Journal of Biomedical Informatics 127 (2022) 104011

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Multi-channel fusion LSTM for medical event prediction using EHRs

Sicen Liu [a], Xiaolong Wang [a], Yang Xiang [b], Hui Xu [c], Hui Wang [c], Buzhou Tang [a,b,*]

[a] Department of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China
[b] Peng Cheng Laboratory, Shenzhen, China
[c] Gennlife (Beijing) Technology Co Ltd, Beijing, China

ARTICLE INFO

Keywords:
Medical event prediction
Electronic health records
Heterogeneous data
Multi-channel fusion
Long short-term memory

ABSTRACT

Automatic medical event prediction (MEP), e.g. diagnosis prediction, medication prediction, using electronic health records (EHRs) is a popular research direction in health informatics. In many cases, MEP relies on the determinations from different types of medical events, which demonstrates the heterogeneous nature of EHRs. However, most existing methods for MEP fail to distinguishingly model the type of event that is highly associated with the prediction task, i.e. task-wise event, which usually plays a more significant role than other events. In this paper, we proposed a Long Short-Term Memory network (LSTM)-based method for MEP, named Multi-Channel Fusion LSTM (MCF-LSTM), which models the correlations between different types of medical events using multiple network channels. To this end, we designed a task-wise fusion module, in which a gated network is applied to select how much information can be transferred between events. Furthermore, the irregular temporal interval between adjacent medical visits is also modeled in an individual channel, which is combined with other events in a unified manner. We compared MCF-LSTM with state-of-the-art methods on four MEP tasks on two public datasets: MIMIC-III and eICU. Experimental results show that MCF-LSTM achieves promising results on AUC(receiver operating characteristic curve), AUPR (area under the precision-recall curve), and top-k recall, and outperforms other methods with high stability.

1. Introduction

With the wide use of electronic health records (EHRs) systems in the past decades, massive EHRs data associated with patients, including demographic information, diagnoses, laboratory tests (labtests), prescriptions, radiological images, clinical notes, and so on, have been collected [1–5]. As a result, EHRs provide abundant materials to the development of advanced machine learning methods, which can be used to assist in healthcare from many aspects of routine medical practice by, e.g., making accurate medical event predictions (MEP), reducing medical errors, and facilitating more coordinated care [1,2,6,7]. However, making medical event predictions on certain tasks is still challenging, even for experienced physicians or relevant practitioners. Some medical events, e.g. diseases, are very difficult to predict or diagnose even given a patient's full history. For example, the diagnosis of some diseases may depend on the interaction of multiple factors and sometimes it is hard to identify the changes of conditions from plenty of dependent variables manually. As stated in [8], the diagnosis of hypogonadism can be easily incorrect or missed because of the interactions of complex interacted variables and the subtle changes they may cause. Although the MEP task is challenged, several studies have proven that predictive models outperform humans [9,10]. It allows us to believe that with the advances of artificial intelligence techniques, the predictive model of MEP would become better soon, and there would be an increasing number o scenarios under which computer-aided predictive models can assist medical practitioners such as physicians [11]. In the past decades, traditional machine learning methods such as logistic regression, support vector machines (SVM), and random forests have been broadly applied to various MEP tasks, including medical incident prediction [5,6], health risk prediction [13–16], unplanned readmission prediction [16–19], length of stay prediction [16,18], mortality prediction [13,14], discharge diagnosis prediction [21–26], prescription prediction [21,22], etc., relying upon EHRs [1,13,23,24]. However, although different types of medical events in EHRs are derived from heterogeneous sources, e.g. drugs are used in the therapeutic and diagnosis codes are made by the coding practitioners, most previous machine learning methods treat them equally and are insuffient in modeling the type information [31,32].

* Corresponding author at: Department of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, China.
E-mail address: tangbuzhou@hit.edu.cn (B. Tang).

https://doi.org/10.1016/j.jbi.2022.104011
Received 5 July 2021; Received in revised form 4 January 2022; Accepted 1 February 2022
Available online 15 February 2022

Recently, a growing number of deep learning methods have been applied to these tasks and have achieved promising performances [16,31–35]. The most representative deep learning-based method in modeling EHRs is recurrent neural network (RNN) [31], which is suitable for handling general time-series data, with a popular variant long short-term memory (LSTM) [32]. However, unlike general time-series data, EHRs are usually heterogeneous, i.e. heterogeneity, and with irregular time intervals, i.e. irregular temporality, as shown in Fig. 1, which cannot be well modeled by the basic LSTM. Concretely, the heterogeneity of EHRs refers to that different types of medical events can be involved, e.g. drugs, diagnoses, and procedures, which might signify different medical processes [36]. And irregular temporality refers to that the time intervals between adjacent visits are usually irregular, which could decay the impact of certain medical events [32,37]. To deal with the specific characteristics of EHRs, some LSTM variants have been proposed for MEP which considers either heterogeneity or irregular temporality. For example, T-LSTM [32], a time-aware LSTM considering the effect of irregular time intervals on information transmission, and DeLSTM [27], a variant LSTM considering temporality correlations between two types of medical events. Even so, studies that distinguishingly model the type of event that is highly associated with the prediction task, i.e. task-wise event, are quite limited. The task-wise events, however, usually play a more significant role than other events.

In this paper, we proposed a novel task-wise method based on LSTM to model correlations among different types of medical events and the irregular temporal interval between adjacent medical visits for MEP in a unified framework, named Multi-Channel Fusion LSTM (MCF-LSTM). In MCF-LSTM, the channel for the type of medical event that needs to be predicted (i.e., task-wise channel) is taken as the primary channel, while the other channels are as auxiliary channels. Further, the irregular temporal interval is taken as an additional type of medical event and modeled using an additional information channel. In order to model correlations between the primary channel and auxiliary channels, a gated task-wise fusion module is introduced over the basic structure of LSTM to select how much information from each auxiliary channel is transferred into the primary channel.

The contributions of this paper can be summarized as follows:

1) We jointly modeled heterogeneity and irregular temporality of EHRs by regarding the irregular temporal information as a special type of medical event and proposed a task-wise fusion module to integrate multiple types of medical events.
2) We updated the basic structure of LSTM with the proposed fusion module to form MCF-LSTM.
3) We conducted experiments on two public datasets and results show that MCF-LSTM is effective on four medical event prediction tasks and is superior to other state-of-the-art methods.

2. Related work

In this section, we first briefly introduced the relevant background of the deep learning method in healthcare, and then sketched some state-of-the-art deep learning models for medical event prediction.

2.1. Deep learning for healthcare

The growth of EHRs is at a lightning-fast speed in the era of big data, it is a promising research topic to assess the future risks of diseases and realize accurate predictions for individuals [3]. EHRs contain multivariate time series data (e.g., prescription, lab tests, produce, diagnosis) that records patients' visits to the hospital, which provide unprecedented opportunities for clinical states prediction [38]. Deep learning-based models have shown the capability on numerous tasks for healthcare such as readmission prediction [3,11,12,30], physiologic decompensation prediction [10,12], in-hospital mortality prediction [12,28], length of stay (LOS) prediction [12,31], phenotyping [12,26,32], prescription prediction [27,28], procedure prediction[42], lab-test prediction and disease risk prediction [8,27,28,36,37]. Due to the complexity of clinical records, given a patient cohort, complex temporal and heterogeneous medical sequences could be obtained, and effectively modeling these data urgently need to be addressed.

2.2. Medical event prediction

In recent years, various deep learning methods have been proposed for MEP on EHRs. Most studies have focused on the modeling of the temporality of EHRs. For example, DoctorAI is a representative method based on RNN [31], which takes the multi-hot vector of various types of medical events and the temporal interval since the last visit as input at each timestamp. T-LSTM, a time-aware LSTM proposed by Baytas et al. [32], applies a subspace decomposition to the cell memory of LSTM to discount the memory content according to irregular time intervals. RetainVis is also an RNN-based method with attention, it incorporates visit temporal intervals as additional features into the input of RNN and uses an attention mechanism to predict future states of patients [44]. StageNet is a stage-level LSTM that considers temporal intervals in the basic LSTM cell and adopts CNN to capture correlations between neighbor stages [14]. HiTANet is a hierarchical time-aware Transformer-based method using hierarchical attention to model temporal information at local and global levels [37]. However, these methods

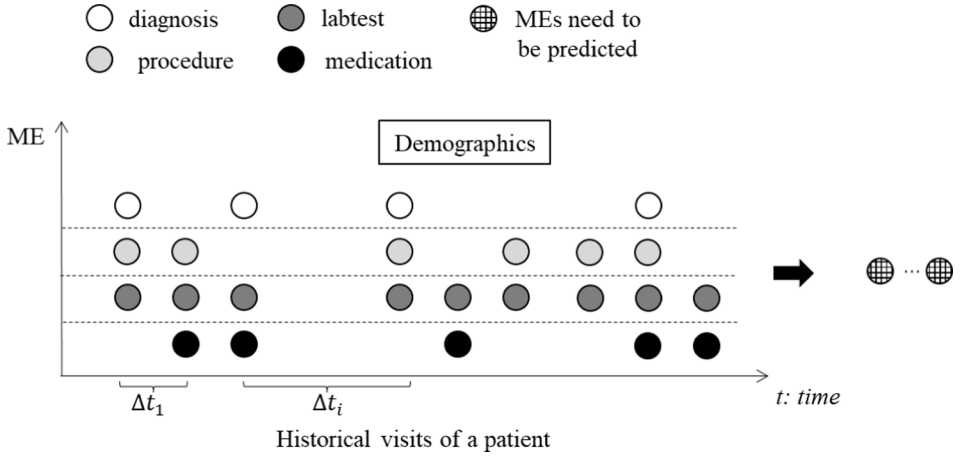

Fig. 1. Medical Event Prediction based on Heterogeneous and Irregular EMRs.

mainly focus on the processing of temporal information and do not distinguish among different types of medical events.

Some studies involve the modeling of heterogeneity but they mostly consider limited types of events or models in very simple ways. For example, derived by T-LSTM, Jin et al. [27] proposed a decomposed heterogeneous LSTM (DeLSTM) to model correlations between two types of medical events in each visit for improvement. In DeLSTM, labtests are regarded as auxiliary information of medications and are integrated into medications for the next prescription prediction. The limitation of DeLSTM is that it is not time-aware, and cannot model MEP tasks with multiple types of medical events. RGNN-TG-ATT is an extension of DeLSTM by introducing an additional time-aware graph neural network that links medical events in neighbor visits with edges weighted by temporal intervals [28]. Although RGNN-TG-ATT considers not only the irregular temporal characteristic of EHRs but also correlations among two types of medical events, it still suffers the limitation of DeLSTM that can not explore multiple types of medical events. Choi et al. took the multilevel structure of EHRs into account and learned multilevel embeddings from the hierarchical structure of EHRs [30]. However, there are obvious flaws of this method, e.g. not all EHRs have a satisfied hierarchical structure. None of the methods mentioned above are conducive to scaling to the scenarios of multiple types of medical events.

To summarize, methods that effectively model heterogeneity of EHRs are still insufficient. Intuitively, the type of event that is highly associated with the prediction task, e.g. the drug event in the drug recommendation task, would play a more significant role in modeling. Therefore, in this study, we proposed a method based on LSTM for MEP to emphasize the impact of the task-wise event.

3. Method

In this section, we first introduce the notations to represent medical event prediction with temporal EHRs data, and then MCF-LSTM in detail.

3.1. Medical event prediction with temporal EHR data

For a patient $p_i \in P(i = 1, \cdots, n)$ with demographic information S_i and T visits $(v_{i1}, \cdots, v_{it}, \cdots, v_{iT})$, each visit v_{it} may contain multiple diagnoses v_{it}^d, procedures v_{it}^p, laboratory tests (labtests) v_{it}^l and medications v_{it}^m, where $v_{it}^d = (x_{d_{it}}^1, \cdots, x_{d_{it}}^d, \cdots, x_{d_{it}}^D)^T$, and x_{it}^d denotes whether the d-th diagnosis code appears at time t (1-yes and 0-no), $v_{it}^p = (x_{p_{it}}^1, \cdots, x_{p_{it}}^p, \cdots, x_{p_{it}}^P)^T$, $x_{p_{it}}^p \in X_P$ denotes whether the p-th procedure code appears at time t (1-yes and 0-no), $v_{it}^l = (x_{l_{it}}^1, \cdots, x_{l_{it}}^l, \cdots, x_{l_{it}}^L)^T$, $x_{l_{it}}^l \in X_L$ denotes whether the l-th labtest code appears at time t (1-yes and 0-no), and $v_{it}^m = (x_{m_{it}}^1, \cdots, x_{m_{it}}^m, \cdots, x_{m_{it}}^M)^T$, $x_{m_{it}}^m \in X_M$ denotes whether the m-th medication code appears at time t (1-yes and 0-no). The goal of the MEP task is to predict any of these types of events (i.e., diagnosis, procedure, labtest or medication) at the next visit of p_i, $v_{i(T+1)}$, given the previous T visit of p_i. Fig. 2 gives the formal description of the medical event prediction task (take medication prediction as an example).

In this paper, information from each type of medical event is regarded as an information channel, the channel with the same medical event type of the prediction task is denoted as the primary channel, and the other channels are denoted as auxiliary channels. For example, the medication channel is the primary channel and the other channels (e.g. diagnosis channel) are the auxiliary channels for the medication prediction task. In order to model the irregular temporal characteristic of EHRs data, we also regard the temporal information as a special auxiliary channel. In this way, the history information shown in Fig. 2 is converted into 5-channel EHR data.

3.2. Basic structure of multi-channel fusion LSTM (MCF-LSTM)

The key idea of MCF-LSTM is to make medical event prediction via multi-channel fusion in the LSTM framework, which is centered on the primary channel and projects all auxiliary channels to the primary channel via a task-wise fusion module based on a gated layer. The fusion module can capture correlations between the primary channel and the auxiliary channels effectively. We propose the following two ways for channel fusion. For convenience to present MCF-LSTM, we denote the primary channel of the t-th visit as v_t^*, and the auxiliary channels of t-th visit as $v_t^1, \cdots, v_t^k, \cdots, v_t^{cl}$, where cl is the number of auxiliary channels. Take the medication prediction task as an example, as shown in Fig. 2, v_{it}^m is the primary channel of the t-th visit corresponding to v_t^*, and v_{it}^d, v_{it}^p, v_{it}^l as well as $v_{it}^{\Delta t}$ are auxiliary channels of the t-th visit corresponding to $v_t^1, \cdots, v_t^k, \cdots, v_t^{cl}(cl = 3$ (medical events) $+ 1$ (temporal interval) $= 4)$, where $v_{it}^{\Delta t}$ is the channel corresponding to the irregular temporal interval between visit t-1 and t, and Δt denotes the channel for irregular temporal intervals.

1) **Single-Belt Fusion (SBF)** is a way to fuse channel information on a single belt (shown in Fig. 3), where the primary channel and all auxiliary channels share one LSTM unit for information fusion, and information from all auxiliary channels is projected to the primary channel using a *Fusion* module, each of auxiliary channel message passing through the primary channel belt. The formal description of the basic structure of SBF is:

$$c_{t-1}^k = Fusion\left(c_{t-1}, v_t^k\right) fork = 1, \ldots, cl \tag{1}$$

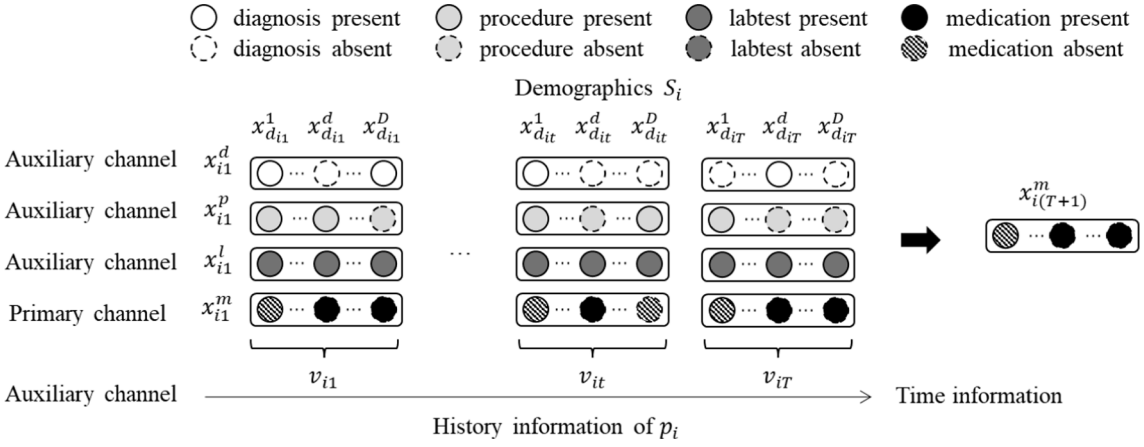

Fig. 2. Formal description of medical event prediction (e.g. medication prediction).

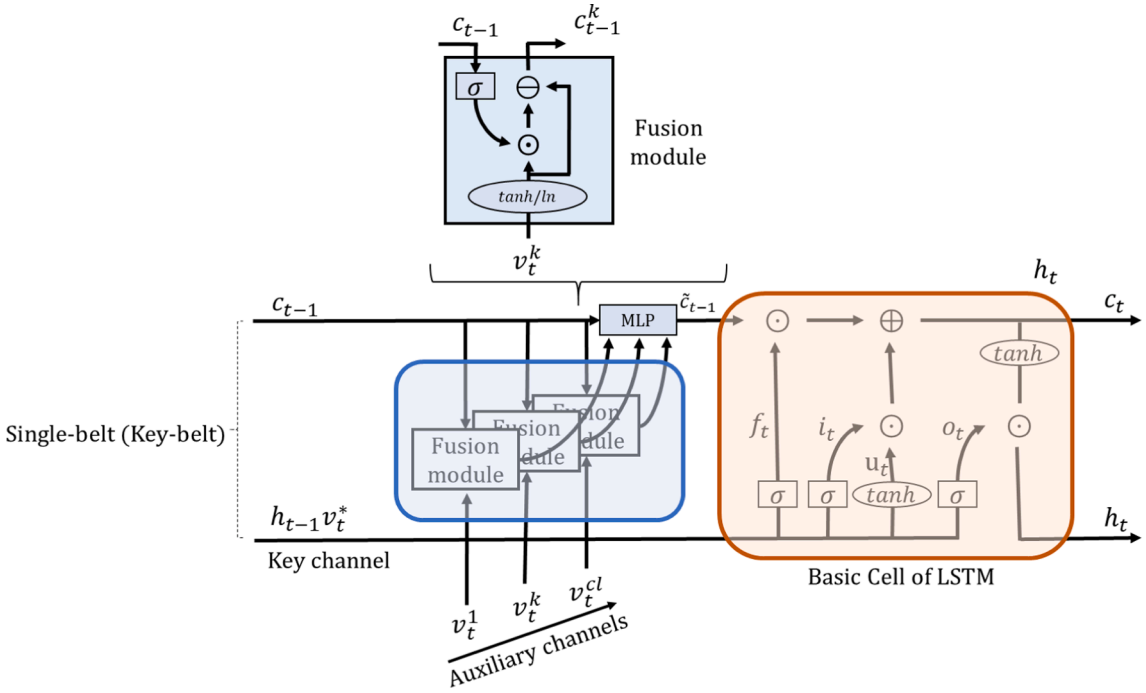

Fig. 3. Basic structure of Sigle-Belt Fusion.

$$c_{t-1} = MLP\left(\left[C_{t-1}^{1}, \cdots, C_{t-1}^{cl}, C_{t-1}\right]\right) \tag{2}$$

$$f_t = \sigma\left(W_f\left[v_t^*, h_{t-1}\right] + b_f\right) \tag{3}$$

$$i_t = \sigma\left(W_i\left[v_t^*, h_{t-1}\right] + b_i\right) \tag{4}$$

$$o_t = \sigma\left(W_o\left[v_t^*, h_{t-1}\right] + b_o\right) \tag{5}$$

$$c_t = f_t{*}c_{t-1} + i_t{*}\tanh\left(W_c\left[v_t^*, h_{t-1}\right] + b_c\right) \tag{6}$$

$$h_t = o_t{*}\tanh(c_t) \tag{7}$$

where *Fusion* is the task-wise fusion module based on a gated network as follows:

$$v_t^{*k} = \begin{cases} 1/ln\left(e + v_t^k\right) for\ temporal\ interval\ channel \\ tanh\left(W_k v_t^k + b_k\right) for\ other\ auxiliary\ channels \end{cases} \tag{8}$$

$$g_t^k = \sigma\left(W_g c_{t-1} + b_g\right) \tag{9}$$

$$c_{t-1}^k = \left(1 - g_t^k\right){*}v_t^{*k} \tag{10}$$

MLP is a multi-layer perceptron to convert the concatenation of c_{t-1}^1, $\cdots, c_{t-1}^{cl}$ and c_{t-1} into a vector of the same dimension as c_{t-1}, denoted by c_{t-1}. f_t, i_t and o_t are the forget gate, input gate, and output gate of the basic LSTM cell. W_f^k, W_i^k, W_o^k, W_c^k, W_k and W_g are weights of gates, b_f^k, b_i^k, b_o^k, b_c^k, b_k and b_g are corresponding biases.

In this SBF cell, each fusion module first selects information from an auxiliary channel according to the history information (Eq. (1)), the selected information from all auxiliary channels is then integrated into the history information (Eq. (2)), and the combined information from all channels is finally input into the LSTM cell (Eq. (3–7)). Each auxiliary channel is projected to the primary channel via the fusion module.

2) **Multi-Belt Fusion (MBF)** is a way to fuse channel information on multiple belts (shown in Fig. 4), where each auxiliary channel has its belt for information fusion. In MBF, all auxiliary channels are first individually fused with the primary channel, and then they are combined with the primary channel, each auxiliary channel possesses its message-passing belt. The *k*-th auxiliary channel is fused with the primary channel in the following way:

$$\widehat{c}_{t-1}^k = Fusion\left(c_{t-1}^k\right) \tag{11}$$

$$f_t^k = \sigma\left(W_f^k\left[v_t^*, h_{t-1}^k\right] + b_f^k\right) \tag{12}$$

$$i_t^k = \sigma\left(W_i^k\left[v_t^*, h_{t-1}^k\right] + b_i^k\right) \tag{13}$$

$$o_t^k = \sigma\left(W_o^k\left[v_t^*, h_{t-1}^k\right] + b_o^k\right) \tag{14}$$

$$c_t^k = f_t^k{*}\widehat{C}_{t-1}^k + i_t^k{*}\tanh\left(W_c^k\left[v_t^*, h_{t-1}^k\right] + b_c^k\right) \tag{15}$$

$$h_t^k = o_t^k{*}\tanh\left(c_t^k\right) \tag{16}$$

where W_f^k, W_i^k, W_o^k and W_c^k are weights of gates, b_f^k, b_i^k, b_o^k and b_c^k are corresponding biases.

After all auxiliary channels are fused with the primary channel, we combine them with the primary channel using a multi-layer perceptron:

$$h_t = MLP\left(\left[h_t^1, \cdots, h_t^k, \cdots h_t^{cl}, h_t^*\right]\right) \tag{17}$$

Compared with SBF, MBF first uses an individual belt for each auxiliary channel with the same fusion module to select information from the auxiliary channel according to the history information, and then it combines the information from all channels after the LSTM cell.

3.3. Medical event prediction via MCF-LSTM

Besides the information in different channels, patient demographic information is also important for medical event prediction. Following previous studies [27,28], we incorporate the demographic information (such as gender, age, ethnicity, and region, and we just use that information of patient in the first admission) S into h_T (the hidden state at the last time in MCF-LSTM) in the following way:

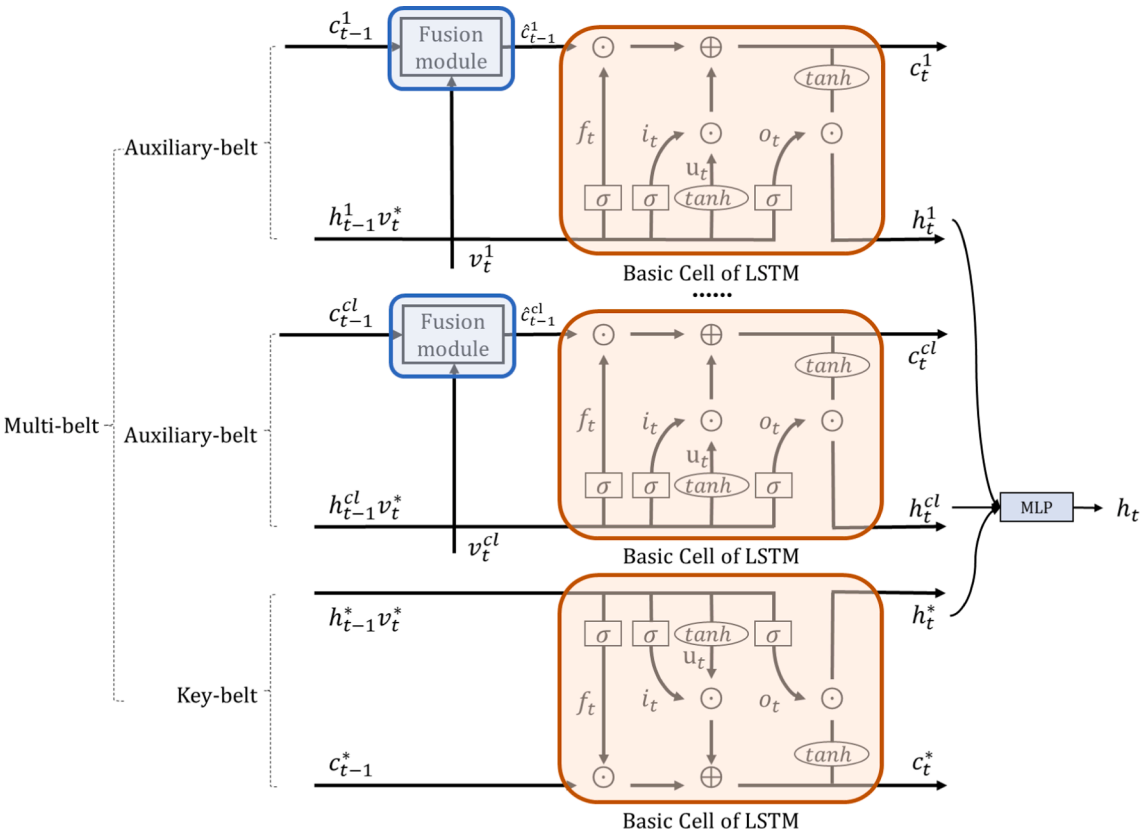

Fig. 4. The basic structure of Multi-Belt Fusion.

$$v_T^{'} = relu(MLP(W_{static}S + b_{static}) + h_T) \tag{18}$$

where W_{static} and b_{static} are trainable parameters.

Then, we use the sigmoid function to predict whether a medical event (e.g. medication code in the medication prediction task) appears at the next visit $T + 1$:

$$y^{'} = \sigma\left(W_y v_T^{'} \pm b_y\right) \tag{19}$$

where W_y and b_y are trainable parameters.

The cross-entropy loss function, take the medication prediction task as an example, is used as the objective function for parameter optimization:

$$L = -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m} y_{i,j}\log\left(y_{i,j}^{'}\right) + \left(1 - y_{i,j}\right)\log\left(1 - y_{i,j}^{'}\right) + \lambda\|W\|^2 \tag{20}$$

where n is the number of patients, m is the number of medication codes for prediction, $y_{i,j}$ and $y_{i,j}^{'}$ are the possibilities of the ground truth and predicted code of the j-th code at the next visit of patient i, λ is a hyperparameter to control the importance of the L_2 regularization of trainable parameters W.

4. Experiments

4.1. EHR datasets

We conducted experiments on the following two real-world datasets:

MIMIC-III is a publicly available dataset about intensive care unit (ICU) patients over 11 years. It consists of EHRs data of 53,423 patients collected at Beth Israel Deaconess Medical Center. Following previous studies [27,28], we removed patients with visits less than twice and only keep the medication codes with a frequency of more than 2000. We finally obtained a cohort of 5438 patients.

eICU the eICU collaborative research database, is a multi-center critical care dataset released by Philips in partnership with the MIT Lab for Computational Physiology. This dataset covers a total of 139,367 patients admitted to critical care units in 2014 and 2015 year. For the medical event prediction task, we chose medication, labtest, and diagnosis information for the experiment. We also removed patients whose visits are less than twice. Finally, we obtained a cohort of 9,215 patients. The statistics of these two datasets are listed in Table 1 in detail.

4.2. Compared methods

We compared MCF-LSTM with a simple rule-based method (called COPY) and the following state-of-the-art methods: **DoctorAI** [31], **T-LSTM** [32], **DeLSTM** [27], **RetainVis** [44], **RGNN-TG-ATT** [28], **StageNet** [14] **and HiTANet** [37]. For a patient, COPY directly uses the medical events in the last visit as the prediction for his/her next visit.

Table 1
statistics of the two datasets for experiments.

Dataset	MIMIC-III	eICU
#Patients	5438	9215
#Diagnosis codes	905	1366
#Procedure codes	529	NA
#Medication codes	346	1,375
#Labtest codes	676	155
Avg. (Max) #visits per patient	2.59(29)	2.19(9)
Avg. (Max) #diagnosis codes per visit	12.92(87)	8.22(106)
Avg. (Max) #procedure codes per visit	4.12(37)	NA
Avg. (Max) #medication codes per visit	32.00(172)	14.73(100)
Avg. (Max) #labtest codes per visit	71.26(252)	38.36(87)

where "#" denotes "the unique number of", "Avg. (Max) #" denotes "the averaged (maximum) number of", and "NA" denotes that the corresponding item does not exist.

4.3. Experimental settings

In this study, we randomly split each dataset into training, validation, and test sets with ratios of 8:1:1. We implemented MCF-LSTM and RGNN-TG-ATT by PyTorch and utilized the released source codes of DoctorAI[1], RetainVis[2], T-LSTM[3], DeLSTM, StageNet[4], and HiTANet[5]. Adam was used as the optimizer of MCF-LSTM and RGNN-TG-ATT. All models were trained on the training set using early stop with 100 epochs, and their parameters were optimized on the validation set. All hyper-parameters of different models were set as default except learning rate and the following ones, where the learning rate was optimized from 10^{-6} to 10^{-3} with step size of 10 times:

- RGNN-TG-ATT/RETAIN/T-LSTM. The number of the hidden units of the LSTM cell was set to 256.
- DoctorAI. The number of the hidden units of the GRU cell was set to 200.
- DeLSTM. The number of the hidden units of the decomposed LSTM cell was set to 256.
- StageNet. The length of the observation window K was set to 10. The number of the hidden units of the LSTM cell was set to 384. The chunk size factor C was set to 128.
- MCF-LSTM. The number of the hidden units of LSTM was set to 512.

We ran the other weights in nets of our experiments with randomly initialized weights 10 times and report their averaged performances.

4.4. Evaluation metrics

Following existing works [14], [27], [28], [31], [32], [37] and [44], we adopted the area under the receiver operating characteristic curve (AUC), the area under the precision-recall curve (AUPR) and top-k recall as model evaluation metrics.

5. Results

The comparison results of MCF-LSTM with other methods on AUC and AUPR values are reported in Table 2, where the highest values in each column are highlighted in bold. We can observe the following conclusions according to the results.

Firstly, the methods that pay special attention to irregular intervals, including RetainVis, T-LSTM, RGNN-TG-ATT, SatgeNet, HiTANet, MCF-LSTM-SBF, and MCF-LSTM-MBF, as well as the methods that pay special attention to heterogeneous data, including RetainVis, DoctorAI, DeLSTM, RGNN-TG-ATT, StageNet, HiTANet, MCF-LSTM-SBF, and MCF-LSTM-MBF, achieves much better performance than the rule-based method, i.e., COPY. This result likely demonstrates that all these deep learning methods are able to capture the complex relationships among heterogeneous and irregular medical time series data.

Secondly, DeLSTM considerably outperforms T-LSTM on medication prediction, indicating that the relations between labtests and medications perhaps are much stronger than that between the temporal information and medications.

Thirdly, HiTANet achieves better performance than other baselines in most cases because it considers both the heterogeneous characteristic and the irregular temporal characteristic of EHRs using Transformer with an attention mechanism. It concatenates heterogeneous data from different channels and models temporal dependencies in the sequence of patient visits via an attention mechanism.

Fourthly, our proposed methods, MCF-LSTM-SBF and MCF-LSTM-MBF, perform not only better but also more stable than all other methods in most cases. Compared with DeLSTM, MCF-LSTM learns dynamic temporal dependencies in the sequence of patient visits. Compared with T-LSTM, MCF-LSTM considers heterogeneous data. Compared with RetainVis, DoctorAI, RGNN-TG-ATT, StageNet, and HiTANet, MCF-LSTM takes full advantage of the correlations among information from different channels. In the case of the two proposed methods, MCF-LSTM-MBF is better than MCF-LSTM-SBF in most cases.

Finally, most deep learning methods perform well on all medical event prediction tasks except the medication prediction task on the MIMIC-III dataset. It indicates that 1) medication prediction is more dependent on the types of medical services than other medical event prediction tasks; 2) medication prediction in ICU might be easy than in other situations.

In order to compare our methods with other methods more comprehensively, we also show the top-k recalls (denoted by recall@k) of all methods on the medication prediction task in Table 3. We set k to 10, 20, 30, 40 and 50. We can see that MCF-LSTM outperforms other methods consistently and is more stable than other methods for all different k. In the case of the two proposed methods, MCF-LSTM-MBF achieves higher top-k recalls on both datasets.

To investigate the effect of auxiliary channels on the prediction of the primary channel, we conduct ablation study by removing auxiliary channels individually or all at once and comparing the variants with the original MCF-LSTM-MBF. Experimental results on the medication prediction task on MIMIC-III are shown in Table 4, where "w/o" denotes "without". From the results, we can see that the performance of MCF-LSTM-MBF decreases when any auxiliary channel is removed, indicating that any one of the four auxiliary channels considered in this study is complementary to the primary channel. MCF-LSTM-MBF degrades into basic LSTM with the worst performance when all auxiliary channels are removed. It indicates that the auxiliary channels are complementary to each other, and the task-wise fusion module can model correlations between the primary channel and auxiliary channels effectively. In addition, we also investigate another variant of MCF-LSTM-MBF that considers multiple auxiliary channels equally as one auxiliary channel, denoted by MCF-LSTM-MBF(ONE). As shown in Table 4, MCF-LSTM-MBF outperforms MCF-LSTM-MBF(ONE), indicating that the type information of medical events is important for medical event predication, which cannot be ignored.

6. Discussion

Gaining knowledge and effectively guiding the clinical decision from complex and heterogeneous EHRs remains a challenge in health care. However, the large availability of EHRs data brings tremendous opportunities to healthcare research. In this section, we analyzed the fusion network and explain the benefits of the MCF-LSTM we have proposed. A number of studies have attempted to use the heterogeneity or irregularity of EHRs to model the medical event prediction task. These studies utilize the sequence of EHRs built to predict by a time variable or by a limited type of medical event. Our goal is to make medical prediction models more universal in complex healthcare scenarios. As shown in Fig. 3 and Fig. 4, we explored the limited types of medical events to multiple types, and then treated the irregular temporal information as a special type of medical event. This simple but powerful operation has been demonstrated to be useful as well. Furthermore, this operation will make it easier to adapt our model to complex healthcare scenarios, since the larger number of medical events in prediction simply requires expanding an auxiliary channel in the model. Also, the experimental results above illustrate that MCF-LSTM is able to model the multiple types and irregular temporal characteristics of EHRs in a framework that makes effective use of MEP task, it demonstrates that the proposed task-wise fusion module provides an effective way to capture correlations among different types of medical events and temporal

[1] https://github.com/mp2893/doctorai.
[2] https://github.com/minjechoi/RetainVis.
[3] https://github.com/illidanlab/T-LSTM.
[4] https://github.com/v1xerunt/StageNet.
[5] https://github.com/HiTANet2020/HiTANet.

Table 2
AUC and AUPR of all methods on different medical event prediction tasks.

Dataset	Model	Medication prediction		Diagnosis prediction		Procedure prediction		Labtest prediction	
		AUC(%)	AUPR(%)	AUC(%)	AUPR(%)	AUC(%)	AUPR(%)	AUC(%)	AUPR(%)
MIMIC-III	COPY	66.46 ± 0.00	20.54 ± 0.00	70.93 ± 0.00	20.81 ± 0.00	64.68 ± 0.00	8.69 ± 0.00	84.65 ± 0.00	52.47 ± 0.00
	RetainVis	82.09 ± 0.13	33.69 ± 0.39	92.85 ± 0.06	28.22 ± 0.24	92.82 ± 0.26	22.48 ± 0.64	96.71 ± 0.03	83.47 ± 0.11
	DoctorAI	84.15 ± 0.16	35.69 ± 0.26	93.83 ± 0.02	25.37 ± 0.29	94.16 ± 0.07	24.23 ± 0.27	96.92 ± 0.03	82.36 ± 0.38
	T-LSTM	79.72 ± 0.65	29.77 ± 0.97	91.71 ± 0.41	24.89 ± 1.04	90.85 ± 0.76	19.13 ± 1.04	96.00 ± 0.22	81.33 ± 0.73
	DeLSTM	84.08 ± 0.17	36.70 ± 0.15	NA	NA	NA	NA	NA	NA
	RGNN-TG-ATT	85.06 ± 0.34	39.78 ± 0.61	NA	NA	NA	NA	NA	NA
	StageNet	83.93 ± 0.20	36.26 ± 0.37	94.03 ± 0.04	30.27 ± 0.41	94.05 ± 0.12	23.12 ± 1.14	97.20 ± 0.03	84.21 ± 0.15
	HiTANet	83.48 ± 0.13	39.21 ± 0.27	94.59 ± 0.04	38.91 ± 0.20	94.48 ± 0.07	28.48 ± 0.41	97.18 ± 0.02	85.00 ± 0.15
	MCF-LSTM-SBF	85.62 ± 0.02	39.40 ± 0.15	94.92 ± 0.02	39.04 ± 0.17	94.70 ± 0.06	27.72 ± 0.43	97.26 ± 0.02	85.23 ± 0.10
	MCF-LSTM-MBF	**86.06 ± 0.03**	**40.68 ± 0.22**	**95.01 ± 0.04**	**39.95 ± 0.27**	**94.97 ± 0.04**	**28.97 ± 0.34**	**97.36 ± 0.01**	**85.63 ± 0.11**
eICU	COPY	62.40 ± 0.00	11.28 ± 0.00	90.34 ± 0.00	57.16 ± 0.00	NA	NA	83.60 ± 0.00	68.30 ± 0.00
	RetainVis	96.65 ± 0.06	24.27 ± 0.31	93.48 ± 0.08	37.01 ± 1.18	NA	NA	95.54 ± 0.04	86.95 ± 0.11
	DoctorAI	92.94 ± 1.14	13.13 ± 2.02	93.68 ± 0.34	22.96 ± 2.96	NA	NA	95.16 ± 0.02	84.31 ± 0.22
	T-LSTM	89.29 ± 1.14	17.56 ± 1.06	92.78 ± 0.07	22.24 ± 1.41	NA	NA	94.80 ± 0.11	85.19 ± 0.28
	DeLSTM	95.47 ± 5.52	22.58 ± 6.17	NA	NA	NA	NA	NA	NA
	RGNN-TG-ATT	97.74 ± 0.04	27.43 ± 0.38	NA	NA	NA	NA	NA	NA
	StageNet	97.78 ± 0.07	31.13 ± 0.55	94.94 ± 0.14	42.17 ± 1.57	NA	NA	95.79 ± 0.05	87.80 ± 0.14
	HiTANet	97.32 ± 0.38	29.08 ± 0.80	96.15 ± 0.41	54.41 ± 2.75	NA	NA	96.01 ± 0.03	88.18 ± 0.13
	MCF-LSTM-SBF	97.94 ± 0.02	32.21 ± 0.20	96.07 ± 0.06	55.46 ± 0.50	NA	NA	96.15 ± 0.03	88.46 ± 0.15
	MCF-LSTM-MBF	**98.00 ± 0.01**	**32.14 ± 0.24**	**96.56 ± 0.04**	**59.26 ± 0.25**	NA	NA	**96.31 ± 0.02**	**88.72 ± 0.13**

Table 3
Top-*k* recall of all methods on the medication prediction task.

Dataset	Method	Recall@10	Recall@20	Recall@30	Recall@40	Recall@50
MIMIC-III	COPY	0.1680 ± 0.0000	0.2963 ± 0.0000	0.3742 ± 0.0000	0.4323 ± 0.0000	0.4756 ± 0.0000
	RetainVis	0.1829 ± 0.0027	0.3004 ± 0.0029	0.3912 ± 0.0028	0.4640 ± 0.0035	0.5248 ± 0.0027
	DoctorAI	0.1924 ± 0.0006	0.3110 ± 0.0018	0.4036 ± 0.0024	0.4803 ± 0.0021	0.5449 ± 0.0028
	T-LSTM	0.1654 ± 0.0044	0.2697 ± 0.0070	0.3523 ± 0.0079	0.4212 ± 0.0092	0.4812 ± 0.0089
	DeLSTM	0.1981 ± 0.0025	0.3162 ± 0.0029	0.4084 ± 0.0029	0.4842 ± 0.0025	0.5474 ± 0.0035
	RGNN-TG-ATT	0.1948 ± 0.0027	0.3184 ± 0.0028	0.4149 ± 0.0032	0.4933 ± 0.0030	0.5576 ± 0.0037
	StageNet	0.1963 ± 0.0024	0.3147 ± 0.0028	0.4038 ± 0.0025	0.4799 ± 0.0027	0.5421 ± 0.0037
	HiTANET	0.2051 ± 0.0014	0.3310 ± 0.0019	0.4264 ± 0.0025	0.5037 ± 0.0026	0.5680 ± 0.0025
	MCF-LSTM-SBF	0.2083 ± 0.0012	0.3342 ± 0.0021	0.4316 ± 0.0017	0.5098 ± 0.0019	0.5764 ± 0.0012
	MCF-LSTM-MBF	**0.2119 ± 0.0017**	**0.3407 ± 0.0020**	**0.4398 ± 0.0012**	**0.5199 ± 0.0016**	**0.5852 ± 0.0013**
eICU	COPY	0.1914 ± 0.0000	0.2503 ± 0.0000	0.2716 ± 0.0000	0.2821 ± 0.0000	0.2904 ± 0.0000
	RetainVis	0.2676 ± 0.0056	0.4198 ± 0.0047	0.5299 ± 0.0064	0.6151 ± 0.0056	0.6827 ± 0.0055
	DoctorAI	0.1841 ± 0.0215	0.2845 ± 0.0312	0.3635 ± 0.0400	0.4283 ± 0.0485	0.4835 ± 0.0547
	T-LSTM	0.2138 ± 0.0065	0.3278 ± 0.0164	0.4116 ± 0.0151	0.4787 ± 0.0178	0.5277 ± 0.0198
	DeLSTM	0.2584 ± 0.0501	0.4050 ± 0.0776	0.5150 ± 0.0980	0.5975 ± 0.1143	0.6653 ± 0.1257
	RGNN-TG-ATT	0.3089 ± 0.0065	0.4741 ± 0.0046	0.5914 ± 0.0034	0.6754 ± 0.0180	0.7400 ± 0.0036
	StageNet	0.3040 ± 0.0063	0.4693 ± 0.0059	0.5880 ± 0.0065	0.6758 ± 0.0045	0.7465 ± 0.0052
	HiTANET	0.3170 + 0.0093	0.4796 + 0.0149	0.5914 + 0.0035	0.6744 + 0.0018	0.7391 + 0.0036
	MCF-LSTM-SBF	0.3216 ± 0.0029	0.4940 ± 0.0045	0.6154 ± 0.0031	0.7037 ± 0.0029	0.7692 ± 0.0034
	MCF-LSTM-MBF	**0.3321 ± 0.0028**	**0.5043 ± 0.0066**	**0.6218 ± 0.0059**	**0.7069 ± 0.0056**	**0.7722 ± 0.0047**

Table 4
Effect of different auxiliary channels on the medication prediction task on the MIMIC-III dataset.

Method	AUC	AUPR
MCF-LSTM-MBF w/o all	0.7881 ± 0.0080	0.2930 ± 0.0181
MCF-LSTM-MBF w/o diagnosis	0.8581 ± 0.0008	0.4014 ± 0.0014
MCF-LSTM-MBF w/o labtest	0.8578 ± 0.0007	0.4010 ± 0.0019
MCF-LSTM-MBF w/o temporal	0.8578 ± 0.0004	0.3994 ± 0.0013
MCF-LSTM-MBF w/o procedure	0.8584 ± 0.0007	0.4023 ± 0.0019
MCF-LSTM-MBF(ONE)	0.8558 ± 0.0008	0.3978 ± 0.0016
MCF-LSTM-MBF	0.8606 ± 0.0003	0.4068 ± 0.0022

dependencies in a patient visit sequence. To illustrate the ability of the fusion module, we compared the feature representations of information from different channels reduced by t-SNE [45], which can project higher-dimensional features to two-dimensional features for a convenient display. The results are shown in Fig. 5 (take MCF-LSTM for the medication prediction task for example), where *medic*, *lab*, *diag* and *proc* in the legend denote the feature representations of medication, labtest, diagnosis, and procedure respectively after passing through the fusion module, *raw_m*, *raw_l*, *raw_d*, and *raw_p* denotes the original feature representations of medication, labtest, diagnosis, and procedure before passing through the fusion module respectively. We can see that the feature representations of different channels are much closer to each other when they pass through the fusion module. From the distribution of *fusion-before* and *fusion-after*, we can infer that information from all auxiliary channels is well projected into the primary channel via the task-wise fusion module, where the gated network adaptively selects how much the information of the auxiliary channel can transfer to the primary channel.

For the other scenarios of healthcare, MCF-LSTM can integrate other types of information such as clinical text, electrocardiograph (ECG), electroencephalogram (EEG), X-ray, MRI, and medical knowledge by adding new channels. It is a case for further improvement in our future work. In addition, MCF-LSTM only considers temporal intervals at the local level. Another possible direction for further improvement is to introduce temporal intervals at the global level into MCF-LSTM.

The potential benefit of our proposed method can be summarized as follows:

Firstly, in any scenario of medical event prediction, our model can select important auxiliary medical events and successfully integrate

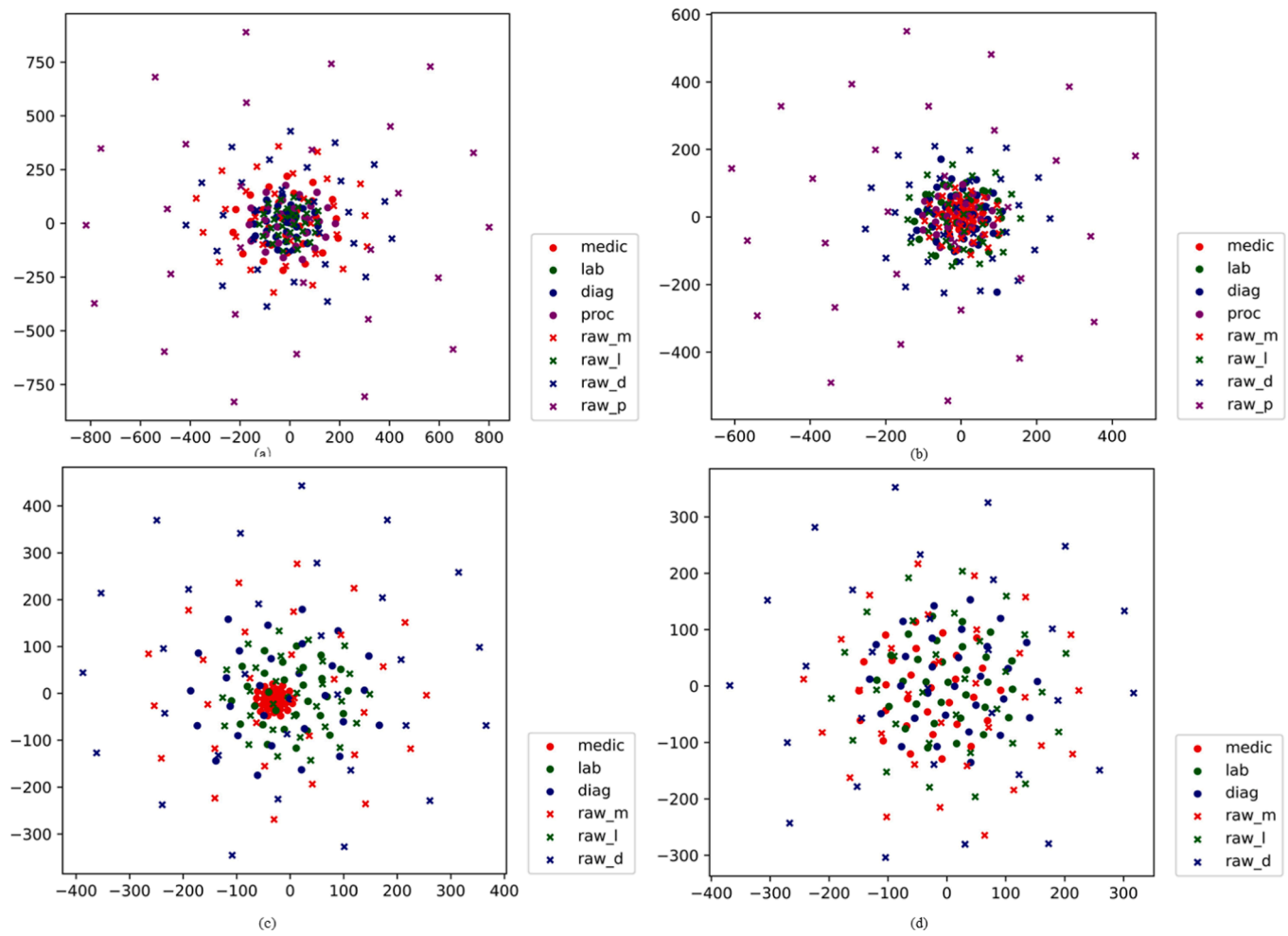

Fig. 5. Feature representations of information from different channels in the reduced two-dimensional coordinate system before/after the fusion module (MCF-LSTM for medication prediction), (a) is MCF-LSTM-SBF on the MIMIC-III dataset, (b) is MCF-LSTM-MBF on the MIMIC-III dataset, (c) is MCF-LSTM-SBF on the eICU dataset, (d) is MCF-LSTM-MBF on the eICU dataset.

them with target medical events, thus it offers an opportunity to explore the generalizability of this predicted model. Additionally, the indefatigable machine learning algorithms may relieve medical practitioners from heavy clinical burdens and reduce mistakes made by humans.

Secondly, it has the potential of mining clinical-specific knowledge and guiding clinical decisions from complex and heterogeneous EHRs by using data-driven deep learning methods. Compared to the traditional modeling approach, it removes the burden of specifying which variables should be used and how they should be combined. It also can learn clinical-specific patterns that cannot be derived from the labor-intensive process of humans. With these successful deep learning techniques, traditional feature learning is transferred to neural network learning. These techniques have a wider range of application scenarios and are more generalizable.

Thirdly, in clinical scenarios, the medical event prediction model maybe effective as a clinical decision-support algorithm. Through continuous data update and iteration of the hospital, the model also can update on the new dataset and be more skilled at mimicking the behavior of experienced physicians, and then serve as clinical auxiliary assistor of the doctor. On the other hand, it can also provide reference information for the inexperienced practitioner.

7. Conclusion

In this study, a novel multi-channel fusion LSTM, called MCF-LSTM, is proposed for medical event prediction. In MCF-LSTM, temporal information is regarded as a new type of medical event, each type of medical event is regarded as a channel, and a task-wise fusion module based on the gated network is introduced into LSTM to fuse information from multiple channels in two different ways. Experiments on two public datasets have shown that MCF-LSTM is an effective task-wise method to model the heterogeneous characteristic and irregular temporal characteristics of medical time series data in a unified framework and outperforms other state-of-the-art methods.

CRediT authorship contribution statement

Sicen Liu: Conceptualization, Methodology, Software, Investigation, Data curation, Validation, Writing – original draft, Visualization. **Xiaolong Wang:** Resources, Supervision, Project administration, Funding acquisition. **Yang Xiang:** Project administration, Funding acquisition. **Hui Xu:** Project administration, Funding acquisition. **Hui Wang:** Project administration, Funding acquisition. **Buzhou Tang:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This paper is supported in part by grants: National Natural Science Foundations of China (U1813215 and 61876052), Special Foundation for Technology Research Program of Guangdong Province (2015B010131010), National Natural Science Foundations of Guangdong, China (2019A1515011158), Science and Technology Planning Project of Shenzhen Municipality (JCYJ20190806112210067), and the Pilot Project in 5G + Health Application of Ministry of Industry and Information Technology & National Health Commission (5G + Luohu Hospital Group: an Attempt to New Health Management Styles of Residents). We also thank Jin's team for sharing their source code of DeLSTM.

References

[1] B. Shickel, P.J. Tighe, A. Bihorac, P. Rashidi, Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis, IEEE J. Biomed. Health Inform. 22 (5) (2018) 1589–1604, https://doi.org/10.1109/JBHI.2017.2767063.

[2] G.S. Birkhead, M. Klompas, N.R. Shah, Uses of electronic health records for public health surveillance to advance public health, Annu. Rev. Public Health. 36 (1) (2015) 345–359, https://doi.org/10.1146/annurev-publhealth-031914-122747.

[3] D.W. Bates, S. Saria, L. Ohno-Machado, A. Shah, G. Escobar, Big data in health care: using analytics to identify and manage high-risk and high-cost patients, Health Aff. 33 (7) (2014) 1123–1131, https://doi.org/10.1377/hlthaff.2014.0041.

[4] G.J. Kuperman, A. Bobb, T.H. Payne, A.J. Avery, T.K. Gandhi, G. Burns, D. C. Classen, D.W. Bates, Medication-related clinical decision support in computerized provider order entry systems: a review, J. Am. Med. Inf. Assoc. 14 (1) (2007) 29–40, https://doi.org/10.1197/jamia.M2170.

[5] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, A guide to deep learning in healthcare, Nat. Med. 25 (1) (2019) 24–29, https://doi.org/10.1038/s41591-018-0316-z.

[6] J.R. Ayala Solares, F.E. Diletta Raimondi, Y. Zhu, F. Rahimian, D. Canoy, J. Tran, A.C. Pinho Gomes, A.H. Payberah, M. Zottoli, M. Nazarzadeh, N. Conrad, K. Rahimi, G. Salimi-Khorshidi, Deep learning for electronic health records: a comparative review of multiple deep neural architectures, J. Biomed. Inf. 101 (2020) 103337, https://doi.org/10.1016/j.jbi.2019.103337.

[7] C.F. Luz, M. Vollmer, J. Decruyenaere, M.W. Nijsten, C. Glasner, B. Sinha, Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies, Clin. Microbiol. Infect. 26 (10) (2020) 1291–1299, https://doi.org/10.1016/j.cmi.2020.02.003.

[8] J. Sterling, A.M. Bernie, R. Ramasamy, Hypogonadism: easy to define, hard to diagnose, and controversial to treat, Canadian Urol. Assoc. J. 9 (1–2) (2015) 65, https://doi.org/10.5489/cuaj.2416.

[9] S. Qiu, P.S. Joshi, M.I. Miller, C. Xue, X. Zhou, C. Karjadi, G.H. Chang, A.S. Joshi, B. Dwyer, S. Zhu, M. Kaku, Y. Zhou, Y.J. Alderazi, A. Swaminathan, S. Kedar, M.-H. Saint-Hilaire, S.H. Auerbach, J. Yuan, E.A. Sartor, R. Au, V.B. Kolachalama, Development and validation of an interpretable deep learning framework for Alzheimer's disease classification, Brain. 143 (6) (2020) 1920–1933, https://doi.org/10.1093/brain/awaa137.

[10] T.A. Buchan, H.J. Ross, M. McDonald, F. Billia, D. Delgado, J.G. Duero Posada, A. Luk, G.H. Guyatt, A.C. Alba, Physician judgement vs model-predicted prognosis in patients with heart failure, Can. J. Cardiol. 36 (1) (2020) 84–91, https://doi.org/10.1016/j.cjca.2019.07.623.

[11] T. Bai, S. Zhang, B.L. Egleston, S. Vucetic, Interpretable representation learning for healthcare via capturing disease progression through time, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, London United Kingdom, 2018, pp. 43–51. https://doi.org/10.1145/3219819.3219904.

[12] L.A. Carrasco-Ribelles, J.R. Pardo-Mas, S. Tortajada, C. Sáez, B. Valdivieso, J. M. García-Gómez, Predicting morbidity by local similarities in multi-scale patient trajectories, J. Biomed. Inf. 120 (2021) 103837, https://doi.org/10.1016/j.jbi.2021.103837.

[13] G. Fleming, M.V.D. Merwe, G. McFerren, Fuzzy expert systems and GIS for cholera health risk prediction in southern Africa, Environ. Modell. Software 22 (4) (2007) 442–448, https://doi.org/10.1016/j.envsoft.2005.12.008.

[14] J. Gao, C. Xiao, Y. Wang, W. Tang, L.M. Glass, J. Sun, StageNet: stage-aware neural networks for health risk prediction, in: Proceedings of The Web Conference 2020, ACM, Taipei Taiwan, 2020, pp. 530–540. https://doi.org/10.1145/3366423.3380136.

[15] Y. An, N. Huang, X. Chen, FangXiang Wu, J. Wang, High-risk prediction of cardiovascular diseases via attention-based deep neural networks, IEEE/ACM Trans. Comput. Biol. Bioinf. 18 (3) (2021) 1093–1105, https://doi.org/10.1109/TCBB.2019.2935059.

[16] H. Harutyunyan, H. Khachatrian, D.C. Kale, G. Ver Steeg, A. Galstyan, Multitask learning and benchmarking with clinical time series data, Sci Data. 6 (2019) 96, https://doi.org/10.1038/s41597-019-0103-9.

[17] D. He, S.C. Mathews, A.N. Kalloo, S. Hutfless, Mining high-dimensional administrative claims data to predict early hospital readmissions, J. Am. Med. Inform. Assoc. 21 (2) (2014) 272–279, https://doi.org/10.1136/amiajnl-2013-002151.

[18] A. Rajkomar, E. Oren, K. Chen, A.M. Dai, N. Hajaj, M. Hardt, P.J. Liu, X. Liu, J. Marcus, M. Sun, P. Sundberg, H. Yee, K. Zhang, Y. Zhang, G. Flores, G.E. Duggan, J. Irvine, Q. Le, K. Litsch, A. Mossin, J. Tansuwan, D. Wang, J. Wexler, J. Wilson, D. Ludwig, S.L. Volchenboum, K. Chou, M. Pearson, S. Madabushi, N.H. Shah, A. J. Butte, M.D. Howell, C. Cui, G.S. Corrado, J. Dean, Scalable and accurate deep learning with electronic health records, Npj Digital Med. 1 (2018) 18, https://doi.org/10.1038/s41746-018-0029-1.

[19] E. Choi, Z. Xu, Y. Li, M. Dusenberry, G. Flores, E. Xue, A. Dai, Learning the graphical structure of electronic health records with graph convolutional transformer, AAAI. 34 (01) (2020) 606–613, https://doi.org/10.1609/aaai.v34i01.5400.

[21] A. Gottlieb, G.Y. Stein, E. Ruppin, R.B. Altman, R. Sharan, A method for inferring medical diagnoses from patient similarities, BMC Med. 11 (1) (2013) 194, https://doi.org/10.1186/1741-7015-11-194.

[22] Z. Jia, X. Zeng, H. Duan, X. Lu, H. Li, A patient-similarity-based model for diagnostic prediction, Int. J. Med. Inf. 135 (2020) 104073, https://doi.org/10.1016/j.ijmedinf.2019.104073.

[23] Z.C. Lipton, D.C. Kale, C. Elkan, R. Wetzel, Learning to Diagnose with LSTM Recurrent Neural Networks, ICLR. (2015). http://arxiv.org/abs/1511.03677 (accessed October 2, 2019).

[24] W. Lee, S. Park, W. Joo, I.-C. Moon, Diagnosis prediction via medical context attention networks using deep generative modeling, in: 2018 IEEE International Conference on Data Mining (ICDM), 2018, pp. 1104–1109. https://doi.org/10.1109/ICDM.2018.00143.

[25] L. Liu, J. Shen, M. Zhang, Z. Wang, J. Tang, Learning the Joint Representation of Heterogeneous Temporal Events for Clinical Endpoint Prediction, Thirty-Second AAAI Conference on Artificial Intelligence. (2018) 109–116.

[26] H. Song, D. Rajan, J.J. Thiagarajan, A. Spanias, Attend and diagnose: clinical time series analysis using attention models, in: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), n.d., p. 8.

[27] B. Jin, H. Yang, L. Sun, C. Liu, Y. Qu, J. Tong, A Treatment Engine by Predicting Next-Period Prescriptions, in: KDD'18, ACM Press, London, United Kingdom, 2018, pp. 1608–1616. https://doi.org/10.1145/3219819.3220095.

[28] S. Liu, T. Li, H. Ding, B. Tang, X. Wang, Q. Chen, J. Yan, Y.i. Zhou, A hybrid method of recurrent neural network and graph neural network for next-period prescription prediction, Int. J. Mach. Learn. Cyber. 11 (12) (2020) 2849–2856, https://doi.org/10.1007/s13042-020-01155-x.

[30] E. Choi, C. Xiao, W. Stewart, J. Sun, MiME: Multilevel Medical Embedding of Electronic Health Records for Predictive Healthcare, in: Advances in Neural Information Processing Systems 31, Curran Associates, Inc., 2018.

[31] E. Choi, M.T. Bahadori, A. Schuetz, W.F. Stewart, J. Sun, Doctor AI: predicting clinical events via recurrent neural networks, JMLR Workshop Conf. Proc. 56 (2016) 301–318.

[32] I.M. Baytas, C. Xiao, X. Zhang, F. Wang, A.K. Jain, J. Zhou, patient subtyping via time-aware LSTM networks, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Halifax NS Canada, 2017, pp. 65–74. https://doi.org/10.1145/3097983.3097997.

[33] S.N. Shukla, B.M. Marlin, Interpolation-prediction networks for irregularly sampled time series, ICLR. (2019).

[34] Q. Tan, M. Ye, B. Yang, S. Liu, A.J. Ma, T.-C.-F. Yip, G.-L.-H. Wong, P. Yuen, DATA-GRU: dual-attention time-aware gated recurrent unit for irregular multivariate time series, AAAI. 34 (01) (2020) 930–937, https://doi.org/10.1609/aaai.v34i01.5440.

[35] X. Xiao, G. Wei, L.i. Zhou, Y.i. Pan, H. Jing, E. Zhao, Y. Yuan, Treatment initiation prediction by EHR mapped PPD tensor based convolutional neural networks boosting algorithm, J. Biomed. Inf. 120 (2021) 103840, https://doi.org/10.1016/j.jbi.2021.103840.

[36] A.E.W. Johnson, T.J. Pollard, L. Shen, L.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R.G. Mark, MIMIC-III, a freely accessible critical care database, Sci Data. 3 (2016), 160035, https://doi.org/10.1038/sdata.2016.35.

[37] J. Luo, M. Ye, C. Xiao, F. Ma, HiTANet: hierarchical time-aware attention networks for risk prediction on electronic health records, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, Virtual Event CA USA, 2020. pp. 647–656. https://doi.org/10.1145/3394486.3403107.

[38] R. Miotto, F. Wang, S. Wang, X. Jiang, J.T. Dudley, Deep learning for healthcare: review, opportunities and challenges, Briefings Bioinf. 19 (2018) 1236–1246, https://doi.org/10.1093/bib/bbx044.

[42] R. Moskovitch, F. Polubriaginof, A. Weiss, P. Ryan, N. Tatonetti, Procedure prediction from symbolic Electronic Health Records via time intervals analytics, J. Biomed. Inf. 75 (2017) 70–82, https://doi.org/10.1016/j.jbi.2017.07.018.

[44] B.C. Kwon, M.-J. Choi, J.T. Kim, E. Choi, Y.B. Kim, S. Kwon, J. Sun, J. Choo, RetainVis: visual analytics with interpretable and interactive recurrent neural networks on electronic medical records, IEEE Trans. Visual. Comput. Graphics. 25 (1) (2019) 299–309, https://doi.org/10.1109/TVCG.2018.2865027.

[45] Laurens, Maaten Van Der, G. Hinton, Laurens vanLaurens van der Maaten, Visualizing data using **t-SNE**, J. Mach. Learn. Res. 9 (2605) (2008) 2579–2605.