELDA-Net are as follows:

- **ELDA-Net-T** removes Feature-level Interaction Learning Module to validate the effectiveness of the Time-level Interaction Learning Module.
- **ELDA-Net-F_{bi}** keeps Feature-level Interaction Learning Module with our proposed Bi-directional Embedding Mod-

- ule, but it drops Time-level Interaction Learning Module.
- **ELDA-Net-F_{fm}** applies the FM-based embedding mechanism (i.e., a linear embedding without bias), which is its difference from ELDA-Net-F_{bi}.

Models with * assign embedding vectors with all ones to the features when the standardized values are zero, which is its difference from the model without *.

B. Main Results

We first evaluate our proposed ELDA on in-hospital mortality prediction task in both datasets, and the comparison results between the ELDA-Net and the baseline methods are shown in Figure 6. In the PhysioNet2012 and the MIMIC-III dataset, ELDA-Net always performs best in the in-hospital mortality prediction task, achieving an improvement of 2.6% and 3.4% in terms of AUC-PR over the most competitive baseline methods. Then, we evaluate ELDA-Net on the LOS prediction task in both datasets. The comparison results illustrated in Figure 6 confirm the effectiveness of ELDA-Net, which achieves an improvement of about 2.5% and 0.5% in terms of AUC-PR score compared with the best-performing baseline models in the PhysioNet2012 and the MIMIC-III dataset, respectively. In a nutshell, in all tasks and all datasets, ELDA-Net outperforms all the baseline methods consistently in terms of all metrics, which confirms the effectiveness of our proposal in facilitating accurate analytics.

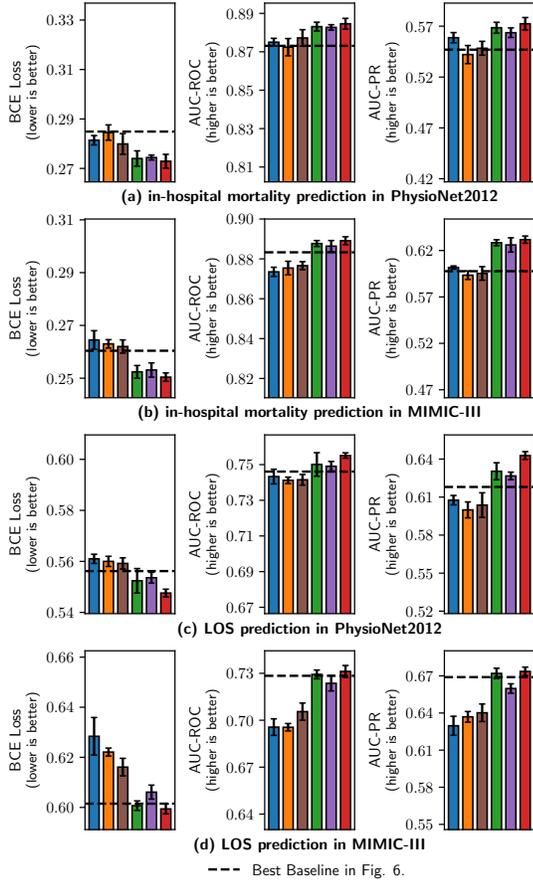


Fig. 7. Experimental results of the ablation study. The dash line in each subfigure denotes the best-performing baseline in Figure 6.

Among the baseline methods that generally handle non-time series data, LR is the most widely adopted interpretable model, but its performance is not satisfactory. In comparison, FM can consistently outperform LR, as FM manages to capture the pair-wise feature interactions that are used as augmented information. Besides, the superior performance of AFM mainly comes from the modeling of attention weights, which can discriminate the importance of interactions, but this improvement is not stable enough in the LOS prediction task.

Different from previous methods, time-series models can always achieve better performance due to the modeling of the dynamics in the EMR data, such as GRU and others. The performance of the StageNet also shows an improvement due to learning the stage of the disease progression via LSTM. Some models apply attention mechanisms to enhance both performance and interpretability. For example, Dipole can re-capture the crucial information in previous time steps via the attention network, which leads to boosted performance, and ConCare achieves competitive performance by modeling the cross-feature interdependence after separately processing different medical features. Their performance is always the best among the baseline models in the in-hospital mortality prediction task, but they achieve suboptimal performance in the LOS prediction task compared with GRU-D.

However, attention-based models cannot always achieve such superior performance. The model design of RETAIN considers the balance between the performance and the interpretability, which may be the reason for its degraded performance. Besides, SANd applies positional encoding and dense interpolation strategies to model the temporal order, which is not as powerful as RNN-based architecture and results in worse performance. Moreover, GRU-D devises an exponential decay mechanism to model the time interval and impute the missing data. Due to such a design, it can always provide satisfactory performance among all the baselines, especially in the LOS prediction task.

Our proposed ELDA consistently outperforms all baseline methods in all metrics, as ELDA-Net targets at explicitly modeling the interactions in both the feature level and the time level for time-series EMR data. In addition, ELDA-Net further devises a novel Bi-directional Embedding Module that can generate informative embedding vectors for numerical medical features. Based on these modules, ELDA-Net is able to generate a more comprehensive representation of a patient’s health condition and hence, achieves superior analytic performance.

C. Ablation Study

We then investigate the effectiveness of each proposed interaction learning module in ELDA-Net, with experimental results shown in Figure 7.

With the Time-level Interaction Learning Module, ELDA-Net-T is able to outperform many baselines. For example, as for the in-hospital mortality prediction task in PhysioNet2012, the AUC-PR score of ELDA-Net-T is 0.559, which is larger than 0.536 and 0.547 achieved by the original GRU and the best-performing baseline model (i.e., Dipole_l). Its superior performance validates that explicitly modeling the time-level interactions is essential to derive a more comprehensive representation of a patient’s dynamic health conditions and contribute to more accurate predictions.

Next, when evaluating the effectiveness of different feature-level interaction modeling mechanisms, Figure 7 also reveals several insightful findings. First, compared with ELDA-Net- F_{fm} , ELDA-Net- F_{fm}^* achieves a small improvement due to the modification of the embedding mechanism. It verifies that assigning separate embedding vectors to zero values is able to make better use of these values compared with assigning zero vectors in ELDA-Net- F_{fm} . Second, ELDA-Net- F_{bi} can consistently outperform both ELDA-Net- F_{fm} and ELDA-Net- F_{fm}^* in terms of all evaluation metrics in all tasks, which confirms that our proposed bi-directional embedding mechanism is more suitable to numerical medical features in the EMR data. Third, we also evaluate our devised module in the same scenario as ELDA-Net- F_{fm}^* and develop ELDA-Net- F_{bi}^* . As shown in Figure 7, there is a degradation in ELDA-Net- F_{bi}^* ’s prediction performance of all analytics tasks compared with ELDA-Net- F_{bi} , as such a modification breaks the consecutiveness of the embedding vectors for continuous values. Overall, ELDA-Net achieves better performance than

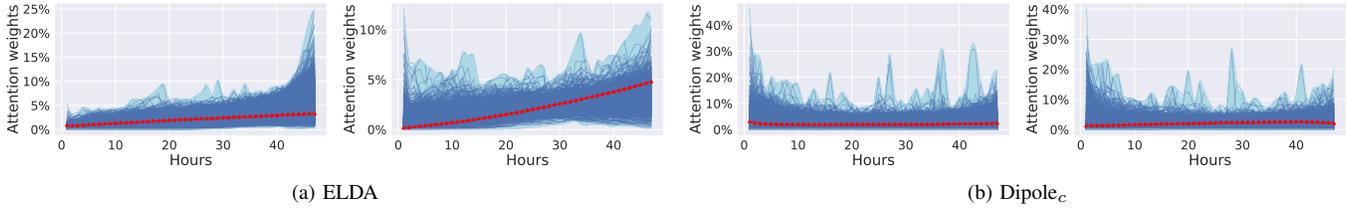


Fig. 8. Attention weights (percentage) of the time-level interactions for survival patients (left) and non-survival patients (right) in each subfigure. Each blue line denotes the attention weights of a certain patient over time. The red line denotes the average attention weights for either the whole group of survival patients or that of non-survival patients.

its variants as illustrated in Figure 7, which confirms that both levels of interactions are complementary to facilitate the learning of patients’ comprehensive representations. In addition, the integration of these two interaction learning modules can also equip ELDA-Net with the interpretability in both the feature level and the time level. We shall elaborate on the detailed interpretation results in Section V-D.

D. Interpretability Study

In this section, we use the PhysioNet2012 dataset to demonstrate how ELDA can provide interpretable results with medical insights for the in-hospital mortality prediction task. Specifically, we analyze the attention weights of the interactions learned in both the time level and the feature level.

ELDA’s Time-level Interaction Interpretation. To intuitively demonstrate the attention weights in the Time-level Interaction Learning Module, we analyze the interpretation results from ELDA for two complementary groups, i.e., survival patients and non-survival patients.

We illustrate the attention weights of the interactions between the last hour and all earlier hours (i.e., 47 hours) in Figure 8. From the red lines in Figure 8a, we can observe that in both groups, the Time-level Interaction Learning Module generally pays more attention to the patients’ latter conditions, because the conditions at these time steps are generally closer to a patient’s final health condition due to the chronological order. Besides, compared with survival patients, the conditions of the non-survival patients are more varied and unstable, and ELDA can focus more on several time steps that are more critical to these patients. In these time steps, the medical records are typically richer, and thereby, indicate the abnormality or sudden deterioration in patients’ health conditions. Besides, according to blue lines in Figure 8a, we also find that our Time-level Interaction Learning Module can effectively identify and assess the crucial time steps for patients where the positions and the lengths of these crucial time steps vary among patients due to patients’ individuality.

To further validate the effectiveness of our proposed module in the time-level interpretability, we introduce $Dipole_c$ for comparison, which can implicitly explore the interactions among patients’ subsequent visits in the time level. The time-level interpretation results of both models are illustrated in Figure 8. Compared with $Dipole_c$ ’s results in Figure 8b, we can observe that with the Time-level Interaction Learning Module, ELDA can effectively differentiate the general trends

in two cohorts of patients. Moreover, ELDA can also depict that the latter time steps’ representations exhibit more attention weights compared with earlier time steps. In a nutshell, ELDA with the Time-level Interaction Learning Module is effective in the time-level interpretability, as ELDA not only highlights crucial time steps but also shows the general trends in different groups of patients. Such auxiliary information is essential for clinicians to facilitate patient management.

ELDA’s Feature-level Interaction Interpretation. To illustrate the functionality of the Feature-level Interaction Learning Module in detail, we investigate a representative DM Patient A with DLA (as introduced in Section I). As a DM patient, Patient A suffers from extremely high values of Glucose as shown in the red line of Figure 10a. We then dive into more details of Patient A and show several essential medical records of Patient A in Table II, including the fraction of inspired oxygen (FiO₂), Glucose, bicarbonate (HCO₃), hematocrit (HCT), heart rate (HR), Lactate, mean arterial blood pressure (MAP), temperature (Temp), PH, and White blood cell (WBC). Due to the limited space, we only illustrate the attention weights of these features in two time steps (i.e., the 13th hour and the 35th hour) in Figure 9a, which are the start of the increase of Glucose and the time when Glucose becomes stable.

According to medical references (e.g., [27]), we know that DLA is a clinical syndrome caused by abnormal biochemical changes of the increase in blood lactic acid and the decrease in PH (<7.35) due to different reasons. In clinical practice, clinicians will do some auxiliary checking for better assessment of the condition because DLA patients typically have low HCO₃, low Temp, deep and big breath, and low blood pressure [27]. As shown in Table II, all the aforementioned symptoms are observed in Patient A, which confirms the development of DLA and reveals the patient’s worsening health condition. From Figure 9a, we can observe that the learned attention weights in the Feature-level Interaction Learning Module demonstrate that Glucose pays more attention to some closely related and abnormal medical features (e.g., FiO₂, HCO₃, HR, Lactate, MAP, Temp). However, some medical features that are irrelevant to DLA (e.g., HCT, WBC, etc.) tend to exhibit relatively low attention scores. All of these observations can be treated as abnormal patterns and are highly consistent with the medical knowledge for diagnosing DLA [27].

To evaluate our proposed ELDA’s interpretability in depicting abnormal patterns among medical features, we conduct a controlled experiment where we modify all the observed

TABLE II
ESSENTIAL MEDICAL FEATURES OF PATIENT A. THE VALUES IN THE TABLE ARE AFTER STANDARDIZATION.

Hours	FiO2	Glucose	HCO3	HCT	HR	Lactate	MAP	Temp	pH	WBC
13	0.85	6.09	-2.37	-0.96	3.32	10.41	-0.50	-2.16	-1.79	0.01
35	0.31	-0.46	-0.88	-0.10	1.25	9.03	0.16	-0.97	0.84	0.51

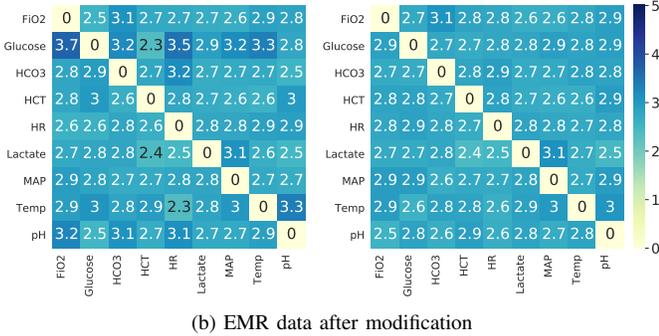
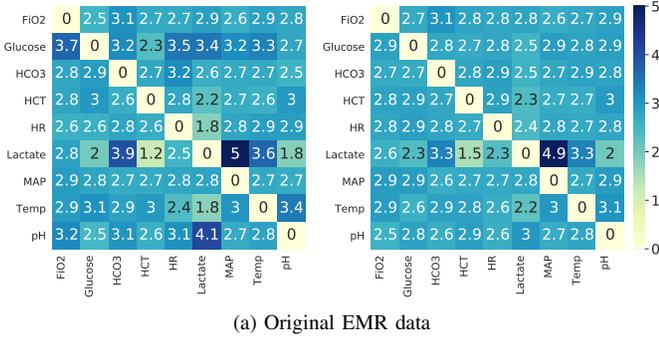


Fig. 9. Attention weights (percentage) of the feature-level interactions for Patient A on 13th hour (left) and 35th hour (right). Take the row pH as an example, values in the row denote the attention weights of interactions between pH and other features when processing the feature pH.

Lactate values in Patient A’s EMR data to normal (i.e., the global mean of the Lactate in all patients in the training dataset), and the corresponding results are shown in Figure 9b. In Figure 9a, we observe that the abnormal values of Lactate also pay more attention to several related medical features (e.g., MAP, Temp), which agree with [56], [57]. Specifically, it is stated in [56] that the increase in Lactate has an inverse correlation with Temp, and the results in [57] show that Glucose and Lactate correlate positively with blood pressure (i.e., MAP) in diabetic patients. Both these two Lactate-related medical features exhibit high attention weights when learning the interactions for the abnormal Lactate, and we can easily find in Figure 9b that these high attention weights are reduced to an average level on the modified EMR data. This controlled experiment confirms that our Feature-level Interaction Learning Module is able to unveil abnormal patterns among medical features. With this module, ELDA-Net can enrich medical feature representations with interactions between a feature itself and other medical features, which can depict different types of abnormality. Further, medical features with higher attention scores in the interactions imply tighter relationships, which has the potential to unveil the underlying interactions among medical features and advance medical research.

Another finding is that the same feature-feature interaction

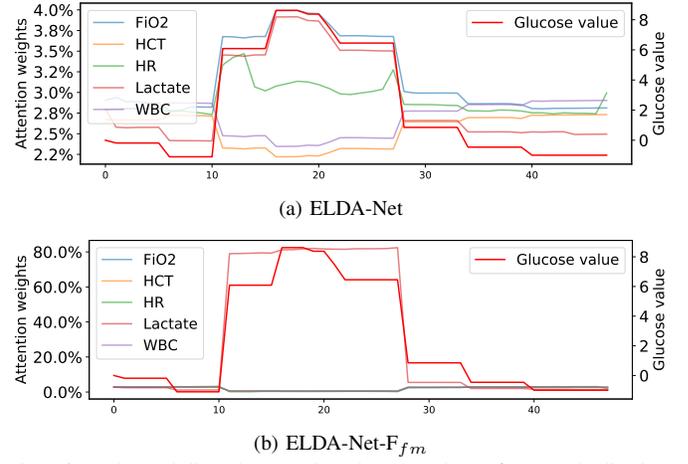


Fig. 10. The red line denotes the glucose value (after standardization) for Patient A. The lines in other colors show the attention weights of the interactions between glucose and other medical features (partial).

will have different attention weights in both features. For example, in the 13th hour, the high value of lactate is the main cause of the low pH value for Patient A, which is reflected as a higher attention weight in the interactions of pH. However, when it comes to lactate, ELDA-Net pays more attention to other abnormal features that are more important than pH, such as MAP, Temp, etc. After careful treatment in ICU for around a day (i.e., 35th hour), glucose reaches a normal level, which leads to a more uniform distribution of attention weights among other medical features (including pH).

Apart from demonstrating the feature-level interactions learned in these two time steps, we also find some interesting insights in the changing of the attention weights over time. While a patient is being hospitalized, the patients’ health conditions are continuously changing which can be reflected via the variation of the attention weights. According to Figure 10, we can observe that with the change of glucose value, different medical features exhibit varied influences on its attention weights. Some closely related medical features (e.g., FiO2, HR, Lactate) always attract more attention weights when they are abnormal, while others that are weakly related to glucose (e.g., HCT, WBC) do not exhibit an apparent influence with the same change in glucose values. Based on such interpretability analysis from the Feature-level Interaction Learning Module of ELDA, clinicians can timely detect the abnormal patterns and make responsive clinical decisions.

To further validate our ELDA-Net’s feature-level interpretability, we compare the interpretation results between ELDA-Net and ELDA-Net- F_{fm} in the feature level. The interpretation results of ELDA-Net- F_{fm} are shown in Figure 10b. As illustrated, we can find that Lactate always exhibits extremely high attention scores (i.e., $>50\%$) when learning interactions between the abnormal Glucose and all other medical features. This is because that the FM-based embedding mechanism employed in ELDA-Net- F_{fm} has a limitation that the scale of embedding vectors is highly related to feature values, which results in high attention scores for features with extreme abnormal values. In comparison to Lactate,

TABLE III
THE NUMBER OF PARAMETERS AND THE RUNTIME

Model	# of param	Training (s)	Prediction (ms)
LR	38	0.8	<0.01
FM	630	138	0.70
AFM	718	148	0.72
SAnD	106k	17	0.08
GRU	20k	9	0.05
RETAIN	13k	14	0.07
Dipole _{<i>i</i>}	40k	9	0.05
Dipole _{<i>g</i>}	56k	10	0.05
Dipole _{<i>c</i>}	44k	10	0.05
StageNet	85k	126	0.92
GRU-D	38k	466	3.23
ConCare	183k	118	0.69
ELDA-Net-T	21k	10	0.05
ELDA-Net-F _{<i>bi</i>}	49k	43	0.21
ELDA-Net-F _{<i>fm</i>}	43k	41	0.22
ELDA-Net	53k	44	0.22

attention scores of other Glucose-related medical features with abnormal values are compressed to a large extent, which leads to the limited capacity in learning patients’ comprehensive representations. Compared with ELDA-Net-F_{*fm*}, our ELDA-Net is able to capture and depict more abnormal patterns as illustrated in Figure 10a. For example, when the value of Glucose increases to an abnormal range, our Feature-level Interaction Learning Module can still attach more attention scores to other related medical features with abnormal values (e.g., FiO2, HR, etc.) besides the Lactate.

In a nutshell, supported by the medically meaningful interpretations in both the time level and the feature level, ELDA can provide valuable auxiliary information for clinicians to analyze patients’ conditions in a time-aware manner, and hence, facilitate personalized and timely treatments.

E. Efficiency Test

Model Complexity. We first illustrate the total number of trainable parameters in each model in Table III to measure the model complexity. As shown, LR, FM, AFM that are widely adopted to model non-time series data, have a small number of parameters (i.e., <1k). Hence, their performance is not competitive compared with time-series models. In time-series models, some state-of-the-art models (e.g., StageNet, Concare, and SAnD) need complex structures to distill information from the EMR data. Hence, they introduce tens of thousands of parameters that leads to their boosted analytic performance.

Compared with these models, our proposed ELDA-Net is able to significantly improve the performance with only a moderate number of parameters, where most additional parameters introduced are to derive the feature embedding in the Bidirectional Embedding Module.

Runtime. We next study the runtime for the offline training and online prediction with experimental results in Table III. Specifically, we report the average runtime (seconds) per batch for training where the batch size is set to 64, and we also report the average runtime (milliseconds) for making a prediction.

For the simplest approach LR, it only needs negligible time to train or predict, whereas the other two non-time series models FM [35] and AFM [38] need much longer time due to

modeling the feature interactions which can only be calculated in an element-wise manner and hence, hard to be optimized by matrix computation. As for time-series models, most of them only take a short time to do the training or prediction, whereas Concare and StageNet take a relatively longer time brought by their complex structure to explore cross-feature interdependencies and disease stages, respectively. In addition, GRU-D needs the longest time in both training and prediction, because it deals with irregular time series which is always longer than regular time series. Compared with most time series models, ELDA-Net-T and ELDA-Net-F_{*bi*}, as the variants of the ELDA-Net, slightly increase the runtime for both training and prediction due to the modeling of the interactions in the time level and the feature level. The latter needs to compute the interaction among all medical features which is more complex than the former, and hence, it introduces a bit more time. In total, the ELDA-Net can not only run faster than most SOTA models (e.g., ConCare, GRU-D, StageNet) but also achieve a remarkable performance improvement.

VI. CONCLUSIONS

In healthcare analytics, the feature-level and time-level interactions have not been explicitly investigated simultaneously. In this paper, we propose a novel end-to-end model ELDA-Net to explicitly capture and analyze these two types of interactions in time-series EMR data. Specifically, in the Feature-level Interaction Learning Module, ELDA-Net manages to preserve the individuality of each feature and depict different abnormal patterns to enrich the feature representation. In the Time-level Interaction Learning Module, ELDA-Net can describe the dynamic change of a patient’s health conditions and automatically identify the crucial time steps that vary among patients. ELDA-Net also includes a novel Bi-directional Embedding Module to generate more informative embeddings for numerical medical features. Based on the holistic design of ELDA-Net to learn a more comprehensive representation of patients’ dynamic health conditions, we further develop a general framework ELDA for more accurate predictions and fine-grained interpretability in time-series EMR data analytics. Extensive experiments show that our proposed ELDA-Net is consistently more accurate than state-of-the-art methods in terms of all evaluation metrics, and ELDA can provide interpretation results with medical insights in both the feature level and the time level.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their constructive comments and suggestions. This research/project is supported by the National Research Foundation, Singapore under its Emerging Areas Research Projects (EARP) Funding Initiative. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

REFERENCES

- [1] B. C. Ooi, K.-L. Tan, S. Wang, W. Wang, Q. Cai, G. Chen, J. Gao, Z. Luo, A. K. Tung, Y. Wang *et al.*, "Singa: A distributed deep learning platform," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 685–688.
- [2] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.
- [3] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.
- [4] M. Boehm, M. Dusenberry, D. Eriksson, A. V. Evfimievski, F. M. Manshadi, N. Pansare, B. Reinwald, F. Reiss, P. Sen, A. Surve, and S. Tatikonda, "Systemml: Declarative machine learning on spark," *Proc. VLDB Endow.*, vol. 9, no. 13, pp. 1425–1436, 2016.
- [5] A. Kumar, R. McCann, J. F. Naughton, and J. M. Patel, "Model selection management systems: The next frontier of advanced analytics," *SIGMOD Rec.*, vol. 44, no. 4, pp. 17–22, 2015.
- [6] W. Wang, M. Zhang, G. Chen, H. Jagadish, B. C. Ooi, and K.-L. Tan, "Database meets deep learning: Challenges and opportunities," *ACM SIGMOD Record*, vol. 45, no. 2, pp. 17–22, 2016.
- [7] M. Stonebraker, P. Brown, D. Zhang, and J. Becla, "Scidb: A database management system for applications with complex analytics," *Computing in Science & Engineering*, vol. 15, no. 3, pp. 54–62, 2013.
- [8] A. L'heureux, K. Grolinger, H. F. Elyamany, and M. A. Capretz, "Machine learning with big data: Challenges and approaches," *Ieee Access*, vol. 5, pp. 7776–7797, 2017.
- [9] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep ehr: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1589–1604, 2017.
- [10] P. Yadav, M. Steinbach, V. Kumar, and G. Simon, "Mining electronic health records (ehrs) a survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–40, 2018.
- [11] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Deepcare: A deep dynamic memory model for predictive medicine," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2016, pp. 30–41.
- [12] K. Zheng, W. Wang, J. Gao, K. Y. Ngiam, B. C. Ooi, and J. W. L. Yip, "Capturing feature-level irregularity in disease progression modeling," in *CIKM*. ACM, 2017, pp. 1579–1588.
- [13] I. M. Baytas, C. Xiao, X. Zhang, F. Wang, A. K. Jain, and J. Zhou, "Patient subtyping via time-aware LSTM networks," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. Part F1296, pp. 65–74, 2017.
- [14] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.
- [15] S. N. Shukla and B. M. Marlin, "Interpolation-prediction networks for irregularly sampled time series," *arXiv preprint arXiv:1909.07782*, 2019.
- [16] K. Zheng, J. Gao, K. Y. Ngiam, B. C. Ooi, and J. W. L. Yip, "Resolving the bias in electronic medical records," in *KDD*. ACM, 2017, pp. 2171–2180.
- [17] Y. Xu, S. Biswal, S. R. Deshpande, K. O. Maher, and J. Sun, "Raim: Recurrent attentive and intensive model of multimodal patient monitoring data," in *Proceedings of the 24th ACM SIGKDD international conference on Knowledge Discovery & Data Mining*, 2018, pp. 2565–2573.
- [18] J. Futoma, S. Hariharan, and K. Heller, "Learning to detect sepsis with a multitask gaussian process rnn classifier," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1174–1182.
- [19] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," *arXiv preprint arXiv:1608.05745*, 2016.
- [20] 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 833–840.
- [21] 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 *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 1903–1911.
- [22] K. Zheng, S. Cai, H. R. Chua, W. Wang, K. Y. Ngiam, and B. C. Ooi, "Tracer: A framework for facilitating accurate and interpretable analytics for high stakes applications," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 2020, pp. 1747–1763.
- [23] K. Zheng, G. Chen, M. Herschel, K. Y. Ngiam, B. C. Ooi, and J. Gao, "PACE: learning effective task decomposition for human-in-the-loop healthcare delivery," in *SIGMOD Conference*. ACM, 2021, pp. 2156–2168.
- [24] X. Liu, M. Lu, B. C. Ooi, Y. Shen, S. Wu, and M. Zhang, "CDAS: A crowdsourcing data analytics system," *Proc. VLDB Endow.*, vol. 5, no. 10, pp. 1040–1051, 2012.
- [25] P. Zimmet, K. Alberti, and J. Shaw, "Global and societal implications of the diabetes epidemic," *Nature*, vol. 414, no. 6865, pp. 782–787, 2001.
- [26] G. E. Umpierrez and A. E. Kitabchi, "Diabetic ketoacidosis," *Treatments in endocrinology*, vol. 2, no. 2, pp. 95–108, 2003.
- [27] P. English and G. Williams, "Hyperglycaemic crises and lactic acidosis in diabetes mellitus," *Postgraduate medical journal*, vol. 80, no. 943, pp. 253–261, 2004.
- [28] H. Song, D. Rajan, J. Thiagarajan, and A. Spanias, "Attend and diagnose: Clinical time series analysis using attention models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [29] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic, "Interpretable representation learning for healthcare via capturing disease progression through time," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 43–51.
- [30] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.
- [31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [32] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [33] X. Zhang, B. Qian, S. Cao, Y. Li, H. Chen, Y. Zheng, and I. Davidson, "Inprem: An interpretable and trustworthy predictive model for healthcare," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 450–460.
- [34] Q. Suo, F. Ma, G. Canino, J. Gao, A. Zhang, P. Veltri, and G. Agostino, "A multi-task framework for monitoring health conditions via attention-based recurrent neural networks," in *AMIA annual symposium proceedings*, vol. 2017. American Medical Informatics Association, 2017, p. 1665.
- [35] S. Rendle, "Factorization machines," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 995–1000, 2010.
- [36] R. Pasricha and J. McAuley, "Translation-based factorization machines for sequential recommendation," in *Proceedings of the 12th ACM Conference on Recommender Systems*, 2018, pp. 63–71.
- [37] T. Chen, H. Yin, Q. V. H. Nguyen, W.-C. Peng, X. Li, and X. Zhou, "Sequence-aware factorization machines for temporal predictive analytics," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020, pp. 1405–1416.
- [38] J. Xiao, H. Ye, X. He, H. Zhang, F. Wu, and T. S. Chua, "Attentional factorization machines: Learning the weight of feature interactions via attention networks," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 3119–3125, 2017.
- [39] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir *et al.*, "Wide & deep learning for recommender systems," in *Proceedings of the 1st workshop on deep learning for recommender systems*, 2016, pp. 7–10.
- [40] W. Cheng, Y. Shen, and L. Huang, "Adaptive factorization network: Learning adaptive-order feature interactions," in *AAAI*. AAAI Press, 2020, pp. 3609–3616.
- [41] R. Wang, G. Fu, B. Fu, and M. Wang, "Deep & cross network for ad click predictions," *2017 AdKDD and TargetAd - In conjunction with the 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2017*, 2017.
- [42] Y. Qu, H. Cai, K. Ren, W. Zhang, Y. Yu, Y. Wen, and J. Wang, "Product-based neural networks for user response prediction," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 1149–1154, 2017.
- [43] S. Cai, K. Zheng, G. Chen, H. V. Jagadish, B. C. Ooi, and M. Zhang, "Arm-net: Adaptive relation modeling network for structured data," in *SIGMOD Conference*. ACM, 2021, pp. 207–220.

- [44] X. Cai, J. Gao, K. Y. Ngiam, B. C. Ooi, Y. Zhang, and X. Yuan, "Medical concept embedding with time-aware attention," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2018-July, pp. 3984–3990, 2018.
- [45] E. Choi, C. Xiao, W. F. Stewart, and J. Sun, "Mime: Multilevel medical embedding of electronic health records for predictive healthcare," *arXiv preprint arXiv:1810.09593*, 2018.
- [46] X. Peng, G. Long, T. Shen, S. Wang, J. Jiang, and M. Blumenstein, "Temporal self-attention network for medical concept embedding," *Proceedings - IEEE International Conference on Data Mining, ICDM*, vol. 2019-Novem, pp. 498–507, 2019.
- [47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.
- [48] J. Gao, C. Xiao, Y. Wang, W. Tang, L. M. Glass, and J. Sun, "Stagenet: Stage-aware neural networks for health risk prediction," in *Proceedings of The Web Conference 2020*, 2020, pp. 530–540.
- [49] Q. Tan, M. Ye, B. Yang, S. Liu, A. J. Ma, T. C.-F. Yip, G. L.-H. Wong, and P. Yuen, "Data-gru: Dual-attention time-aware gated recurrent unit for irregular multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 930–937.
- [50] Y. Xiang, J. Xu, Y. Si, Z. Li, L. Rasmy, Y. Zhou, F. Tiryaki, F. Li, Y. Zhang, Y. Wu, X. Jiang, W. J. Zheng, D. Zhi, C. Tao, and H. Xu, "Time-sensitive clinical concept embeddings learned from large electronic health records," *BMC Medical Informatics and Decision Making*, vol. 19, no. Suppl 2, 2019.
- [51] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [52] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.
- [53] 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, p. 96, 2019. [Online]. Available: <https://doi.org/10.1038/s41597-019-0103-9>
- [54] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 233–240.
- [55] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.
- [56] C. V. Herbert and D. C. Jackson, "Temperature effects on the responses to prolonged submergence in the turtle *chrysemys picta bellii*. i. blood acid-base and ionic changes during and following anoxic submergence," *Physiological Zoology*, vol. 58, no. 6, pp. 655–669, 1985.
- [57] P. Jansson, A. Larsson, and P. Lönnroth, "Relationship between blood pressure, metabolic variables and blood flow in obese subjects with or without non-insulin-dependent diabetes mellitus." *European journal of clinical investigation*, vol. 28, no. 10, pp. 813–818, 1998.