Clinical Phenotyping Prediction via Auxiliary Task Selection and Adaptive Shared-Space Correction

Xiao Yang[1], Ning Liu[1(✉)], Jianbo Qiao[1], Haitao Yuan[2], Teng Ma[3], Yonghui Xu[1], and Lizhen Cui[1]

[1] Shandong University, Jinan, Shandong 250100, China
liun21cs@sdu.edu.cn, clz@sdu.edu.cn
[2] Beijing University of Posts and Telecommunications, Beijing 100091, China
[3] Tsinghua University, Beijing 100084, China

Abstract. Clinical Phenotyping is a fundamental task in clinical services, which assessments whether a patient suffers a medical condition of interest. Existing works focus on learning better patients' representations. Recently, multi-task learning has been proposed to transfer knowledge from different tasks and achieved promising performance. However, the existing multi-task models still suffer from the serious negative transfer and slow convergence problem when multiple phenotype tasks are trained together. Meanwhile, phenotype relatedness is ignored, limiting to boost the performance of the multi-task learning for the phenotype prediction. To address these issues, we propose a private-shared multi-task framework with auxiliary task selection and adaptive shared-space correction for phenotype prediction (MTL_AC). To start with, we design an auxiliary task selection method to find the most compatible phenotype task against one task by using phenotype relatedness. And then, a novel adaptive shared-space correction mechanism is proposed to address the negative transfer and slow convergence problem when two tasks are jointly trained under the private-shared multitask framework. The experimental results show that the proposed method performs better on various phenotype prediction tasks.

Keywords: Clinical phenotyping · Multi-task learning · Negative transfer · Auxiliary task

1 Introduction

Clinical Phenotyping is a basic clinical process, which aims to figure out whether a patient suffers from a medical condition of interest and is commonly used as the first step to facilitate multiple medical services [19]. For example, medical experts can group the patients via predefined phenotypes for precision medicine [20]. Actually, physicians should classify the patient's phenotype via complex health records, which requires a high level of clinical experience and knowledge.

L. Fang et al. (Eds.): CICAI 2022, LNAI 13605, pp. 438–449, 2022.
https://doi.org/10.1007/978-3-031-20500-2_36

With the rising complexity of patients' data collected, clinical phenotyping can be a challenging task for medical experts [24,26].

Recently, the adoption of electronic health records has collected quantities of health-related data, which offers opportunities for designing data-driven methods for automatic phenotype prediction [3,14,24].

Many works have utilized various deep learning technologies to model different aspects of the patients' data, including the recurrent neural networks [2,4,29], attention mechanism [17,22,30], graph-based methods [6,15], et al. These works are mainly focusing on learning the better representations of patients from multiple aspects. Inspired by the human learning activities, multi-task learning is designed to transfer knowledge among the tasks to improve the performance and has achieved promising performance in the clinical domain compared to training only for one task [9,21]. In spite of the success of multi-task learning in the clinical domain, they have certain limitations when performing the clinical phenotype prediction. Firstly, existing work has shown that multi-task learning may degrade the performance of some relevant tasks [7,28]. The existing multi-task models still suffer from the serious negative transfer problem caused by the multi-times updating of the shared space such that task-specific bias is introduced. Secondly, the relationship among different phenotypes is ignored, limiting to boost the performance of the multi-task learning for the phenotype prediction. In the clinical setting, one patient may suffer from different phenotypes, which may provide additional information on the phenotype relatedness.

In this paper, to boost the performance of clinical phenotype prediction, we formulate the clinical phenotype prediction task as a multi-task learning problem where each phenotype task is referred to as an independent task and propose a private-shared multitask framework with auxiliary task selection and adaptive shared-space correction for phenotype prediction (MTL_AC). Firstly, to distinguish the task-specific and task-independent information, we adopt a private-shared multi-task framework with two types of representations learned. Secondly, we propose an auxiliary task selection method to select the most compatible task with the consideration of phenotype co-occurrence. Thirdly, to diminish the negative transfer of the shared space, we design a novel adaptive shared-space correction method to adaptive change the optimization direction of the shared space. We conduct a comprehensive evaluation of real phenotype prediction tasks. Experimental results show that the proposed method outperforms the previous methods for most of the phenotype prediction tasks. In summary, the major contributions are listed as follows:

- We formulate the phenotype prediction task under the multi-task learning formulation and propose a private-shared multi-task framework with auxiliary task selection and adaptive shared-space correction.
- We design an auxiliary task section method via the co-occurrence of multiple phenotypes to select the most compatible phenotype.

- We design an adaptive shared-space correction method under the private-shared multi-task learning framework to reduce the bias introduced by the multi-times updating of the shared space.
- We conduct experiments on real clinical phenotype prediction tasks and the experimental results show the advantages of our method.

2 Related Work

2.1 Deep Learning for Clinical Phenotyping

Recently, deep learning methods have been successfully applied to clinical phenotyping. Previous works focus on learning better representations by considering the different aspects of data in electronic health data. As the information of patients is recorded sequentially in the EHR, Recurrent Neural Networks (RNN) are widely used as a temporal encoder to model the sequential information [23,27]. For example, [12] proposes an explainable deep learning system for healthcare, using RNN and attention mechanism to help medical staff to interpret, thus building a newcomer in deep learning systems. [18] propose a temporal deep learning model that performs bidirectional representation learning on EHR sequences for phenotype prediction, and can handle heterogeneous data, achieving excellent results in the prediction of chronic diseases. [1] introduced a semi-supervised learning method into phenotype prediction, using binary Markov process and Gaussian process for modeling, effectively using unlabeled EHR data to achieve high-precision prediction. These methods only focus on a single clinical phenotype task and cannot maintain the original performance in the face of multiple phenotype task predictions.

2.2 Multi-task Learning

Multi-task learning [5] is an approach that combines multiple tasks for training, aiming to exploit potential correlations and common features between tasks to improve performance. [16] proposed an adversarial sharing-private model, which uses an adversarial generation method to ensure that the information learned in the shared space and private space does not converge. In the field of clinical phenotype prediction, there is also a related study that introduces auxiliary tasks in the multi-task field into phenotype prediction. [7] proposed a method to randomly select auxiliary tasks between clinical phenotype tasks, which improved the prediction effect. [13] proposed an auxiliary task extraction method based on feature similarity, which is significantly better than random sampling.

3 Methodology

As is illustrated in Fig. 1, our proposed method MTL_AC is composed of two major components. The Auxiliary Task Selection, described in Sect. 3.1, is aimed at constructing the relatedness matrix according to the co-occurrence of multiple

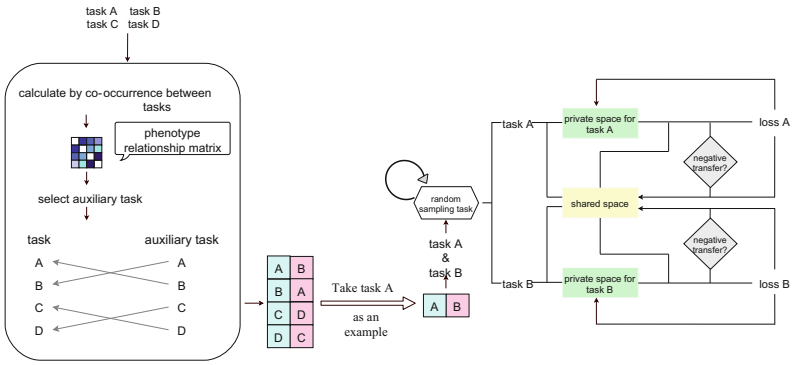

Fig. 1. MTL_AC: Taking tasks A, B, C, and D as an example, all tasks are sent to the auxiliary task selection module, each task finds the most suitable auxiliary task and then forms a task group, each group of tasks will be sent to the private-shared framework and an adaptive shared space correction mechanism is used to prevent negative transfer between tasks in task groups.

predefined phenotypes and finding the most compatible auxiliary task for each phenotype. And the Adaptive Shared-Space Correction, described in Sect. 3.2 is proposed to the re-correct bias of shared space under the private-shared multi-task framework.

3.1 Auxiliary Task Selection

Auxiliary task selection aims to find the best compatible phenotype for each phenotype. We argue that the comorbidity of phenotype plays an essential role when analyzing the patients' status, indicating the co-occurrence of phenotype may be of great help to dig out the correct phenotype with the help of the most similar phenotype task. In this section, we provide an effective way of digging out the most related phenotype task based on the intrinsic co-occurrence feature of the patients' phenotype.

Initially, we define the relationship matrix among the phenotype tasks as follows:

$$C_{ij} = \frac{\mathcal{N}(i,j)}{\mathcal{N}(i)} \tag{1}$$

where $\mathcal{N}(i,j)$ is the number of patients which suffer both of the i^{th} and j^{th} phenotype.

In order to find the best auxiliary tasks, we need to try our best to eliminate the effects of other phenotypes. Therefore, a penalty factor is defined for each phenotype as follows:

$$P_i = \frac{\sum_j C_{ij} - 1}{N - 1} \tag{2}$$

Following Eq. 2, the penalised relationship matrix CO^p is defined as the follow:

$$COMatrix_i = C_i - P_i \tag{3}$$

The final symmetrical phenotype relationship matrix reflects the degree to which tasks are suitable as auxiliary tasks. After the construction of the symmetrical phenotype relationship matrix, the auxiliary task can be selected from the relationship matrix by various distance measurements.

3.2 Private-Shared Framework with Adaptive Shared Space Correction

In this section, we first illustrate the Private-Shared Framework for phenotype prediction, and then we describe the detailed procedure of the Adaptive Shared Space Correction to reduce the bias in the shared space.

Private-Shared Framework. Given the selected auxiliary task pairs, we adopt the Private-Shared Multi-task Framework and design two layers for task-specific and task-independent feature learning.

Specifically, given the pair-wise phenotype dataset $D = \{\{D_k^i, y_k^i\}_{k=1}^{k=M}, \{D_k^j, y_k^j\}_{k=1}^{k=N}\}$, where y_k^i denotes the ground-truth label of each patient for a phenotype and $D_k^j \in R^{(T \times D)}$ is the feature matrix of k^{th} patient for the j^{th} phenotype with D features and T time-slots collected.

To get the representation of a patient, an encoder (e,g, LSTM, etc.) is utilized to transformer the patient's features into the embedding space, which can be obtained by:

$$x_i^k = Encoder(D_i^k; \theta) \tag{4}$$

And then, we use two different linear layers parameterized by θ^s and θ^p along with the no-linear activation functions (act) to transformer the embedding of the patient into the task-shared and task-specific space. Noted that, θ^s is shared among the tasks while θ^p is task-specific parameters. Formally, the task-specific vector $P(x_i^k)$ and the task-shared vector $S(x_i^k)$ can be computed by:

$$P(x_i^k) = f(x_i^k; \theta^p, act) \tag{5}$$

$$S(x_i^k) = f(x_i^k; \theta^s, act) \tag{6}$$

where $f(a; b, act) = act(a \cdot b)$ is the transformation function.

After that, the final patient's embedding $E(x_i^k)$ is computed by the concatenation of the two vectors, which is defined as:

$$E(x_i^k) = Concat(P(x_i^k), S(x_i^k)) \tag{7}$$

For optimization of a specific phenotype task, Binary Cross Entropy loss is adopted as the loss function.

Algorithm 1. Adaptive Shared-Space Correction

Require: Tasks: (T_i, T_j), Train data set: $x_{train} = \{x^i_{train}, x^j_{train}\}$, Validation data set: $x_{val} = \{x^i_{val}, x^j_{val}\}$, Training Parameters: $\theta = \{\theta_i \mid \theta_i = (\theta^s, \theta^p_i), i \in T\}$
Ensure: Network parameters: θ

1: **while** Not Converge **do**
2: Random sampling task T_i, the left task is T_j
3: Random sample a batch x^i_{batch} from the training set x^i_{train} of T_i
4: Random sample a batch x^j_{batch} from the training set x^j_{train} of T_j
5: Calculate the $loss_{pre}$ of T_j using θ_j and x^j_{batch}
6: update θ_i using x^i_{batch}
7: Calculate the $loss_{after}$ of T_j using θ_j and x^j_{batch}
8: **if** $loss_{pre} + threshold < loss_{after}$ **then**
9: update Shared space parameters θ^s using $loss_{after}$
10: **end if**
11: **if** loss on x_{val} stop fall within limited steps **then**
12: break;
13: **end if**
14: **end while**

Adaptive Shared Space Correction. As is illustrated in Sect. 3.2, the learned features can be divided into two groups, namely shared features and task-private features. However, due to the implicit separation of private and shared features when adopting the multi-task framework for phenotype prediction, the information learned in the shared space may be biased when training a task, resulting in the performance degradation of other tasks.

For any task i and j under the shared-private multi-task framework, the update process of the shared part θ^s is computed by:

$$\theta^s = \theta^s - \alpha \frac{\partial L_i(\theta^s)}{\partial \theta^s} \tag{8}$$

$$\theta^s = \theta^s - \alpha \frac{\partial L_j(\theta^s)}{\partial \theta^s} \tag{9}$$

From Eq. 8, 9, the shared parameter θ^s is updated multi-times according to the loss of the selected task while ignoring the effectiveness of other tasks, which may inject the task-specific information into the shared space. In order to separate the task-specific information during the training, we design a novel correction mechanism, namely Shared Space Correction Mechanism, to re-optimize the shared space by utilizing the information of other tasks, which is described in Algorithm 1. The aim of the Shared Space Mechanism is to ensure the correction optimization direction of the shared space. Therefore, we calculate the loss of other tasks except for the training task as an indicator to measure that the shared space is optimized in the correct direction (Algorithm 1, Line 5). If the indicator doesn't perform well on other tasks, the shared space will be re-optimized (Algorithm 1, Line 8–10).

4 Experiment

In this section, we introduce the empirical results of our MTL_AC framework on different phenotype classification tasks. And we use the area under the ROC curve (AUC) and F-score as the evaluation metrics.

4.1 Experiment Setup

Dataset. We evaluate the effectiveness of our framework on the phenotype classification tasks and report the average performance. The data set comes from a subset of the MIMIC-III database, which is open for public clinical research and covers 42276 ICU hospitalization records [11]. Following [9], a total of 17 clinical variables are selected and 25 phenotype prediction tasks are constructed. We randomly divided these data sets into training sets, validation sets, and testing sets with the proportion of 70%, 15%, and 15% respectively. We select the data within 24 h of admission for prediction.

Comparison Methods. We categorise the comparison methods as the follows:

Basic Encoder. As the data collected are sequential and multidimensional, we adopt four types of commonly-used models to capture the temporal information for clinical phenotyping.

- **LSTM:** The approach is proposed by [10], which is the standard Long-ShortLSTM [10].
- **Bi-Attention:** The approach learns the forward and backward timing information in the patient's representation vectors and predicts the patient's disease by utilizing the attention mechanism [25].
- **T-LSTM:** The approach is proposed by [2], which handles irregular time intervals in Healthcare Field by adding time decay. We modify this model into a supervised learning model.
- **SAnD:** The approach is first proposed by [22], which employs the masked, self-attention mechanism.

Phenotype Prediction Framework. We adopt three different training schemes for the phenotype prediction, which are listed as the following:

- **Baseline:** The approach formulates the phenotype prediction task as a multi-label classification problem.
- **MTNN:** A multi-task framework designed for Electronic phenotyping task [7].
- **cFSGL:** A multi-task framework designed based on accelerated gradient method (AGM) [8].

Ablation Models. In order to figure out the effectiveness of Auxiliary Task Selection and Adaptive Shared Space Correction modules, we design the ablation models as the follows:

- **SP-MTL:** The basic Shared-Private Model illustrated in Sect. 3.2.
- **MTL_AC-G:** SP-MTL with the Auxiliary Task Selection illustrated in Sect. 3.1.
- **MTL_AC-C:** SP-MTL with the Adaptive Shared Space Correction illustrated in Algorithm 1.

4.2 Experimental Settings

For the parameters in the attention mechanism in the T-LSTM model, we randomly initialize them from a uniform distribution in $(-0.1, 0.1)$. For other parameters, we adopt the default initialization strategy. And the models are trained with backpropagation using Adam optimizer. The detailed settings of hyper-parameters are shown in Table 1.

Table 1. Settings of hyper-parameters

Hyper-parameters 1	Settings
Initial learning rate	3e−4
Batch Size	32
Number of Early Stop	10
Dropout	0.5
Embedding Size of EHR data	128

4.3 Analysis of Results

Table 2. Comparison of framework MTL_AC and framework cFSGL and MTNN on basic encoders

Model	Baseline		cFSGL		MTNN		SP-MTL		MTL_AC	
	Auc	F1	Auc	F1	Auc	F1	Auc	F1	Auc	F1
LSTM	0.7054	0.5001	0.7088	0.4988	0.7081	0.4989	0.6963	0.4868	**0.7188**	**0.5166**
Bi-Attention	0.7012	0.4888	0.7029	0.4885	0.7105	0.5056	0.6981	0.4984	**0.7219**	**0.5270**
T-LSTM	0.7114	0.4978	0.7134	0.4987	0.7156	0.5049	0.7059	0.4924	**0.7264**	**0.5264**
SAnD	0.6643	0.5131	0.6796	0.5217	0.6853	0.5349	0.6681	0.5307	**0.6971**	**0.5464**

Effectiveness of the Proposed MTL_AC. Table 2 shows the prediction results of MTL_AC based on MTNN. Firstly, The predictive effect of treating phenotype as an independent task is generally better than treating phenotype as a multi-label task. There is even a situation where MTNNs prediction effect exceeds that of cFSGL. Our framework MTL_AC is based on MTNN, the accuracy has increased by 0.011 on average, and the F1 score has increased by 0.018

on average. And achieved a comprehensive and stable improvement on all models. When we use the SP-MTL model, compared with the single-task MTNN, the accuracy is reduced by 0.0128 on average, and the F1 score is reduced by 0.009 on average. It can be found that compared with the single-task, the multi-task model SP-MTL has a very serious negative transfer phenomenon.

We used T-LSTM, which achieved the best results among the four models, to show the improvement in all 25 clinical phenotype tasks, as shown in Fig. 2. The framework MTL_AC has achieved improvements in 22 phenotypes. Compared with framework cFSGL, which has only achieved improvements in 13 phenotypes, our framework has achieved very significant improvements, especially for chronic diseases. This helps to improve the difficulty of chronic disease prediction in clinical phenotype tasks.

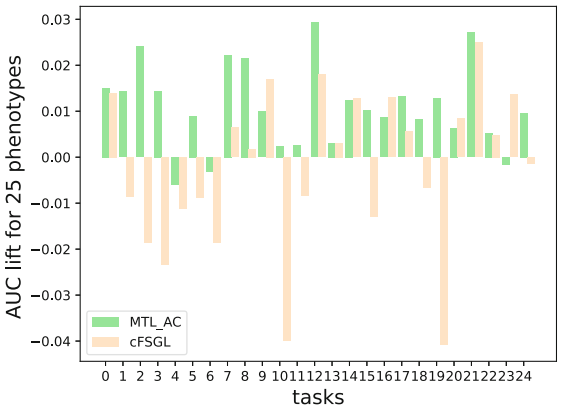

Fig. 2. Comparison of framework cFSGL and our framework MTL_AC

Table 3. Ablation of the proposed MTL_AC

Model	MTL_AC-G		MTL_AC-C	
	Auc	F1	Auc	F1
LSTM	0.7137	0.5131	0.6985	0.4855
Bi-Attention	0.7131	0.5117	0.7016	0.4993
T-LSTM	0.7224	0.5204	0.7129	0.4986
SAnD	0.6910	0.5430	0.6768	0.5275

Ablation Study. Our method is divided into finding one-to-one auxiliary tasks and a correction mechanism. We separately count the effects of each method. As shown in Table 3, after selecting auxiliary tasks (MTL_AC-G), both AUC and F1 scores are higher than MTNN. Only using Algorithm 1, Compared with SP-MTL, the accuracy is increased by 0.0535 on average, and the F1 score is

increased by 0.00065 on average. The improvement effect is not as good as the one-to-one auxiliary task, and the effect is still worse than the single-task MTNN.

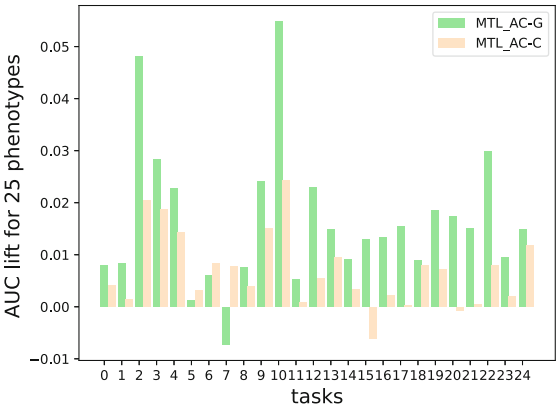

Fig. 3. Improvement of MTL_AC-G and MTL_AC-C over SP-MTL

Figure 3 shows the improvement of the two algorithms compared to the basic SP-MTL model. It can be seen that the improvement of the two algorithms for the basic multi-task model is very obvious.

5 Conclusion

In this paper, We propose the MTL_AC framework for clinical phenotypic task prediction. It exploits the co-occurrence between tasks to find the best one-to-one auxiliary phenotype task for each phenotype task, and further designs a self-correcting mechanism to prevent the negative transfer of tasks in the task-auxiliary task. Four models were tested on the MIMIC-III dataset and compared with another multi-task framework. The experimental results prove that our frameworks have produced better results. We further explore other ways of combining phenotypic tasks with other clinical tasks to gain more potential information.

Acknowledgement. This work is supported by the Fundamental Research Funds of Shandong University and partially supported by the NSFC (No. 91846205) the National Key R&D Program of China (No. 2021YFF0900800), the major Science and Technology Innovation of Shandong Province grant (No. 2021CXGC010108).

References

1. Ahuja, Y., Hong, C., Xia, Z., Cai, T.: Samgep: a novel method for prediction of phenotype event times using the electronic health record. medRxiv (2021)
2. Baytas, I.M., Xiao, C., Zhang, X., Wang, F., Jain, A.K., Zhou, J.: Patient subtyping via time-aware lstm networks. In: SIGKDD, pp. 65–74 (2017)
3. Birkhead, G.S., Klompas, M., Shah, N.R.: Uses of electronic health records for public health surveillance to advance public health. Annu. Rev. Public Health **36**, 345–359 (2015)
4. Cao, Y., et al.: Kdtnet: medical image report generation via knowledge-driven transformer. In: DASFAA, p. 117–132 (2022)
5. Caruana, R.: Multitask learning. Mach. Learn. **28**(1), 41–75 (1997)
6. Choi, E., Bahadori, M.T., Song, L., Stewart, W.F., Sun, J.: Gram: graph-based attention model for healthcare representation learning. In: SIGKDD, pp. 787–795 (2017)
7. Ding, D.Y., Simpson, C., Pfohl, S., Kale, D.C., Jung, K., Shah, N.H.: The effectiveness of multitask learning for phenotyping with electronic health records data. In: BIOCOMPUTING 2019: Proceedings of the Pacific Symposium, pp. 18–29 (2018)
8. Emrani, S., McGuirk, A., Xiao, W.: Prognosis and diagnosis of parkinson's disease using multi-task learning. In: SIGKDD, pp. 1457–1466 (2017)
9. Harutyunyan, H., Khachatrian, H., Kale, D.C., Ver Steeg, G., Galstyan, A.: Multitask learning and benchmarking with clinical time series data. Sci. Data **6**(1), 96 (2019)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
11. Johnson, A.E., et al.: Mimic-iii, a freely accessible critical care database. Sci. data **3**(1), 1–9 (2016)
12. Khedkar, S., Gandhi, P., Shinde, G., Subramanian, V.: Deep learning and explainable ai in healthcare using ehr. In: Deep Learning Techniques for Biomedical and Health Informatics, pp. 129–148 (2020)
13. Kung, P.N., Yin, S.S., Chen, Y.C., Yang, T.H., Chen, Y.N.: Efficient multi-task auxiliary learning: selecting auxiliary data by feature similarity. In: EMNLP, pp. 416–428 (2021)
14. Liu, N., Lu, P., Zhang, W., Wang, J.: Knowledge-aware deep dual networks for text-based mortality prediction. In: ICDE, pp. 1406–1417 (2019)
15. Liu, N., Zhang, W., Li, X., Yuan, H., Wang, J.: Coupled graph convolutional neural networks for text-oriented clinical diagnosis inference. In: DASFAA, pp. 369–385 (2020)
16. Liu, P., Qiu, X., Huang, X.: Adversarial multi-task learning for text classification. arXiv preprint arXiv:1704.05742 (2017)
17. Ma, F., Chitta, R., Zhou, J., You, Q., Sun, T., Gao, J.: Dipole: diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In: SIGKDD, pp. 1903–1911 (2017)
18. Meng, Y., Speier, W.F., Ong, M.K., Arnold, C.: Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. IEEE J. Biomed. Health Inform. (2021)
19. Oellrich, A., et al.: The digital revolution in phenotyping. Brief. Bioinform. **17**(5), 819–830 (2016)
20. Robinson, P.N.: Deep phenotyping for precision medicine. Hum. Mutat. **33**(5), 777–780 (2012)

21. Sadek, R.M., et al.: Parkinson's disease prediction using artificial neural network (2019)
22. Song, H., Rajan, D., Thiagarajan, J.J., Spanias, A.: Attend and diagnose: clinical time series analysis using attention models. In: AAAI (2018)
23. Wang, L., Zhang, W., He, X.: Continuous patient-centric sequence generation via sequentially coupled adversarial learning. In: DASFAA, pp. 36–52 (2019)
24. Wei, W.Q., Teixeira, P.L., Mo, H., Cronin, R.M., Warner, J.L., Denny, J.C.: Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J. Am. Med. Inform. Assoc. **23**(e1), e20–e27 (2016)
25. Yang, Y., Zheng, X., Ji, C.: Disease prediction model based on bilstm and attention mechanism. In: BIBM, pp. 1141–1148 (2019)
26. Yu, F., Cui, L., Cao, Y., Liu, N., Huang, W., Xu, Y.: Similarity-aware collaborative learning for patient outcome prediction. In: DASFAA, pp. 407–422 (2022)
27. Yu, X., Li, G., Chai, C., Tang, N.: Reinforcement learning with tree-lstm for join order selection. In: ICDE (2020)
28. Yuan, H., Li, G.: Distributed in-memory trajectory similarity search and join on road network. In: ICDE, pp. 1262–1273 (2019)
29. Yuan, H., Li, G., Bao, Z., Feng, L.: Effective travel time estimation: when historical trajectories over road networks matter. In: SIGMOD, pp. 2135–2149 (2020)
30. Yuan, H., Li, G., Bao, Z., Feng, L.: An effective joint prediction model for travel demands and traffic flows. In: ICDE (2021)