Readmission Prediction with Knowledge Graph Attention and RNN-based Ordinary Differential Equations

Su Pei[1], Ke Niu[1], Xueping Peng[2], and Jingni Zeng[1]

[1] Beijing Information Science and Technology University, Beijing, China
{peisu1128, niuke, zjn981223}@bistu.edu.cn
[2] Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia
xueping.peng@uts.edu.au

Abstract. Predicting the readmission risk within 30 days on the Electronic Health Record (EHR) has been proven crucial for predictive analytics in healthcare domain. Deep-learning-based models are recently utilized to address this task since those models can relatively improve prediction performance and work as decision aids, which helps reduce unnecessary readmission and recurrence risk. However, existing prediction models, limited by fuzzy relevance of patient data, are unable to get higher prediction accuracy due to data noise generated by patients with different disease types. To solve this problem, we propose an end-to-end model called GROM, which integrates knowledge graph to alleviate the interference of data noise generated in the processing of irregularity dynamic clinical data with neural ordinary differential equation (ODE). The experimental results show that our model achieved the highest average precision and proved that the graph attention mechanism is suitable to improve performance of predicting the risk of readmission within 30 days.

Keywords: Deep learning · Knowledge Graph · Electronic Health Record · ICU Readmission Prediction.

1 Introduction

In recent years, with the continuous development and advancement of medical informatization technology, a large quantity of electronic data, such as Electronic Health Record (EHR) [17], have been generated. How to effectively utilize the valuable information hidden behind these data to benefit a large number of patients has raised the attentions from both researchers and practitioners [10][12][11]. One of the numerous analytical tasks is to predict the future readmission [18] based on a patient's historical EHR data. Readmission prediction can assist doctors to make clinical decisions, reduce the cost of readmission and the risk of relapse after discharge. According to previous research [5], about 10% of the critically ill patients may re-enter ICU, which is a negative indicator to

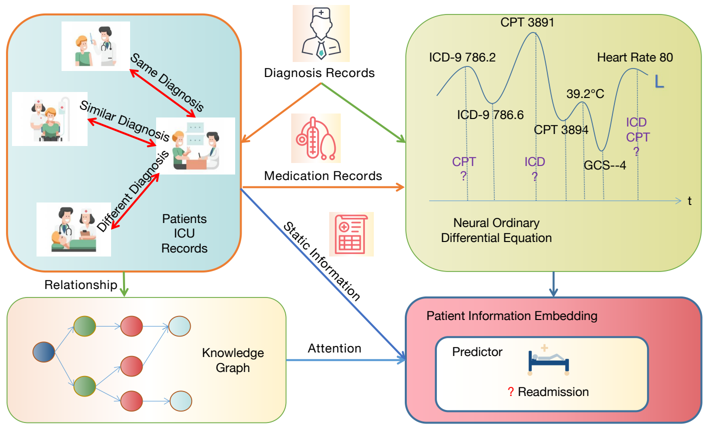

Fig. 1. How to use the relationship between patients' condition to predict the accuracy of readmission.

therapeutic effect. Therefore, building predictive models with ICU-related data to provide doctors with auxiliary diagnosis is an important issue of high application value. In this paper, we apply deep learning techniques to ICU readmission prediction.

With the development of deep learning technology, the research on the prediction of ICU readmission continue to develop. As shown on the upper right of Fig 1, some researches utilized discrete data with irregular time intervals, such as physiological measurements and procedure codes. The absence of these data at certain time points made it impossible to directly use them to represent the complete patient treatment process. To solve this problems, some researchers represented timestamp codes by adding time-related information("embedding") to digital vectors, and modified the internal workings of recursive cell method with ordinary differential equations [4]. However, the model was still insufficient in predicting. When using electronic health records to predict, the interaction between static variables and the nonlinear correlation between static variables and predicted risks were not considered [1]. As shown on the left of Fig 1, existing readmission researches lacked effective treatment of the relationship between patients with different disease types and could not obtain reliable representation of the relationship between patients. Given the correlation between nonlinear static variables, researchers often rely on additional information provided by experts in the hierarchical information of diagnostic codes to construct knowledge graphs to strengthen the connections between data. Current researches together show that when the data quantity is limited, graph attention model uses figure of the parent-child relationship to learn robust representation [3, 9], connecting different patients' health information, enhancing the influence of each other, and finally improving the forecasting accuracy under the conditions of using similar disease information.

Considering that we need to use static data to interact with each other, graph attention can use ontological information related to data volumes to determine the specificity of medical concepts. When there were fewer medical concepts observed in the data, their ancestors would gain more weight and thus be able to understand the data more accurately and to provide general (coarse-grained) information about their children. It could be seen that this method was suitable for us to make up the existing defects. Therefore, as shown in Fig 1, in order to solve the problems existing in previous algorithms, we proposed an end-to-end approach called GROM (**G**raph Attention and **R**NN-based Neural **O**rdinary Differential Equations **M**odel), which integrates RNN-based ODE model with graph-based attention mechanism to improve prediction performance. The proposed model constructs a knowledge graph through the diagnostic code scoring mechanism, strengthens the relationship between patients, and provides help for prediction and patient information to readjust similar diagnostic results.

In order to verify whether the effect of our model on ICU readmission meets our expectations, we used Medical Information Mart for Intensive Care III (MIMIC-III) data sets [6] in experiments. Through experimental comparison, our model can solve the lack of data correlation in the original research well, and achieve better prediction accuracy using the graph attention mechanism. Our main contributions are summarized as follows:

- We investigate the relationship among patient conditions to predict the accuracy of readmission according to patient's static and dynamic data.
- We propose GROM, an end-to-end, robust model to accurately predict patients' future readmission with mutual integration of medical knowledge graph and RNN-based ODE.
- We evaluate the proposed model on a real-world data set, while demonstrates that the GROM is superior to all the comparative methods.

The remainder of this paper is organized as follows. Details about our model are presented in Section 2. And next, in Section 3, we demonstrate the experimental results conducted on real-world dataset. Lastly, we conclude our work in Section 4.

2 Method

In our research, GROM, a model we proposed based on graph attention with RNN-based ODE, is used to predict the risk of readmission within 30 days in ICU. The RNN-based ODE is the basis model which uses multilayers of the network to process time series associated with patient information in prediction. And graph attention mechanism is utilized in this model to learn the knowledge graph of patient diagnostic code to improve the accuracy of model predictions. The overall architecture of the proposed model is shown in Fig 2, the left part of the model is the graph attention module. The knowledge graph obtained from the

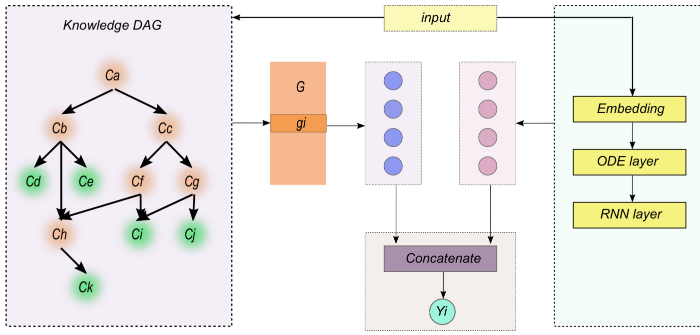

Fig. 2. The graph attention matrix G operates on all entities with diagnostic code and is fully joined to the ODE&RNN model.

input data according to the Clinical Classifications Software (CCS) classification standard [15] is embedded into the graph matrix, and then the graph attention matrix is embedded into the right RNN sequence. In this section, we will describe each layer of the model in detail and Algorithm 1 describes the overall training procedure of the proposed GROM.

2.1 RNN-based ODE Layer

There are a large volume of irregular sampling to obtain patient-related time information, including diagnostic and procedural codes, as well as medication and vital sign codes, which are mapped to the corresponding "embeddings". Diagnoses&procedures and medication&vital signs are processed separately, as they are measured on different time scales [13]. It is difficult for such information to be directly applied to the neural network, so we use the Shen Chang differential equation model proposed by Chen [2], which is very attractive to the processing of time series. In ODEs, the continuously defined dynamic information can be naturally incorporated into any data arriving at any given time [1]. So, we calculate the time-aware code embedded in the ODE dynamic simulation, and neural nodes are used to simulate the dynamic process of embedding.

To make better use of the data's timestamp information and be able to make predictions at any point in time, the neural ODE models the time series as a continuous trajectory of change. Each trajectory is determined by the local initial state S_{t_0} and the potential dynamic global set shared by all time series. Given observation t_0, t_1, ..., t_n and initial state S_{t_0}, an ODE solver produces S_{t_1}, ..., S_{t_n}, which describe the latent state at each observation. We define this generative model formally through a sampling procedure:

$$\begin{aligned} s_{t_0} &\sim p(s_{t_0}) \\ s_{t_1}, s_{t_2}, ..., s_{t_N} &= \texttt{ODESlove}(s_{t_0}, f, \theta f, t_0, ..., t_N) \\ each x_{t_i} &\sim p(x|s_{t_i}, \theta x) \end{aligned} \tag{1}$$

Function f is a time-invariant function that takes the value s at the current time step and outputs the gradient: $\partial s(t)/\partial t = f(s(t), \theta f)$. This function is parametrized using a neural network. Because f is time-invariant, given any latent state s(t), the entire latent trajectory is uniquely defined. Extrapolating this latent trajectory lets us make predictions arbitrarily far forwards or backwards in time.

In GROM, we, using adjunction sensitivity to calculate gradients, treat the ODE approximately as a black box layer to process an irregularly sampled time series in the data. After the ODE layer processes the time information, the processed information is passed to the RNN layer for further processing. Bidirectional RNN is used to overcome the drawback that the prediction accuracy decreases with the increase of sequence length, and to deal with the gradient disappearance.

2.2 Graph Attention Layer

In GROM, we implement a layer based on graph attention mechanism. Graph attention layer is introduced into GROM as a layer, and the result is splicing with the result vector generated by the ODE layer to produce the output after further processing.

To make better use of patients' limited treatment information, the graph attention layer was introduced in this study. In robust representation of medical code, the embedded sequences of the relationship among the medical ontology encoding, hierarchical clinical constructs and medical concepts are usually of arbitrary length and need to be integrated into a fixed-size vector for further processing. Therefore, for the directed acyclic graph with ICD-9 [14] relation obtained by CCS classification, each medical concept node is assigned a basic embedding vector E_I, and the basic embedding is combined with its ancestor nodes through graph-based attention mechanism to obtain the final embedding vector M_I of the i-th medical code. Graph attention mechanisms, such as dot product attention, calculate the weighted average of embedded code, and higher weights are assigned to the most relevant code. Information can be integrated for further processing by using the final memory state of the recursive unit or by applying a graph attention mechanism to the output vector set:

$$g_i = \sum_{j \in A(i)} a_{ij} e_j, \quad a_{ij} \geq 0 \quad for \quad j \in A(i) \tag{2}$$

In this work, recursive cells are realized by bidirectional gated recursive unit. The information related to the graph is embedded into the vectors by applying the exponential decay dot product to the graph weight matrix. In the equation above, g_i is the final representation of code c_i, $A(i)$ is the index of code c_i and its ancestors, e_j is the basic embedding of code c_j, and when calculating g_i, a_{ij} is the weight of concern for embedding e_j.

$$a_{ij} = \frac{exp(f(e_i, e_j))}{\sum_{k \in A(i)} exp(f(e_i, e_k))} \tag{3}$$

Algorithm 1 The GROM model

Randomly initialize diagnoses and procedures, medications and vital signs, and related time information embedding matrix DP, CP;
repeat
  for visit v_t in X **do**
    Calculate the knowledge graph embedded matrix G (see Section 2.2);
    Fusion state information, DP, CP and G;
    Make prediction yˆt using Softmax function;
  end for
  Calculate the prediction loss L;
  Update parameters according to the gradient of L;
until convergence

The $f(e_i, e_j)$ is a scalar value representing the basic embedding compatibility of e_i and e_k, and calculation formula of $f(e_i, e_j)$ is as follows:

$$f(e_i, e_j) = u_a^T tanh(W_a \begin{bmatrix} e_i \\ e_j \end{bmatrix} + b_a) \tag{4}$$

In the equation above, W_a is the weight matrix splicing e_i and e_j, b is the bias vector, and u_a is the weight vector generating scalar values. All the obtained g_i are connected to obtain the embedded representation of the required diagnostic code, and then the embedding matrix G is sent to the graph attention layer of the model for processing. To help with subsequent interpretation without changing network capacity, the vector of fixed size generated by the graph attention mechanism is reduced to a fraction of two scalar values (one related to diagnosis&procedure and the other related to medication&vital signs). Use a fully connected layer with linear activation functions.

3 Experiment

In this section, we performed several comparison experiments on the large public medical electronic medical records MIMIC-III data set[1] to evaluate the performance in ICU readmission prediction of the proposed GROM. This section includes three parts: Data Description, Experiments Setup and Results and Discuss.

3.1 Data Description

Data set. The algorithm was evaluated on publicly available MIMIC-III data set (no ethical approval is required). In our experiment, the supervised learning task consist of predicting whether the patient will be readmitted to the ICU within 30 days from discharge for a given ICU stay. The final data set comprised of 45,298 ICU stays for 33,150 patients, labelled as either positive (N=5,495) or

[1] https://mimic.physionet.org/

negative (N=39,803) depending on whether a patient did or did not experience readmission within 30 days from discharge. To develop and evaluate algorithms, patients based on patient identification were randomly subdivided into training and validation (90%) and test sets 10%.

Data Pre-processing. The data of a patient can be represented as a set of significant static variables and timestamped codes. In our research, static variables included the patient's gender, age, ethnicity, insurance type, marital status, the previous location of the patient prior to arriving at the hospital, and whether the patient was admitted for elective surgery. The importance of static data had a similar characteristic proportion in both positive and negative samples. Data types of timestamped codes included international classification of diseases and related health problems (ICD-9) diagnose and procedure codes, prescribed medications, and patient vital signs. Overall, the models were trained using 23 static variables, 992 unique ICD-9 diagnostic codes, 298 unique ICD-9 program codes, 586 unique medication therapy codes, and 32 codes related to vital signs. The record for each patient contained up to 552 ICD-9 diagnose and procedure codes, as well as 392 medications and vital signs codes related to the current ICU hospitalization.

3.2 Experiment Setup

Baseline Models. To verify the predictive performance of the proposed GROM, we compared it with the following four methods:

GROM. Dynamics in the time of embeddings were modelled using graph attention layer and neural ODEs, embeddings were passed to RNN layers, the final memory states were used for further processing.

ODE+RNN+Attention. Dynamic time in the patient embedding information was modelled using the neural ODE, the modeled information was passed through the RNN layer, the final memory state of the RNN was used for further processing.

ODE+RNN. Dynamics in the time of embeddings were modelled using neural ODEs, embeddings were passed to RNN the layers, the final memory states were used for further processing.

SVM. Support Vector machines.

RNN. The embedding information of patient was directly through the RNN layer, the final memory state was used for further processing.

Implementation. To compare several neural network architecture's classification accuracy in training, the maximum likelihood estimation of network parameters were obtained using log-loss cost function in the training data. The RNN and graph attention layer were embedded by dropout of 0.5, and the Adam optimizer with random gradient descent was used (batch size was 128 and the learning rate was 0.001) [7]. Considering the imbalance between classes, the proportionally increased misclassification overhead was allocated to fewer classes [16]. The training ended after 80 epochs because with the additional training of epochs (based on the average accuracy of the validation data), the over-fitting of the training data became apparent.

Ablation Study Design. In order to obtain the results of data ablation, the model deals with the detailed structure of different types of data. For dynamic data of diagnosis, prescription information sequence and event information sequence, ordinary differential equation can effectively improve the accuracy of the model. In addition, for the patient static information represented by the ICD-9 diagnostic code, the graph attention mechanism is used to reduce adjacent noise. To validate the results of data ablation, our experiment use static data, dynamic data, and graph attention information constructed using diagnostic code, which in turn contain sequences of diagnostic information and program code, as well as sequences of prescription information and event information.

3.3 Results and Discussions

Table 1. Summary statistics for the different algorithms used to predict readmission within 30 days of discharge from the intensive care unit.

	Average Precision	AUROC	F1	Sensitivity	Specificity
GROM	0.375 [0.366,0.384]	0.786 [0.366,0.384]	0.422 [0.416,0.427]	0.74 [0.734,0.746]	0.707 [0.7,0.713]
ODE + RNN + Attention	0.314 [0.306,0.321]	0.739 [0.736,0.741]	0.376 [0.371,0.381]	0.685 [0.666,0.704]	0.697 [0.681,0.711]
ODE + RNN	0.331 [0.323,0.339]	0.739 [0.737,0.742]	0.372 [0.367,0.377]	0.672 [0.659,0.686]	0.697 [0.683,0.711]
RNN	0.196 [0.19,0.203]	0.602 [0.599,0.605]	0.251 [0.248,0.254]	0.582 [0.561,0.603]	0.582 [0.561,0.603]
SVM	0.265 [0.256,0.274]	0.655 [0.651,0.658]	0.303 [0.297,0.309]	0.565 [0.552,0.577]	0.679 [0.668,0.691]

Overall Performance. Table 1 reports the average accuracy, AUROC, F_1-SCORE, sensitivity and specificity of deep learning architectures and support vector machines. GROM obtained the highest average accuracy of 0.375, the highest average AUROC of 0.786 and the highest average F1 score of 0.422. In general, the prediction accuracy of neural network was significantly higher

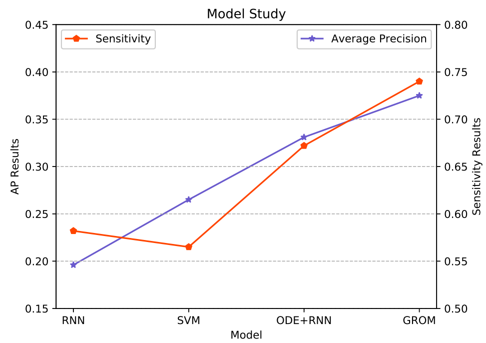

Fig. 3. We use SVM as a reference model to compare it with RNN, ODE+RNN and GROM, verifying that ODE module has processed dynamic data and graph attention mechanism reduces the RNN sensitivity of data noise.

than that of baseline models such as SVM. As shown in Fig 3, it can be seen that the sensitivity of the RNN was higher than that of SVM, and the effect of its other indicators was significantly lower than that of SVM. This result showed that the sensitivity of RNN in missing detection result was better than that of SVM. However, fitting issues had more false positive examples, causing data noise interference to prediction and further resulting in that the highest average accuracy of RNN was significantly lower than that of SVM. And we could deal with dynamic data through the ODE module and utilize graph attention mechanism to reduce noise, obtaining a good prediction improving effect.

The results of the ODE&RNN and RNN showed that the deep learning model RNN performed better in processing sequential data, but it lacked the ability to process irregular interval information. It is seen that the combination of the ODE module and the RNN achieved higher precision than the RNN model alone. Particularly, we noted that the ODE module had a significant increase in the accuracy of readmission prediction, with an average accuracy increase of 13.5% and an accuracy increase of nearly 69%. This allows us to believe that the ODE module is suitable for processing dynamic data in the model, making full use of valuable information for patient readmission prediction. Therefore, it can be concluded that the introduction of ODE based on RNN can take advantage of the modeling capability of irregular interval to better play the role of sequential data processing.

Secondly, GROM was compared with the ODE&RNN, and we can see that GROM achieved better precision than ODE&RNN model. Particularly, we noted that the GROM had the highest accuracy in readmission prediction, which gave us confidence in using graph knowledge to understand patient relationships in the absence of sufficient data. In addition, it is clear that the graph attention mechanism provided valuable information with embedding of CCA information in the prediction of patient readmission. Specifically, GROM improved the prediction

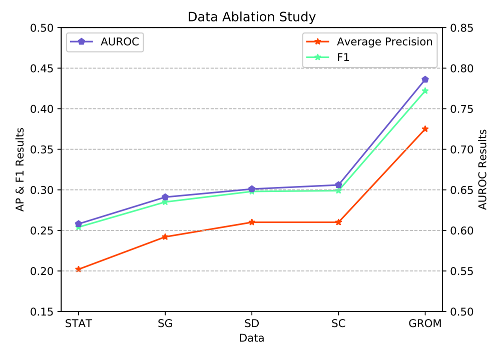

Fig. 4. We gradually add dynamic data and external prior information to static data to analyze the applicability of each part of the data in the model.

accuracy of readmission by 4.4%, indicating that the robust representation of the codes had learned significantly improved the prediction accuracy of readmission, and that the introduction of graphs contributed to data balance. It provided a more effective attention mechanism on existing information, and made better use of the value of information.

Finally, we compared GROM with ODE&RNN&Attention baseline model, the effect of adding an attention layer to the time series was poor, indicating that the use of knowledge graphs instead of the underlying attention mechanism was effective. The applicability of the time series attention structure in predicting reentry risk was poor, and the introduction of this layer in the model may lead to the performance decline of the model. The goal of this model was to extract knowledge from a given knowledge graph using attention mechanisms rather than adding attention mechanisms to past visits. Models with a graph attention layer (average precision range : 0.366-0.384) were slightly better than models based on a time attention layer (average precision range : 0.306-0.321). Instead of directly using the final memory state of RNN, a graph attention layer was applied to the output of RNN at each time step, increasing the association between records and improving prediction performance through data balance. This verifies the validity of the theory that the proposed graph attention mechanism can enhance the model prediction performance by enhancing the correlation between data.

Data Ablation Study. Fig 4 shows that based on static data, we used the attention layer building by the diagnostic codes associated by external CCS prior knowledge, with the prediction precision increased by 4%, and the usage of dynamic data set could effectively increase the precision by 5.8%. Moreover, combined with the static data and dynamic data, and introduced to an external priori knowledge, a complete GROM model could effectively enhance prediction precision, F1-score and AUROC, increasing the precision by about 86%, compared with a model merely using static data. Analysis of the experimental results showed static patient information alone was the worst predictive factor, because static information is only some of the static attributes of patients entering ICU and has a poor correlation with readmission. And patients' dynamic physiologic measurements had the greatest impact on the readmission prediction because these dynamic physiologic measurements are recorded in the patient's hospitalization, and changes in the patient's condition are closely related, which are valuable readmission prediction information. In addition, the patient's diagnostic information had a significant impact on the prediction task, as these diagnostic codes are given by the physician based on the patient's current condition and are highly relevant to the patient's readmission. Finally, the result shows that the introduction of external knowledge graph improved the prediction performance of the model, which makes us have reason to believe that external knowledge graph does contribute to the prediction of patient readmission.

Discussions. There are three major limitations in this study that could be addressed in future research. First, since all data in the MIMIC-III data set are shifted to protect patient confidentiality, it is not possible to ascertain which patients are admitted after 2001 and have at least 12 months of prior data, possibly leading to some incorrect values for the number of ICU admissions in the year preceding discharge. Second, information from clinical notes [8] is not included and the simplifying assumption is made that various diagnose and procedure-related codes are available immediately at the time of discharge. Third, the weak interpretability due to the continuous processing of patients' dynamic data by neural ODEs is difficult to analyze the influence of patient characteristics on the prediction effect.

4 Conclusion

In this paper, we proposed a based graph attention ordinary differential equation recurrent neural network (GROM) to predict readmission in ICU. The model framework was comprised of a recurrent neural network used to be the basic prediction model, adding neural ODE to process the irregular interval sequence information. Besides, our model also introduced a graph attention mechanism for using external knowledge to learn robust and reasonable representations of patient diagnostic codes to reduce noise interference between data. As demonstrated by experiment results, GROM produced better representations, which was validated by being used in a large open source MIMIC-III data set, effectively improving the prediction performance of readmission in ICU.

Acknoweledment

This work was supported in part by the Beijing Information Science and Technology University Research Level Improvement Project under grant no. 2020KYNH214, Qin Xin Talents Cultivation Program under grant no. QXTCPC202112, and Beijing Educational Science Planning Project of China under grant no. CHCA2020102.

References

1. Barbieri, S., Kemp, J., Perez-Concha, O., Kotwal, S., Jorm, L.: Benchmarking deep learning architectures for predicting readmission to the icu and describing patients-at-risk. Scientific Reports **10**(1) (2020)
2. Chen, R., Rubanova, Y., Bettencourt, J., Duvenaud, D.: Neural ordinary differential equations (2018)
3. Choi, E., Bahadori, M.T., Song, L., Stewart, W.F., Sun, J.: Gram: Graph-based attention model for healthcare representation learning. In: the 23rd ACM SIGKDD International Conference (2016)
4. Dupont, E., Doucet, A., Teh, Y.W.: Augmented neural odes (2019)
5. Garland, A., Olafson, K., Ramsey, C.D., Yogendran, M., Fransoo, R.: Epidemiology of critically ill patients in intensive care units: a population-based observational study. Critical care (London, England) **17**(5), R212 (2013)
6. Johnson, A., Pollard, T.J., Shen, L., Lehman, L.W.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: Mimic-iii, a freely accessible critical care database. Scientific Data
7. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. Computer Science (2014)
8. Liu, J., Zhang, Z., Razavian, N.: Deep ehr: Chronic disease prediction using medical notes (2018)
9. Ma, F., You, Q., Xiao, H., Chitta, R., Jing, G.: Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In: the 27th ACM International Conference (2018)
10. Peng, X., Long, G., Pan, S., Jiang, J., Niu, Z.: Attentive dual embedding for understanding medical concepts in electronic health records. In: 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2019)
11. Peng, X., Long, G., Shen, T., Wang, S., Jiang, J.: Self-attention enhanced patient journey understanding in healthcare system. arXiv preprint arXiv:2006.10516 (2020)
12. Peng, X., Long, G., Shen, T., Wang, S., Jiang, J., Blumenstein, M.: Temporal self-attention network for medical concept embedding. In: 2019 IEEE International Conference on Data Mining (ICDM). pp. 498–507. IEEE (2019)
13. Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M.: Scalable and accurate deep learning for electronic health records. npj Digital Medicine **1**(1), 18 (2018)
14. SLEE, Vergil, N.: The international classification of diseases: Ninth revision (icd-9). Annals of Internal Medicine **88**(3), 424–426 (1978)
15. Stearns, M.Q., Price, C., Spackman, K.A., Wang, A.Y.: Snomed clinical terms: overview of the development process and project status. Proceedings / AMIA ... Annual Symposium. AMIA Symposium p. 662 (2001)
16. Weiss, G.M., Mccarthy, K., Zabar, B.: Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs? In: International Conference on Data Mining (2007)
17. Xiao, C., Edward, C., Sun, J.: Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Journal of the American Medical Informatics Association (10), 10 (2018)
18. Xue, Y., Klabjan, D., Luo, Y.: Predicting icu readmission using grouped physiological and medication trends. Artificial intelligence in medicine **95**(APR.), 27–37 (2019)