Published in final edited form as:

ACM BCB. 2021 August ; 2021: . doi:10.1145/3459930.3469513.

KGDAL: Knowledge Graph Guided Double Attention LSTM for Rolling Mortality Prediction for AKI-D Patients

Lucas Jing Liu

Department of Computer Science, University of Kentucky, Lexington, Kentucky, USA

Victor Ortiz-Soriano, Javier A. Neyra

Division of Nephrology, Bone and Mineral Metabolism, University of Kentucky Medical Center, Lexington, Kentucky, USA

Jin Chen

Department of Internal Medicine, Department of Computer Science, University of Kentucky, Lexington, Kentucky, USA

Abstract

With the rapid accumulation of electronic health record (EHR) data, deep learning (DL) models have exhibited promising performance on patient risk prediction. Recent advances have also demonstrated the effectiveness of knowledge graphs (KG) in providing valuable prior knowledge for further improving DL model performance. However, it is still unclear how KG can be utilized to encode high-order relations among clinical concepts and how DL models can make full use of the encoded concept relations to solve real-world healthcare problems and to interpret the outcomes. We propose a novel knowledge graph guided double attention LSTM model named KGDAL for rolling mortality prediction for critically ill patients with acute kidney injury requiring dialysis (AKI-D). KGDAL constructs a KG-based two-dimension attention in both time and feature spaces. In the experiment with two large healthcare datasets, we compared KGDAL with a variety of rolling mortality prediction models and conducted an ablation study to test the effectiveness, efficacy, and contribution of different attention mechanisms. The results showed that KGDAL clearly outperformed all the compared models. Also, KGDAL-derived patient risk trajectories may assist healthcare providers to make timely decisions and actions. The source code, sample data, and manual of KGDAL are available at <https://github.com/lucasliu0928/KGDAL>.

Keywords

Deep Learning; Knowledge Graph; Attention Mechanism; Rolling Mortality Prediction

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

jlui@uky.edu.

1 INTRODUCTION

Acute kidney injury (AKI) is a common complication of hospitalized patients and the incidence increase in patients admitted to the intensive care unit (ICU) [33, 34]. AKI that results in the need for dialysis (AKI-D) is associated with a high risk of hospital mortality [13], and for survivors a risk of incident or progressive chronic kidney disease (CKD) [5, 6, 12, 35], cardiovascular disease [11, 26, 39] or end-stage renal disease (ESRD) [15, 16, 23]. By identifying mortality risk factors from patient individual and population data, providers can implement early intervention strategies leading to better health care and substantially reducing the cost of care.

Numerous factors may influence in-hospital mortality including acute anemia, respiratory failure, electrolytes disarrangements, hemodynamic instability, and demographic information. There is a critical need to identify and correlate these patients and dialysis-specific parameters with inpatient mortality in this specific population. Moreover, accurate prediction of mortality over time (i.e., rolling prediction) in real-world healthcare settings for critically ill patients with AKI-D is needed for better utilization of hospital resources, such as intensifying therapies when is needed, or transitioning patients with a high risk of mortality to comfort care [25, 30].

Two general approaches have been used for mortality predictions in the ICU. The first approach uses clinical scores, including Acute Physiology and Chronic Health Evaluation (APACHE) and Sequential Organ Failure Assessment (SOFA), to identify at-risk patients at any time point [29]. The second approach employs machine learning (ML) methods, such as random forest [22] and SVM [37], to predict mortality risks using electronic health record (EHR) data. With the rapid development of ML techniques, ML-based mortality prediction attracts much attention recently. Nevertheless, ML-based methods mainly focus on mortality prediction at the end of the treatment [21], and the clinical needs for rolling mortality prediction is often overlooked [17, 19]. Traditional ML models rely heavily on feature engineering, requiring not only a deep understanding of the domain knowledge but also tremendous efforts on manual feature extraction and model tuning [32].

In recent years, deep learning (DL) models, including Transformer [36], Long-Short Term Memory (LSTM) [14] and gated recurrent neural networks [8], have shown promising performance on end-to-end patient risk prediction using large-scale EHR data [31]. Doctor AI [9] applies RNN on visit-based medical codes to predict the diagnosis and medication categories for the subsequent visit. However, critical long-term dependency could be lost during RNN training [36]. To address the long-term dependency problem, an attention mechanism was introduced [1] on sequence-to-sequence modeling tasks, where an additive attention mechanism is used in an encoder-decoder model to learn where to pay attention in the input sequence. In the following years, multiple types of attention mechanisms have been proposed [4, 24, 36]. Zhou et al. proposed an attention mechanism along the time-step dimension for relation classification [41], and Transformer [36] was developed to learn attentions without using recurrent network units.

Applications of ML models in clinical settings require a certain level of interpretability. To this end, medical ontologies have been incorporated into DL models using graph-based attention mechanisms. GRAM [10] and DG-RNN [40] are graph-based attention models that are designed to obtain attention from medical knowledge graph (KG) and then use the attention to adjust vector representations of medical codes. However, since attention is purely obtained from KG, the concept completeness and correctness are not considered. In fact, KG is rarely completed and does not have 100% correctness [3]. Furthermore, the graph-based attention model is only feasible on discrete data (e.g., diagnosis codes) but not on continuous temporal data (e.g., blood pressure measurements).

In this study, we propose a Knowledge-Graph Guided Double Attention LSTM (KGDAL) model, aiming to make precise rolling mortality predictions in a real-world healthcare setting for critically ill patients with AKI-D. To our knowledge, KGDAL is the first KG-guided model that extracts both time and feature attention on continuous temporal data. KGDAL has the following advantages:

- KGDAL obtains two-dimensional attention in both the time and feature spaces for improved prediction power and enhanced model interpretability.
- The attention mechanism in the feature space is automatically derived based on the KG rather than manual curation.
- KGDAL can model both continuous and discrete temporal EHR data types.
- KGDAL can make precise rolling mortality predictions for AKI-D patients on two independent clinical datasets.

2 METHOD

The goal of this study is to conduct rolling mortality prediction to assist clinicians in making timely decisions and actions [21]. Mathematically, for each particular subsequence of a patient's EHR data, KGDAL predicts the patient's outcome in the next K hours, where K is the time granularity (e.g., 72 hours). The outcome could be mortality or survival. KGDAL's overall architecture, as shown in Figure 1, contains three phases:

Phase 1. EHR Data Extraction.: Patient clinical features are extracted from patient's EHR. Each feature is matched to a set of concepts in a Knowledge-Graph (KG), followed by manual validation by clinicians. Patient subsequences with random starting time points and varying lengths are generated. The label of each subsequence is "mortality" or "survival" in the next K hours. Static features of each patient are consistent for all the subsequences of the same patient.

Phase 2. Knowledge-Graph Extraction.: The entire KG is used to learn the concept embeddings for every concept identified in Phase 1. The concept embeddings are grouped based on the KG's hierarchical structure, resulting in multiple concept embedding groups. Subsequently, the KG-embedding distances (including both pairwise distances between concept embedding groups and the distances from a concept-of-interest to every concept embedding group) are computed.

Phase 3. Knowledge-Graph Guided Double Attention.: All the temporal features are grouped based on their corresponding concept embedding groups. An LSTM is assigned to each temporal feature group. All the LSTM models are trained simultaneously to minimize the overall loss. Both feature and time attentions are learned using fully connected layers followed by softmax. Double attention is formed using both feature and time attention. The double attention is adopted to adjust feature embeddings for the final prediction and to regularize the prediction loss that minimizes the discrepancy between the attention-based distance and the KG-embedding distance.

2.1 Phase 1. Data Extraction

Given a patient's EHR data $\{S_1, S_2, \dots, S_t, \dots, S_{N_t}\}$, where $S_t \in \mathbb{R}^{N_f}$ is a set of features (e.g. clinical measurements) at time point $t \in \{1, 2, \dots, N_t\}$, N_f is the total number of features, and N_t is the total number of time points. For each patient, we generate N_s subsequences with randomly starting time and random length between 24 and 72 hours (see Figure 1 Phase 1). The outcome label of each subsequence is whether the patient dies or is alive in the next K hours.

Patient clinical features are matched to a set of corresponding concepts in a KG. For features that can be directly matched to a concept in a KG (e.g., a diagnosis code), the corresponding concept will be used. Otherwise, all the associated concepts in the same KG are extracted and then filtered by clinicians. For example, the corresponding concepts of “systolic blood pressure” in HPO [20] are “elevated systolic blood pressure (HP:0004421)” and “decreased systolic blood pressure (HP:0500105)”, both of which are valid and kept for later use.

2.2 Phase 2. Knowledge-Graph Extraction

In phase 2, KGDAL generates concept embeddings and computes the group-wise KG-embedding distances in a KG.

2.2.1 Concept Embedding.—Let E_c be the set of concepts and E_r be all possible relationship types in a KG, a concept relation O can be represented using a triplet denoted as (C_h, r_i, C_t) , where $C_h, C_t \in E_c$ are the head and tail concepts respectively, and $r_i \in E_r$ represents the relationship from C_h to C_t . For example, triplet (“Diabetes mellitus type 1”, “is-a”, “diabetes mellitus”) in SNOMED-CT represents that the Diabetes mellitus type 1 is a subtype of diabetes mellitus.

All the triplets in a KG are used to learn the concept embeddings. The basic idea is to make the learned embeddings of tail concept C_t be close to the sum of the embeddings of head concept C_h and the embeddings of relation r_i . Here, we use TransE [2], one of the most representative translational distance model, to learn the concept embeddings by formulating the problem as follows: given a concept triplet $O = (C_h, r_i, C_t)$ in a KG, we learn the embedding triplets denoted as $G = (\mathbf{h}, \mathbf{l}, \mathbf{t})$, where $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{d_e}$ represents the head and tail concepts embeddings respectively, $\mathbf{l} \in \mathbb{R}^{d_l}$ is the relation vector between $\mathbf{h}$ and $\mathbf{t}$. TransE is trained with negative sampling to learn the embeddings that minimize a margin-based ranking loss function:

$$L_{KG} = \sum_{(h,l,t) \in O} \sum_{(h',l,t') \in O'} \max(0, \gamma + d(h+l, t) - d(h'+l, t')) \quad (1)$$

where O represents the positive samples and O' represents the negative samples that were randomly generated by replacing the head or tail concepts of positive samples. d represents any distance metrics, $\gamma > 0$ is a margin hyper-parameter. Equation 1 shows that the distances in positive samples are minimized where the distances in negative samples are maximized. In this step, the concept embedding denoted as $E_{concept}$ are obtained for each concept associated with the patient clinical features.

2.2.2 Embedding Grouping.—The hierarchical level in a KG represents classes of concepts holding similar characteristics. To capture the commonality within a class and the difference between classes, embedding group E is formed by taking the sum of concept embeddings $E_{concept}$ in each group based on the hierarchical structure of a KG. The number of groups N_g is determined by the number of classes in the user-specified KG level. In general, using a higher level will form more general concept groups, and using a lower level will form more specific concept groups.

2.2.3 Embedding Group-wise Distance.—In order to measure the difference between concept embedding groups as well as how much each group are related to the outcome (i.e., mortality) and to let a DL model pay more attention to the closely related concept embedding groups, two KG-embedding distances are computed. The pairwise distance between concept embedding groups is computed using Eq 2.

$$KG_{dist}(i, j) = dist(E_i, E_j) \quad (2)$$

where E represents a set of concept embedding groups, the subscript represents the group indexes, and $dist$ denotes a distance metric, such as the Euclidean distance.

The distance between each concept embedding group and the concept-of-interest are computed using Eq 3.

$$KG_Target_{dist}(E_i) = dist(E_i, E_{target}) \quad (3)$$

where $KG_Target_{dist}(i)$ represents the distance from the target embedding E_{target} to the i th concept embedding group.

2.3 Phase 3. Knowledge-Graph Guided Double Attention

2.3.1 LSTM Embedding.—By assigning each temporal feature to its corresponding concept embedding group in a KG, a patient's subsequence can be denoted as $\{x_1^i, x_2^i, \dots, x_t^i\}$, where $x_t^i \in \mathbb{R}^{n_i}$ is a list of feature values in the corresponding concept embedding group i at time step t , where i is the group index ($0 \leq i < N_g$), $t \in \{1, 2, \dots, T\}$ is the time step, and n_i is the number of features at each time step for the i th group.

KGDAL consists of N_g LSTM [14] models, each for a feature group. For simplicity, we assume all the feature groups have the equal number of features, the subscript i of n is omitted in the following text. LSTM has three gates, namely the forget gate $\mathbf{f}_t$, the input gate $\mathbf{i}_t$, and the output gate $\mathbf{o}_t$, where $\mathbf{f}_t, \mathbf{i}_t, \mathbf{o}_t \in \mathbb{R}^m$, and m is the dimension of the hidden vectors. Using $\mathbf{c}_t$ and $\mathbf{h}_t$ to represent the cell state vector and the hidden state vector, ($\mathbf{c}_t, \mathbf{h}_t \in \mathbb{R}^m$), the updated LSTM cell in KGDAL can be represented as follows:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{h}_{t-1} + \mathbf{U}_f \mathbf{x}_t + \mathbf{b}_f) \quad (4)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{h}_{t-1} + \mathbf{U}_i \mathbf{x}_t + \mathbf{b}_i) \quad (5)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{h}_{t-1} + \mathbf{U}_o \mathbf{x}_t + \mathbf{b}_o) \quad (6)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{h}_{t-1} + \mathbf{U}_c \mathbf{x}_t + \mathbf{b}_c) \quad (7)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad (8)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (9)$$

where $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_o, \mathbf{W}_c \in \mathbb{R}^{m \times m}$, $\mathbf{U}_f, \mathbf{U}_i, \mathbf{U}_o, \mathbf{U}_c \in \mathbb{R}^{m \times n}$, $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_o, \mathbf{b}_c \in \mathbb{R}^m$ are learnable parameters, σ is a sigmoid function, and $\odot$ is the Hadamard product.

As a result, each LSTM layer outputs a hidden state matrix for each feature group denoted as $\mathbf{H}^{(i)} = [\mathbf{h}_1^{(i)}, \mathbf{h}_2^{(i)}, \dots, \mathbf{h}_T^{(i)}]$, where i is the group index. The hidden state metrics for each feature group is called the feature embedding matrix. Then the feature embedding matrix learned using multiple LSTM layers can be denoted as $\{\mathbf{H}^{(0)}, \mathbf{H}^{(2)}, \dots, \mathbf{H}^{(N_g-1)}\}$, where $\mathbf{H}^{(i)} \in \mathbb{R}^{m_i \times T}$ represents the feature embedding matrix of group i , and m_i is the dimension of hidden state vector from the i th LSTM for group i .

2.3.2 KG-Guided Double Attention.—To model the latent dependencies between different feature groups and at different time steps, KGDAL learns the attentions in both time and feature spaces guided by a KG. The detailed architecture of KGDAL is in Figure 2.

Time Attention.: All the N_g feature embedding matrices are concatenated into one matrix denoted as $\mathbf{H}^C$ with the dimension of $\mathbb{R}^{(m_1 + m_2 + \dots + m_{N_g}) \times T}$. For simplicity, we assume all LSTM layers have equal dimensions of hidden vectors. Hence, the superscript or subscript i of m is omitted in the following text. and the dimension of $\mathbf{H}^C$ is now $\mathbb{R}^{(N_g m) \times T}$. The time attention is computed as follows:

$$\mathbf{M}_\alpha = \tanh(\mathbf{H}^C) \quad (10)$$

$$\alpha = \text{softmax}(\mathbf{M}_\alpha^T \mathbf{w}_\alpha) \quad (11)$$

where $\alpha \in \mathbb{R}^T$ is the time attention, $\mathbf{w}_\alpha \in \mathbb{R}^{N_{g^m}}$ is the learnable parameter, and $\mathbf{M}_\alpha^T$ is the transpose of $\mathbf{M}_\alpha \in \mathbb{R}^{(N_{g^m}) \times T}$.

The time attention mechanism is inspired from and is similar to the work by Zhou et al. [41], but there are two key differences: 1) the input to LSTM is the grouped feature sequences and 2) the time attention mechanism is applied on the output of multiple LSTMs.

Feature Attention.: A similar attention mechanism is used to compute the feature attentions. However, there are two changes. Firstly, we transpose $\mathbf{H}^C$ and use it as the input so that the attention mechanism will be applied on the feature space instead of the time space. Secondly, the KG-embedding distances (KG_Target_{dist}) between the concept embedding groups and the concept-of-interest are used to weight the raw feature attentions. Mathematically, the raw feature attentions is computed using:

$$\mathbf{M}_\beta = \tanh((\mathbf{H}^C)^T) \quad (12)$$

$$\beta = \text{softmax}(\mathbf{M}_\beta^T \mathbf{w}_\beta) \quad (13)$$

where $\beta \in \mathbb{R}^{N_{g^m}}$ is the raw feature attention, $\mathbf{w}_\beta \in \mathbb{R}^T$ is the learnable parameter, and $\mathbf{M}_\beta^T$ is the transpose of $\mathbf{M}_\beta \in \mathbb{R}^{(T) \times N_{g^m}}$.

Then the KG-adjusted feature attention ϵ is equals to β weighted by the KG-embedding distances at corresponding position, which is computed using:

$$\epsilon_{(pos, i)} = \beta_{(pos, i)} \otimes KG_Target_{dist}(i) \quad (14)$$

where $\epsilon \in \mathbb{R}^{N_{g^m}}$, $\otimes$ represents the outer product. pos represents the corresponding positions of feature attentions for each group, where $pos = [im : (i+1)m - 1]$, i is the group index, m is the dimension of the hidden vectors. For example, if the first m raw feature attentions β are obtained for the first feature group, then these m values are weighted by the distance from the first KG-embedding group to the target embedding.

2.3.3 Double Attended Representations.—We combine the attention on both the time and feature spaces by taking the outer product of the time attentions α and a scaled version of KG-adjusted feature attention ϵ to obtain the double attention. In addition, the feature embeddings are then adjusted by the obtained double attention. The double attention Θ and the adjusted feature embeddings $\mathbf{R}$ are computed as follows:

$$\Theta = \epsilon_S \otimes \alpha \quad (15)$$

$$\mathbf{R} = \mathbf{H}^C \odot \Theta \quad (16)$$

where Θ and $\mathbf{R} \in \mathbb{R}^{N_{gm} \times T}$. $\epsilon_S \in \mathbb{R}^{N_{gm}}$ is the scaled feature attention which is computed by taking the ratio between each KG-adjusted feature attention to the first KG-adjusted feature attention ϵ_0 so that the proportion of attentions are maintained, as shown in Equation 17.

$$\epsilon_S = \epsilon / \epsilon_0 \quad (17)$$

Finally, the double-attention adjusted feature embeddings $\mathbf{R}$ are passed into the pooling layer for taking the sum/max over each time step, followed with a dense layer with sigmoid function for final predictions.

2.3.4 Loss Function.—To consolidate the concept relations from KG in attentions, a regularization term is added to the original prediction loss function. The new regularization term minimizes the discrepancy between the pairwise attention-based distances and the pairwise KG-embedding distances.

Let the ground truth label be y and the predicted label be $\hat{y}$, we use the cross-entropy for the original prediction loss denoted as L_{pred} and the regularization term is denoted as L_{reg} , the final loss L is defined as:

$$L = L_{pred} + L_{reg} \quad (18)$$

$$L_{pred} = \sum_{k=1}^{N_s} -(y_k \log(\hat{y}_k) + (1 - y_k) \log(1 - \hat{y}_k)) \quad (19)$$

$$L_{reg} = \sum_{i=1}^{N_g-1} \sum_{j=i+1}^{N_g} (dist(\Theta_i, \Theta_j) - KG_{dist}(i, j)) \quad (20)$$

where Θ represents the double attentions, KG_{dist} is the KG-embedding distance discussed in 2.2.3, and i, j are the group indexes, N_s is the number of samples, k is the sample index.

3 EXPERIMENTS

The performance of KGDAL was evaluated using two (proprietary and public) AKI-D datasets.

3.1 Data Preprocessing

3.1.1 Proprietary EHR Data.—The proprietary EHR data include 608 AKI-D patients who were admitted to the University of Kentucky (UK) HealthCare from January 2009 to October 2019. Among them, 247 (41%) died in the hospital and 361 (59%) survived.

This cohort excluded patients who were less than 18 years old, or were diagnosed with end-stage kidney disease (ESKD) at the time of index hospital admission, or were recipients of kidney transplant. The EHR records during renal replacement therapy (RRT), including both haemodialysis (HD) and continuous renal replacement therapy (CRRT), were extracted. The duration of RRT was limited from 72 hours to 2,000 hours. Any records beyond this range were excluded.

Twelve types of temporal features were collected from EHR, which were systolic blood pressure, diastolic blood pressure, creatinine, bicarbonate, hematocrit, potassium, bilirubin, sodium, temperature, white blood cells (WBC) count, heart rate, and respiratory rate. Six types of static features were also collected, which were demographics (age, race, and gender), admission weight, body mass index (BMI), and Charlson comorbidity score. Three status flags were constructed, which indicated the on or off of CRRT or HD, and the status of being in the ICU or not. In total, 21 features were included in the UK data and the average missing rate of the temporal features was 58.7%.

For each feature, outlier values greater than 97.5 percentile or below 2.5 percentile were both excluded. The temporal granularity of the temporal features was set to one value (median) per hour. Linear interpolation was employed to fill the gaps between two actual measurements if needed. The only exception is creatinine, for which we only kept one value every six hours to maintain the in-practice frequency.

3.1.2 Public EHR Data.—The public data were extracted from the MIMIC-III [18]. We first identified all the AKI patients by the presence of ICD-9 codes of 584.5 to 584.9, then we identified AKI-D patients with the additional presence of ICD-9 procedure codes of 3995 as well as diagnosis codes of V45.11 and V561 [38]. Applying the same cohort exclusion criteria resulted in the MIMIC-III data with 170 AKI-D patients. Among them, 66 (39%) died in the hospital and 104 (61%) survived. The temporal features of MIMIC-III were the same as the UK data except for WBC and temperature, since neither of them was available in the RRT duration. The average missing rate of the temporal features was 49.4%. The same static features and status flags as those in the UK data were included in the MIMIC-III data. In total, 19 features were included in the MIMIC-III data. The same data extraction and outlier detection procedures were conducted.

3.1.3 Knowledge Map.—The Human phenotype ontology (HPO) was used as the clinical knowledge map to learn the concept embeddings. HPO, a widely used biomedical ontology, provides a standardized vocabulary of phenotypic abnormalities encountered in human disease [27]. Concepts in HPO are organized in hierarchies. Among the six sub-hierarchies on the top level, we focused on the “Phenotypic abnormality” sub-hierarchy since it includes most of the concepts of abnormalities related to the selected features in this experiment. To represent the strength of the relations between any two concepts, we counted the number of the common ancestors of the two concepts. Due to the large number of unique descendants in the “Phenotypic abnormality” sub-hierarchy (15,560), we computed the relations between every concept and “Acute Kidney Injury” (HP:0001919) and 100 random selected concepts. Finally, we obtained 1,566,363 identical concept triplets (see definition

in Section 2.2) from HPO, where the number of common ancestors was considered as the relation strength between two concepts.

3.1.4 Experimental Data Generation.—From each patient’s EHR sequence during RRT, we randomly generated at most 30 subsequences with their lengths varying from 48 to 96 hours. For each subsequence in the UK data, the possible class labels are whether the patient died (positive) or survived (negative) 24, 48, or 72 hours after the end of that subsequence. For the MIMIC-III data, the possible class labels of each subsequence are whether the patient died (positive) or survived (negative) 48 or 72 hours after the end of that subsequence. Note that due to the small study cohort size, the “24 hours” label was absent from the MIMIC-III data. In summary, the UK data include 14,757, 15,468, and 16,660 subsequences labeled as negative (alive) and 3,455, 2,744, and 1,552 subsequences labeled as positive (mortality) in the next 72, 48 and 24 hours respectively. For the MIMIC-III data, 4,391 and 4,574 subsequences were labeled as negative (alive); 727 and 544 subsequences were labeled as positive (mortality) in the next 72 and 48, respectively. The ratio between positive and negative ranged between 10% and 25%.

Table 1 showed the data used for training, validation, and testing. All the subsequences of 50 randomly selected patients (25 died vs. 25 alive) were used for validation. All the rest patient data were randomly split into training (90%) and testing (10%). The data split was patient-wise so that the subsequences from the same patient only appeared in one of the three datasets.

3.2 Experiment Results

3.2.1 Concepts Embedding and Concepts Grouping.—Figure 3 shows the HPO concepts matched to the temporal features in both patient cohorts. These concepts were clustered into four groups based on the structure of the “phenotypic abnormality” sub-hierarchy in HPO, which were “Abnormality of the cardiovascular system (Cardiovascular)”, “Abnormality of metabolism/homeostasis (Metabolism)”, “Abnormality of blood and blood-forming tissues (Blood)”, and “Abnormality of the respiratory system (Respiratory)”. Figure 3A shows the partial hierarchical structure of HPO where colors indicate concept groups.

The concept embeddings for all the selected concepts were obtained by training the TransE model using all the HPO concept triplets. The resulting concept embeddings were used to compute two types of pairwise distances, i.e., distances between any two concept embedding groups and distances from the concept-of-interest “AKI” to each concept embedding group. The concept-wise distances were visualized using the t-SNE plot in Figure 3B. “Respiratory” is mostly related to “AKI” (averaged distance 0.02), while “Cardiovascular” is the most distant from “AKI” (averaged distance 0.98).

3.2.2 Performance Comparison.—We compared KGDAL with various mortality rolling prediction models on both the UK data and the MIMIC-III data. In addition, an ablation study was conducted to test whether KGDAL’s KG-adjusted feature attentions were critical in mortality rolling prediction by 1) only using the time attention mechanism, or 2) removing from the loss the KG adjustment that minimizes the discrepancy between the

pairwise attention based distance and the pairwise KG-embedding distances. We compared all the models' performance on both the balanced and imbalanced test sets, the ratio of positive samples (died) to negative samples (survived) are 1:1 and 1: 2, respectively. By randomly sampling five times for each case, we reported the averaged performance on all the evaluation metrics. All the compared models and their inputs are described as following:

- **Random Forest:** The input to this model is the un-grouped temporal features where each temporal feature at each time step is appended as a column, and each static feature is a column. Column-wise mean imputation is used to fill the missing values.
- **Boosted Tree:** We use Extreme Gradient Boosting (XG-Boost) [7] as the second baseline model. The input to this model is the same as Random Forest.
- **LSTM:** The input to a LSTM is the un-grouped temporal features and static features concatenated at each time point.
- **Transformer:** The input to a Transformer is the same as LSTM. We use the encoder part of the original transformer with a dense layer for the prediction task.
- **KGDAL _{α} :** This KGDAL model only uses the time attention mechanism. By removing feature attention, it explores the usefulness of the feature attention mechanism. The inputs are the grouped (KG-guided) features.
- **KGDAL _{$\alpha\beta$} :** This KGDAL model uses both time attention and feature attention, but it removes the KG-adjusted attention and the KG-adjusted loss from KGDAL. It explores the usefulness of the KG-adjusted attention mechanism. The inputs are the grouped (KG-guided) features.

The performance of mortality rolling prediction in the next 72, 48, and 24 hours on the UK data are listed in Table 2, Table 3, and Table 4 respectively. The performance of mortality rolling prediction in the next 72 and 48 hours on MIMIC-III data are listed in Table 5 and Table 6 respectively. In all the tables, precision (PREC), recall (REC), and F1 scores are for the positive (died) class.

As a baseline, a random model achieved ROCAUC 0.52, accuracy 0.52, precision 0.52, recall 0.51, and F1 0.52. The random forest model predicted that almost every patient survived in all experiments. Its average performance on all experiments are ROCAUC 0.50, accuracy 0.58, precision 0.25, recall 0.01 and F1 0.02. The XGBoost model performed better than random forest. However, it is only slightly better than the random model. The performance of deep learning models, including LSTM and Transformer, performed significantly better than the compared traditional machine learning models, indicating that for temporal data based rolling prediction, deep learning models can better capture the critical temporal patterns, resulting in better performance than that of traditional machine learning methods.

Table 2 and Table 3 show that KGDAL has the best performance on 72 and 48-hour rolling prediction on the UK data on almost all evaluation metrics on both balanced and imbalanced test data. The second-best model is KGDAL _{α} followed by KGDAL _{$\alpha\beta$} and Transformer.

However, Table 4 shows that KGDAL_{α} outperformed KGDAL on both the balanced and imbalanced test data for the 24-hour mortality rolling prediction, even though KGDAL has the best recall. It suggests that for shorter prediction windows, the time attention mechanism has more contribution than the feature attention mechanism with/without KG adjustment, and KG-adjusted feature attention can improve the overall model performance slightly. In summary, all the experiments on UK data show that the attention-based models including Transformer, KGDAL_{α} , $\text{KGDAL}_{\alpha\beta}$ have better overall performance than LSTM, indicating the attention per se is critical for rolling prediction tasks.

On the MIMIC-III data, $\text{KGDAL}_{\alpha\beta}$ has the best overall performance for both 72 and 48-hour rolling mortality prediction. While LSTM has the highest ROCAUC and KGDAL achieves better recall, $\text{KGDAL}_{\alpha\beta}$ has higher scores on accuracy, precision, and F1. Surprisingly, KGDAL_{α} , which is the second-best model on the UK data, has the lowest performance among the three in all the experiments on the MIMIC-III data. The performance of KGDAL_{α} on this experiment suggests that it is feature attentions rather than time attentions that play the critical role in MIMIC-III rolling mortality prediction.

In summary, we found that KGDAL's time and feature attentions play an essential role in rolling mortality prediction. While on one data, time attention is more important; feature attention could dominate the model performance on another data. The KG-guided grouping is crucial for all experiments and datasets. Nevertheless, other factors such as sample size and patient distribution may affect model performance as well.

3.2.3 Mortality Risk Trajectory Analysis.—A patient's mortality risk trajectory is a series of predicted risks of all the subsequences of that patient ordered by time. An example trajectory is shown in Figure 5. Given all the mortality risk trajectories, we computed the pairwise trajectory distances using dynamic time wrapping [28], and used hierarchical clustering to identify similar risk trends among all the correctly predicted patients. Two trajectory clusters shown in Figure 4 revealed multiple episodes of increasing risks for non-survivors and quick decreasing risks for survivors. The trend-based analysis may assist healthcare providers in making early decisions before the risk increases.

As a case study, we visualized the mortality risk trajectory of an AKI-D patient in Figure 5. In the figure, the x-axis is the days before outcome event (survival (end of follow-up) / mortality), the y-axis is the predicted risk of death, and the risk scores range from 0 (survived) and 1 (died). Every blue point is the predicted mortality risk from a subsequence of the same patient. All the points were fitted to a smooth curve using polynomial regression showing the trajectory of predicted mortality risks. In the case study, the patient finally survived. However, the risk score was not monotonically decreasing. Starting with a high risk, KGDAL predicted that the risk was gradually decreased for 5 days. The risk trajectory then stayed roughly at 50% for 10 days with mild fluctuations. Finally, the trajectory decreased quickly in the last 5 days of RRT.

The corresponding KG-adjusted attentions are shown in Figure 6, where each row represents the attentions of a feature group, each column is a time point (the darker the color, the high the score). The risk trajectory aligned well with the KG-adjusted attention. The attention

hit three times the maximum in the early RRT duration, indicating a high mortality risk at that time. The figure also reveals that “Respiratory” and “Blood” had overall higher attentions than the other concept groups. This is well-aligned with the observations in the knowledge graph that “Respiratory”, “Blood”, and “AKI” are closely related concepts. The high agreement between the risk trajectory and the time and feature attentions suggests that the attentions obtained from KGDAL may be useful to explain to clinicians the potential risks and why the risk is high or low so that interventions can be taken in place timely.

4 CONCLUSION

In this work, we developed a novel model called KGDAL for rolling mortality prediction for AKI-D patients. KGDAL uses a knowledge graph to guide the generation of 2D attention in both time and feature spaces. KGDAL and its variations achieved the best performance on both the UK data and the MIMIC-III data. Using a case study, we demonstrated the interpretability of KGDAL and the capability of using KGDAL for assisting timely decisions for clinicians.

For future work, we will further optimize the architecture of KGDAL, adjust the training and testing processes, and visualize the resulting trajectories for clinicians. Also, we will incorporate a customized knowledge graph to better capture the relationships between the temporal features for patients from different institutions and further validate KGDAL on large-scale cross-institutional datasets.

ACKNOWLEDGMENTS

This work is supported by NIDDK R56 DK126930 (PI JAN) and P30 DK079337.

REFERENCES

- [1]. Bahdanau Dzmitry, Cho Kyunghyun, and Bengio Yoshua. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).
- [2]. Bordes Antoine, Usunier Nicolas, Garcia-Duran Alberto, Weston Jason, and Yakhnenko Oksana. 2013. Translating embeddings for modeling multi-relational data. In Neural Information Processing Systems (NIPS). 1–9.
- [3]. Cao Yixin, Wang Xiang, He Xiangnan, Hu Zikun, and Chua Tat-Seng. 2019. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In The world wide web conference. 151–161.
- [4]. Chaudhari Sneha, Polatkan Gungor, Ramanath Rohan, and Mithal Varun. 2019. An attentive survey of attention models. arXiv preprint arXiv:1904.02874 (2019).
- [5]. Chawla Lakhmir S, Amdur Richard L, Amodeo Susan, Kimmel Paul L, and Palant Carlos E. 2011. The severity of acute kidney injury predicts progression to chronic kidney disease. *Kidney international* 79, 12 (2011), 1361–1369. [PubMed: 21430640]
- [6]. Chawla Lakhmir S, Eggers Paul W, Star Robert A, and Kimmel Paul L. 2014. Acute kidney injury and chronic kidney disease as interconnected syndromes. *New England Journal of Medicine* 371, 1 (2014), 58–66.
- [7]. Chen Tianqi and Guestrin Carlos. 2016. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 785–794.

- [8]. Cho Kyunghyun, Merriënboer Bart Van, Bahdanau Dzmitry, and Bengio Yoshua. 2014. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014).
- [9]. Choi Edward, Bahadori Mohammad Taha, Schuetz Andy, Stewart Walter F, and Sun Jimeng. 2016. Doctor ai: Predicting clinical events via recurrent neural networks. In Machine learning for healthcare conference. PMLR, 301–318.
- [10]. Choi Edward, Bahadori Mohammad Taha, Song Le, Stewart Walter F, and Sun Jimeng. 2017. GRAM: graph-based attention model for healthcare representation learning. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. 787–795.
- [11]. Gammelager Henrik, Christiansen Christian Fynbo, Johansen Martin Berg, Tønnesen Else, Jespersen Bente, and Sørensen Henrik Toft. 2014. Three-year risk of cardiovascular disease among intensive care patients with acute kidney injury: a population-based cohort study. *Critical care* 18, 5 (2014), 1–10.
- [12]. Heung Michael, Steffick Diane E, Zivin Kara, Gillespie Brenda W, Banerjee Tanushree, Hsu Chi-yuan, Powe Neil R, Pavkov Meda E, Williams Desmond E, Saran Rajiv, et al. 2016. Acute kidney injury recovery pattern and subsequent risk of CKD: an analysis of veterans health administration data. *American Journal of Kidney Diseases* 67, 5 (2016), 742–752. [PubMed: 26690912]
- [13]. Hickson LaTonya J, Chaudhary Sanjay, Williams Amy W, Dillon John J, Norby Suzanne M, Gregoire James R, Albright Robert C Jr, McCarthy James T, Thorsteinsdottir Bjorg, and Rule Andrew D. 2015. Predictors of outpatient kidney function recovery among patients who initiate hemodialysis in the hospital. *American Journal of Kidney Diseases* 65, 4 (2015), 592–602. [PubMed: 25500361]
- [14]. Hochreiter Sepp and Schmidhuber Jürgen. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780. [PubMed: 9377276]
- [15]. Hsu Raymond K and Hsu Chi-yuan. 2016. The role of acute kidney injury in chronic kidney disease. In *Seminars in nephrology*, Vol. 36. Elsevier, 283–292. [PubMed: 27475659]
- [16]. Ishani Areef, Xue Jay L, Himmelfarb Jonathan, Eggers Paul W, Kimmel Paul L, Molitoris Bruce A, and Collins Allan J. 2009. Acute kidney injury increases risk of ESRD among elderly. *Journal of the American Society of Nephrology* 20, 1 (2009), 223–228. [PubMed: 19020007]
- [17]. Johnson Alistair EW and Mark Roger G. 2017. Real-time mortality prediction in the Intensive Care Unit. In *AMIA Annual Symposium Proceedings*, Vol. 2017. American Medical Informatics Association, 994. [PubMed: 29854167]
- [18]. Johnson Alistair EW, Pollard Tom J, Shen Lu, Lehman Li-wei H, Feng Mengling, Ghassemi Mohammad, Moody Benjamin, Szolovits Peter, Celi Leo Anthony, and Mark Roger G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3 (2016), 160035. [PubMed: 27219127]
- [19]. Kim Soo Yeon, Kim Saehoon, Cho Joongbum, Kim Young Suh, Sol In Suk, Sung Youngchul, Cho Inhyeok, Park Minseop, Jang Haerin, Kim Yoon Hee, et al. 2019. A deep learning model for real-time mortality prediction in critically ill children. *Critical care* 23, 1 (2019), 1–10. [PubMed: 30606235]
- [20]. Köhler Sebastian, Gargano Michael, Matentzoglou Nicolas, Carmody Leigh C, Lewis-Smith David, Vasilevsky Nicole A, Danis Daniel, Balagura Ganna, Baynam Gareth, Brower Amy M, et al. 2021. The Human Phenotype Ontology in 2021. *Nucleic acids research* 49, D1 (2021), D1207–D1217. [PubMed: 33264411]
- [21]. Koyner Jay L, Carey Kyle A, Edelson Dana P, and Churpek Matthew M. 2018. The development of a machine learning inpatient acute kidney injury prediction model. *Critical care medicine* 46, 7 (2018), 1070–1077. [PubMed: 29596073]
- [22]. Lin Ke, Hu Yonghua, and Kong Guilan. 2019. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *International journal of medical informatics* 125 (2019), 55–61. [PubMed: 30914181]
- [23]. Lo Lowell J, Go Alan S, Chertow Glenn M, McCulloch Charles E, Fan Dongjie, Ordoñez Juan D, and Hsu Chi-yuan. 2009. Dialysis-requiring acute renal failure increases the risk of progressive chronic kidney disease. *Kidney international* 76, 8 (2009), 893–899. [PubMed: 19641480]

- [24]. Luong Minh-Thang, Pham Hieu, and Manning Christopher D. 2015. Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025 (2015).
- [25]. Malhotra Rakesh, Kashani Kianoush B, Macedo Etienne, Kim Jihoon, Bouchard Josee, Wynn Susan, Li Guangxi, Ohno-Machado Lucila, and Mehta Ravindra. 2017. A risk prediction score for acute kidney injury in the intensive care unit. *Nephrology Dialysis Transplantation* 32, 5 (2017), 814–822.
- [26]. Odutayo Ayodele, Wong Christopher X, Farkouh Michael, Altman Douglas G, Hopewell Sally, Emdin Connor A, and Hunn Benjamin H. 2017. AKI and long-term risk for cardiovascular events and mortality. *Journal of the American Society of Nephrology* 28, 1 (2017), 377–387. [PubMed: 27297949]
- [27]. Robinson Peter N, Köhler Sebastian, Bauer Sebastian, Seelow Dominik, Horn Denise, and Mundlos Stefan. 2008. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics* 83, 5 (2008), 610–615. [PubMed: 18950739]
- [28]. Sardá-Espinosa Alexis. 2019. Time-Series Clustering in R Using the dtwclust Package. *The R Journal* (2019). 10.32614/RJ-2019-023
- [29]. Schoe Abraham, Bakhshi-Raiez Ferishta, de Keizer Nicolette, van Dissel Jaap T, and de Jonge Evert. 2020. Mortality prediction by SOFA score in ICU-patients after cardiac surgery; comparison with traditional prognostic-models. *BMC anesthesiology* 20, 1 (2020), 1–8. [PubMed: 31898488]
- [30]. Shawwa Khaled, Ghosh Erina, Lanius Stephanie, Schwager Emma, Eshelman Larry, and Kashani Kianoush B. 2020. Predicting acute kidney injury in critically ill patients using comorbid conditions utilizing machine learning. *Clinical Kidney Journal* (2020).
- [31]. Shickel Benjamin, Tighe Patrick James, Bihorac Azra, and Rashidi Parisa. 2017. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE journal of biomedical and health informatics* 22, 5 (2017), 1589–1604. [PubMed: 29989977]
- [32]. Singh Anima, Nadkarni Girish, Gottesman Omri, Ellis Stephen B, Bottinger Erwin P, and Gutttag John V. 2015. Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration. *Journal of biomedical informatics* 53 (2015), 220–228. [PubMed: 25460205]
- [33]. Susantitaphong Paweena, Cruz Dinna N, Cerda Jorge, Abulfaraj Maher, Alqahtani Fahad, Koulouridis Ioannis, and Jaber Bertrand L. 2013. World incidence of AKI: a meta-analysis. *Clinical Journal of the American Society of Nephrology* 8, 9 (2013), 1482–1493. [PubMed: 23744003]
- [34]. Szczech Lynda A, Harmon William, Hostetter Thomas H, Klotman Paul E, Powe Neil R, Sedor John R, Smedberg Paul, and Himmelfarb Jonathan. 2009. World Kidney Day 2009: problems and challenges in the emerging epidemic of kidney disease.
- [35]. Thakar Charuhas V, Christianson Annette, Himmelfarb Jonathan, and Leonard Anthony C. 2011. Acute kidney injury episodes and chronic kidney disease risk in diabetes mellitus. *Clinical journal of the American Society of Nephrology* 6, 11 (2011), 2567–2572. [PubMed: 21903988]
- [36]. Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, Kaiser Lukasz, and Polosukhin Illia. 2017. Attention is all you need. arXiv preprint arXiv:1706.03762 (2017).
- [37]. Verplancke Thierry, Van Looy Stijn, Benoit Dominique, Vansteelandt Stijn, Depuydt Pieter, De Turck Filip, and Decruyenaere Johan. 2008. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. *BMC medical informatics and decision making* 8, 1 (2008), 1–8. [PubMed: 18171485]
- [38]. Waikar Sushrut S, Wald Ron, Chertow Glenn M, Curhan Gary C, Winkelmayer Wolfgang C, Liangos Orfeas, Sosa Marie-Anne, and Jaber Bertrand L. 2006. Validity of international classification of diseases, ninth revision, clinical modification codes for acute renal failure. *Journal of the American Society of Nephrology* 17, 6 (2006), 1688–1694. [PubMed: 16641149]
- [39]. Wu Vin-Cent, Wu Che-Hsiung, Huang Tao-Min, Wang Cheng-Yi, Lai Chun-Fu, Shiao Chih-Chung, Chang Chia-Hsui, Lin Shuei-Liong, Chen Yen-Yuan, Chen Yung-Ming, et al. 2014. Long-

term risk of coronary events after AKI. *Journal of the American Society of Nephrology* 25, 3 (2014), 595–605. [PubMed: 24503241]

- [40]. Yin Changchang, Zhao Rongjian, Qian Buyue, Lv Xin, and Zhang Ping. 2019. Domain Knowledge guided deep learning with electronic health records. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 738–747.
- [41]. Zhou Peng, Shi Wei, Tian Jun, Qi Zhenyu, Li Bingchen, Hao Hongwei, and Xu Bo. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*. 207–212.

CCS CONCEPTS

• Computing methodologies → Neural networks; • Applied computing → Health informatics.

Figure 1:

The three phases of the Knowledge-Graph Guided Double Attention (KGDAL) model for rolling mortality prediction for critically ill patients with AKI-D in real-world healthcare settings.

Figure 2:
The detailed architecture of the Knowledge-Graph Guided Double Attention LSTM (KGDAL) model.

Figure 3:

A: The partial hierarchical structure of the Human Phenotype Ontology (HPO) that includes the following concepts. Colors indicate different concept groups (Red: Acute Kidney injury (AKI); Orange: “Cardiovascular”; Green: “Metabolism”; Blue: “Blood”; Pink: “Respiratory”). B: The similarities of the same selected features in a projected space generated using t-SNE.

Figure 4:
Two risk trajectory clusters with different endings.

Figure 5:
The risk trajectory of a survival patient.

Figure 6:
An example of the KG-adjusted 2-D attentions.

Table 1:

Training, validation and testing data at the subsequence level.

	Proprietary EHR dataset (UK data)						Public EHR data (MIMIC-III)					
	In next 72 hours		In next 48 hours		In next 24 hours		In next 72 hours		In next 48 hours		In next 24 hours	
	Alive	Death	Alive	Death	Alive	Death	Alive	Death	Alive	Death	Alive	Death
Train	12275	2643	12817	2101	13717	1201	2725	391	2831	285	1337	155
Validation	1077	419	1147	349	1305	191	1281	211	406	104		
Test	1405	393	1504	294	1638	160	385	125				

Table 2:

Performance of morality prediction in the next 72 hours during RRT (UK data)

	$N_{pos} : N_{neg} = 1 : 1$						$N_{pos} : N_{neg} = 1 : 2$					
	ROCAUC	ACC	PREC	REC	F1		ROCAUC	ACC	PREC	REC	F1	
XGBoost	0.50	0.50	0.51	0.35	0.42		0.50	0.55	0.33	0.35	0.34	
LSTM	0.62	0.61	0.60	0.67	0.63		0.63	0.60	0.44	0.67	0.53	
Transformer	0.70	0.64	0.64	0.64	0.64		0.69	0.63	0.46	0.64	0.53	
KGDAL _{α}	0.75	0.66	0.64	0.77	0.69		0.75	0.64	0.47	0.77	0.59	
KGDAL _{$\alpha\beta$}	0.70	0.63	0.63	0.63	0.63		0.70	0.63	0.46	0.63	0.53	
KGDAL	0.76	0.71	0.66	0.87	0.75		0.74	0.64	0.48	0.87	0.62	

Table 3:

Performance of morality prediction in the next 48 hours during RRT (UK Data)

	$N_{pos} : N_{neg} = 1 : 1$						$N_{pos} : N_{neg} = 1 : 2$					
	ROCAUC	ACC	PREC	REC	F1		ROCAUC	ACC	PREC	REC	F1	
XGBoost	0.50	0.49	0.49	0.35	0.41		0.50	0.55	0.33	0.35	0.34	
LSTM	0.61	0.60	0.59	0.67	0.63		0.62	0.59	0.43	0.67	0.52	
Transformer	0.69	0.63	0.62	0.66	0.64		0.70	0.63	0.46	0.66	0.54	
KGDAL _{α}	0.74	0.67	0.63	0.80	0.70		0.75	0.63	0.47	0.80	0.59	
KGDAL _{$\alpha\beta$}	0.70	0.64	0.63	0.66	0.64		0.72	0.64	0.47	0.66	0.55	
KGDAL	0.73	0.68	0.63	0.88	0.74		0.74	0.63	0.47	0.88	0.61	

Table 4:

Performance of morality prediction in the next 24 hours during RRT (UK Data)

	$N_{pos} : N_{neg} = 1 : 1$						$N_{pos} : N_{neg} = 1 : 2$					
	ROCAUC	ACC	PREC	REC	F1		ROCAUC	ACC	PREC	REC	F1	
XGBoost	0.50	0.51	0.51	0.37	0.43		0.50	0.55	0.34	0.37	0.35	
LSTM	0.62	0.62	0.61	0.69	0.65		0.61	0.58	0.42	0.69	0.53	
Transformer	0.72	0.65	0.64	0.71	0.67		0.71	0.63	0.46	0.71	0.56	
KGDAL _{<i>a</i>}	0.78	0.71	0.66	0.89	0.76		0.78	0.65	0.49	0.89	0.63	
KGDAL _{<i>ap</i>}	0.75	0.68	0.66	0.76	0.71		0.75	0.65	0.48	0.76	0.59	
KGDAL	0.75	0.70	0.64	0.94	0.76		0.75	0.62	0.47	0.94	0.62	

Table 5:

Performance of morality prediction in the next 72 hours during RRT (MIMIC-III)

	$N_{pos} : N_{neg} = 1 : 1$						$N_{pos} : N_{neg} = 1 : 2$					
	ROCAUC	ACC	PREC	REC	F1	F1	ROCAUC	ACC	PREC	REC	F1	F1
XGBoost	0.51	0.49	0.47	0.15	0.23	0.23	0.50	0.59	0.28	0.15	0.20	0.20
LSTM	0.70	0.64	0.80	0.38	0.52	0.52	0.69	0.72	0.65	0.38	0.48	0.48
Transformer	0.64	0.60	0.69	0.38	0.49	0.49	0.62	0.67	0.50	0.38	0.44	0.44
KGDAL _a	0.57	0.38	0.08	0.02	0.04	0.04	0.56	0.49	0.04	0.02	0.03	0.03
KGDAL _{ap}	0.69	0.72	0.97	0.46	0.62	0.62	0.68	0.81	0.97	0.46	0.62	0.62
KGDAL	0.65	0.59	0.58	0.62	0.60	0.60	0.63	0.56	0.40	0.62	0.48	0.48

Table 6:

Performance of morality prediction in the next 48 hours during RRT (MIMIC-III)

	$N_{pos} : N_{neg} = 1 : 1$						$N_{pos} : N_{neg} = 1 : 2$					
	ROCAUC	ACC	PREC	REC	F1		ROCAUC	ACC	PREC	REC	F1	
XGBoost	0.50	0.50	0.49	0.17	0.25		0.50	0.60	0.32	0.17	0.22	
LSTM	0.71	0.64	0.79	0.38	0.52		0.70	0.71	0.61	0.38	0.47	
Transformer	0.66	0.59	0.67	0.37	0.47		0.65	0.65	0.48	0.37	0.41	
KGDAL _a	0.59	0.39	0.08	0.02	0.03		0.57	0.49	0.03	0.02	0.02	
KGDAL _{ap}	0.67	0.70	0.92	0.44	0.60		0.66	0.79	0.84	0.44	0.58	
KGDAL	0.62	0.56	0.56	0.59	0.57		0.61	0.53	0.37	0.59	0.46	