Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Special Track on AI for Good

Chronic Disease Management with Personalized Lab Test Response Prediction

Suman Bhoi[1] , **Mong Li Lee**[1] , **Wynne Hsu**[1] , **Hao Sen Andrew Fang**[2] and **Ngiap Chuan Tan**[2]
[1]National University of Singapore
[2]SingHealth Polyclinics
sumanbhoi@u.nus.edu, {leeml, whsu}@comp.nus.edu.sg, {andrew.fang.h.s, tan.ngiap.chuan}@singhealth.com.sg

Abstract

Chronic disease management involves frequent administration of invasive lab procedures in order for clinicians to determine the best course of treatment regimes for these patients. However, patients are often put off by these invasive lab procedures and do not follow the appointment schedules. This has resulted in poor management of their chronic conditions leading to unnecessary disease complications. An AI system that is able to personalize the prediction of individual patient lab test responses will enable clinicians to titrate the medications to achieve the desired therapeutic outcome. Accurate prediction of lab test response is a challenge because these patients typically have co-morbidities and their treatments might influence the target lab test response. To address this, we model the complex interactions among different medications, diseases, lab test response, and fine-grained dosage information to learn a strong patient representation. Together with information from similar patients and external knowledge such as drug-lab interactions and diagnosis-lab interaction, we design a system called KALP to perform personalized prediction of patients' response for a target lab result and identify the top influencing factors for the prediction. Experiment results on real world datasets demonstrate the effectiveness of KALP in reducing prediction errors by a significant margin. Case studies show that the identified factors are consistent with clinicians' understanding.

1 Introduction

Electronic Health Records (EHR) provide a rich repository of patient related information such as disease diagnosis, lab test results, prescribed medications, etc. over a history of visits. Advances in machine learning have led to the large-scale analytics of EHR for disease inference [Ni *et al.*, 2017], mortality prediction [Tan *et al.*, 2019], personalized medication recommendation [Wang *et al.*, 2019; Shang *et al.*, 2019; Bhoi *et al.*, 2021]. At the same time, there is a growing trend towards using patient analytics for decision support [Oei *et al.*, 2021] to improve patient care and clinical outcomes, particularly in chronic disease management such as hypertension and diabetes. Often, these chronic disease patients are prescribed some standard treatment regime for each chronic condition independently and their conditions are monitored periodically based on lab test results such as HbA1c[1]. For patients with co-morbidities, studies have shown that their lab test results are often influenced by other treatments related to their co-morbidities, making it hard to assess the effectiveness of the prescribed treatment [Unnikrishnan *et al.*, 2012]. The ability to predict the target lab test result of a patient, taking into consideration his condition and medications that are prescribed for co-morbidities, would enable the clinician to personalize the treatment regime for a patient. This can help eliminate invasive procedures associated with sample collection for the lab test, contributing to the Sustainable Development Goal of Good Health and Well-being[2].

Existing research use patient specific information such as demographics and past visit records to predict lab test results [Luo *et al.*, 2016; Kang, 2018]. These works do not consider the impact of medications on the target lab test result. For example, patients with high blood pressure are prescribed medications like Propranolol which is known to increase HbA1c [Dornhorst *et al.*, 1985]. In practice, lab test results are often influenced by drugs and diagnosis. However, these drug-lab interactions and diagnosis-lab interactions are largely ignored by current works on lab response predictions. Another aspect that has been overlooked is to leverage on patient analytics where patients with similar demographics and diagnosis tend to have similar lab test responses.

In this work, we use a transformer encoder to capture the patient specific information, while the information of similar patients is modeled using the modified graph attention network (GATv2). With this, we obtain a strong latent patient representation and incorporate fine-grained dosage information to accurately predict patient response to a target lab test in the presence of medication titrations for chronic disease patient management. This is in contrast to existing works which mainly consider medication type. The contributions of this work are as follows:

[1]https://medlineplus.gov/lab-tests/hemoglobin-a1c-hba1c-test/

[2]Specific target indicators being addressed are 3.4.1 and 3.d.1 under the third SDG.

- We design a personalized deep learning system that combines the sequential information in patient visit history and information from similar patients to generate a strong patient representation.
- We augment the patient representation with external knowledge of drug-lab interactions and diagnosis-lab interactions to model the complex relationships among drugs, co-morbidities, and lab test results.
- We identify the factors that influenced the predicted lab test results, thus providing insights to changes (if any) in the lab results.
- Extensive quantitative and qualitative experiments on the benchmark MIMIC-III [Johnson *et al.*, 2016] EHR and a proprietary outpatient dataset demonstrate the effectiveness of the proposed system to significantly lower the prediction errors by a large margin.

2 Related Works

Lab test prediction has been studied using patient EHR and visit information [Luo *et al.*, 2016; Kang *et al.*, 2015; Kang *et al.*, 2017; Kang, 2018]. Learning a strong latent patient representation is key to the accurate prediction of lab test results. Researchers have used recurrent neural network (RNN), bidirectional RNN, time-aware long-short term memory and memory augmented RNNs to handle long-range dependencies in patient visit history when learning patient representations [Choi *et al.*, 2016; Ma *et al.*, 2017; Baytas *et al.*, 2017; Sukhbaatar *et al.*, 2015].

Transformer encoder representations have been shown to be superior over RNN based approaches in capturing the sequential dependency in EHR for disease prediction [Li *et al.*, 2020], mortality prediction [Darabi *et al.*, 2020], medication recommendation [Prakash *et al.*, 2021]. HiTANet [Luo *et al.*, 2020] introduces a time-aware transformer to obtain a patient representation using local and global attention for risk prediction. BEHRT [Li *et al.*, 2020] adapts a Transformer-based architecture [Kenton and Toutanova, 2019] to learn patient representation using patient's diagnostic and demographic information to predict future disease occurrence. All these works do not consider information from other similar patients and do not use fine-grained dosage information.

The work in [Lu *et al.*, 2021] models patient similarity and ontology of diseases to learn disease representation while incorporating clinical notes for the task of diagnosis and heart failure prediction. Our work differs from this in that our notion of patient similarity includes demographics, prescribed medications information in addition to disease diagnosis to predict lab test responses.

3 Methodology

Given a sequence of past patient visits, our task is to predict the result of a target lab test for the patient's current visit. Figure 1 gives an overview of the proposed solution called KALP for Knowledge Augmented Lab test Prediction. There are two key components: (a) learning an effective latent patient representation and (b) augmenting it with knowledge of drug-lab and disease-lab interactions, as well as patients' historical responses to the lab test.

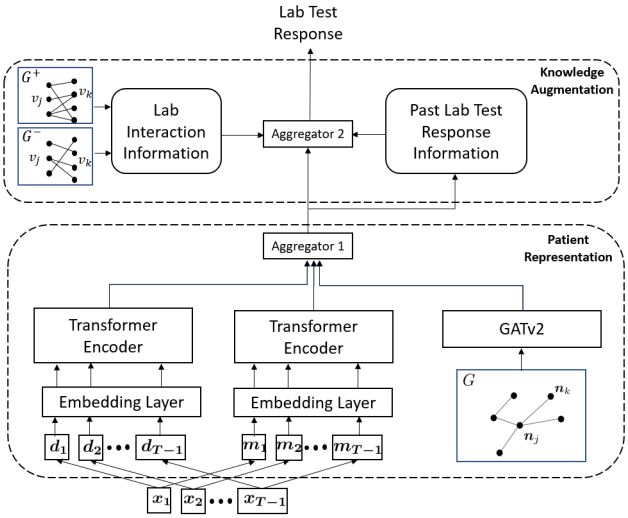

Figure 1: Overview of KALP.

3.1 Patient Representation

Our proposed patient representation takes into account both the sequential dependency in patient diagnosis and medication information, as well as information from similar patients. Suppose a patient has $T-1$ visits prior to his current visit. We represent the i^{th} visit of a patient as

$$\boldsymbol{x}_i = [\boldsymbol{d}_i, \boldsymbol{m}_i, l_i]$$

where $\boldsymbol{d}_i$ is a multi-hot vector depicting the diagnosis, $\boldsymbol{m}_i$ is a vector of (medication, dosage) pairs, and l_i denotes the result of the target lab test, $1 \leq i \leq T-1$.

Sequential Dependency. We utilize two transformers to model the sequential dependency in the patients' diagnosis and medication information over the visits. Having dual transformers enable us to obtain representation for patients who may have missing diagnosis or medication information. We linearly embed $\boldsymbol{d}_i$ and $\boldsymbol{m}_i$, into a low dimensional space using embedding matrices $\boldsymbol{E}^d$ and $\boldsymbol{E}^m$ respectively. These linear embeddings are combined with positional embeddings [Vaswani *et al.*, 2017] to obtain the embeddings for diagnosis $\boldsymbol{d}_i^e$ and medication $\boldsymbol{m}_i^e$ which are then passed to their respective transformers (see Figure 2).

Each transformer consists of multiple layers with each layer having $T-1$ encoders. Each encoder has a position-wise fully connected feed-forward network and a multi-head self-attention mechanism. The position-wise feed-forward network has two linear transformations with a ReLU layer in between. The multi-head attention employs scaled dot-product attention to obtain the weights on the patient visits given the visit history. Since this multi-head attention can attend to future time steps, to ensure that the model's predictions are only conditioned on past visits, we apply a triangular mask to the embedding $\boldsymbol{d}_i^e$ and $\boldsymbol{m}_i^e$. This mask is the same as the one used in the decoder component. The input to the i^{th} encoder in the first layer is the concatenation

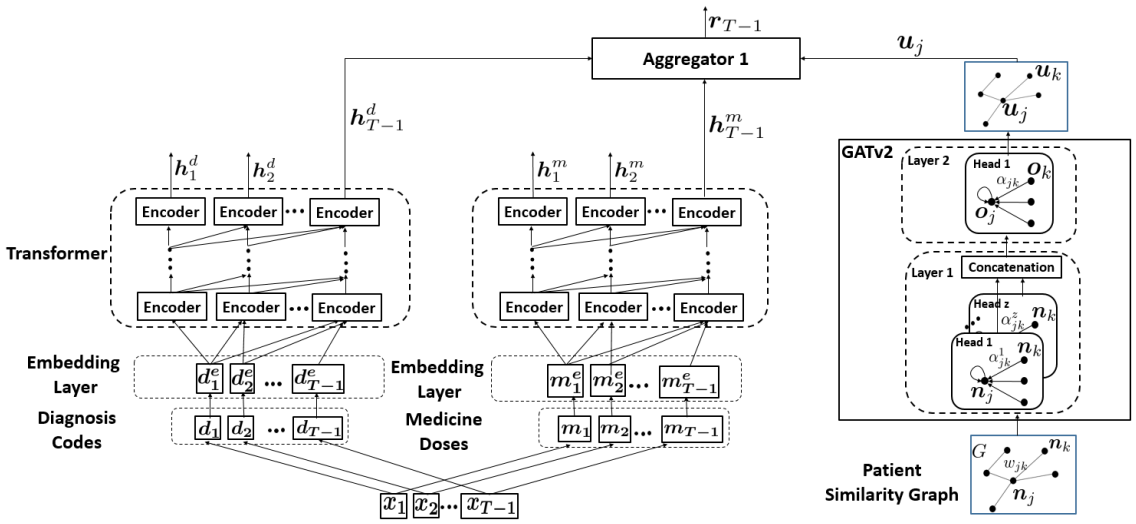

Figure 2: Details of the patient representation component.

of $\boldsymbol{d}_j^e$, $1 \leq j \leq i$. Inputs to the i^{th} encoder in the subsequent layers is the concatenation of the outputs from the first to the i^{th} encoders of the previous layers. We employ residual connection around the self-attention mechanism and the feed-forward network is followed by layer normalization. We also apply dropout to avoid over-fitting. The outputs of the transformers for diagnosis and medications are given by:

$$[\boldsymbol{h}_1^d, \boldsymbol{h}_2^d, ..., \boldsymbol{h}_{T-1}^d] = \text{Transformer}\left([\boldsymbol{d}_1^e, \boldsymbol{d}_2^e, ..., \boldsymbol{d}_{T-1}^e]\right) \quad (1)$$

$$[\boldsymbol{h}_1^m, \boldsymbol{h}_2^m, ..., \boldsymbol{h}_{T-1}^m] = \text{Transformer}\left([\boldsymbol{m}_1^e, \boldsymbol{m}_2^e, ..., \boldsymbol{m}_{T-1}^e]\right) \quad (2)$$

where $\boldsymbol{h}_i^d$ and $\boldsymbol{h}_i^m$ are the outputs of i^{th} encoder in the last layer of the respective transformers. Since our goal is to predict the lab test result for the current visit T, we use the outputs $\boldsymbol{h}_{T-1}^d$ and $\boldsymbol{h}_{T-1}^m$ as the encoded sequential diagnosis and medication information.

Patient Similarity Information. Patient similarity has been used to improve the accuracy of diagnosis prediction [Jia *et al.*, 2020] and treatment recommendation [Wirbka *et al.*, 2020]. The work in [MacDonald *et al.*, 2010] observe that patients with similar demographics and diagnosis tend to have similar lab test responses. As such, we incorporate information from similar patients to learn a more effective patient representation to improve the lab result prediction accuracy.

We construct a weighted patient similarity graph $G = (V, E)$ where each node $\boldsymbol{n}_i \in V$ denotes a patient i, and each labelled edge $(\boldsymbol{n}_i, \boldsymbol{n}_j, w_{ij}) \in E$ denotes that patient i is similar to patient j with a degree of similarity w_{ij}. Here, $\boldsymbol{n}_i$ is based on patient information at their first visit, together with their age, weight, and gender information.

Note that any similarity measure can be used. Here we use cosine similarity between $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$ to compute w_{ij}. We adapt GATv2 [Brody *et al.*, 2021] to learn the node representation of the weighted patient similarity graph G. GATv2 has two layers with z attention heads in the first layer, and 1 attention head in the second layer as shown in Figure 2. The attention weight between nodes $\boldsymbol{n}_j$ and $\boldsymbol{n}_k$ for the b^{th} attention head in the first layer is given by:

$$\alpha_{jk}^b = \frac{w_{jk} \times \text{GATv2}(\boldsymbol{E}^b \cdot [\boldsymbol{n}_j \| \boldsymbol{n}_k])}{\sum_{\boldsymbol{n}_i \in S_j} w_{ij} \times \text{GATv2}(\boldsymbol{E}^b \cdot [\boldsymbol{n}_j \| \boldsymbol{n}_i])} \quad (3)$$

where $\|$ denotes concatenation, S_j is the set of nodes whose similarity with node j is non-zero, $\boldsymbol{E}^b$ is the embedding matrix of the b^{th} attention head, and w_{jk} denotes the edge weight representing the similarity between patients j and k.

The output from the first layer for a node n_j is given by:

$$\boldsymbol{o}_j = \|_{b=1}^{z} \sigma\left(\sum_{\boldsymbol{n}_k \in S_j} \alpha_{jk}^b \boldsymbol{E}^b \cdot \boldsymbol{n}_k\right) \quad (4)$$

where σ is the sigmoid function. With this, the output from the second layer for node n_j can be obtained as follows:

$$\boldsymbol{u}_j = \sigma\left(\sum_{\boldsymbol{n}_k \in \mathcal{S}_j} \alpha_{jk} \boldsymbol{E} \cdot \boldsymbol{o}_k\right) \quad (5)$$

where α_{jk} is calculated by using the average of the attention weight between node j and k from the first layer as new edge weights, i.e., w_{jk} is updated to the average of α_{jk}^b for $1 \leq b \leq z$, $\boldsymbol{E}$ is the embedding matrix for the second layer. Combining this $\boldsymbol{u}_j$ with the diagnosis and medication representation of the most recent visit $\boldsymbol{h}_{T-1}^d$ and $\boldsymbol{h}_{T-1}^m$, we obtain the patient representation $\boldsymbol{r}_{T-1}$:

$$\boldsymbol{r}_{T-1} = \boldsymbol{h}_{T-1}^d + \boldsymbol{h}_{T-1}^m + \boldsymbol{u}_j \quad (6)$$

3.2 Knowledge Augmentation

Next, we augment the obtained patient representation with drug-lab interactions and disease-lab interactions, as well as patients' past responses to the lab test.

Lab Interaction Information. We obtain the drug-lab and disease-lab interactions from *AACC Effects on Clinical Laboratory Tests database*[3], SIDER [Kuhn *et al.*, 2016] and

[3]https://clinfx.wiley.com/aaccweb/aacc/

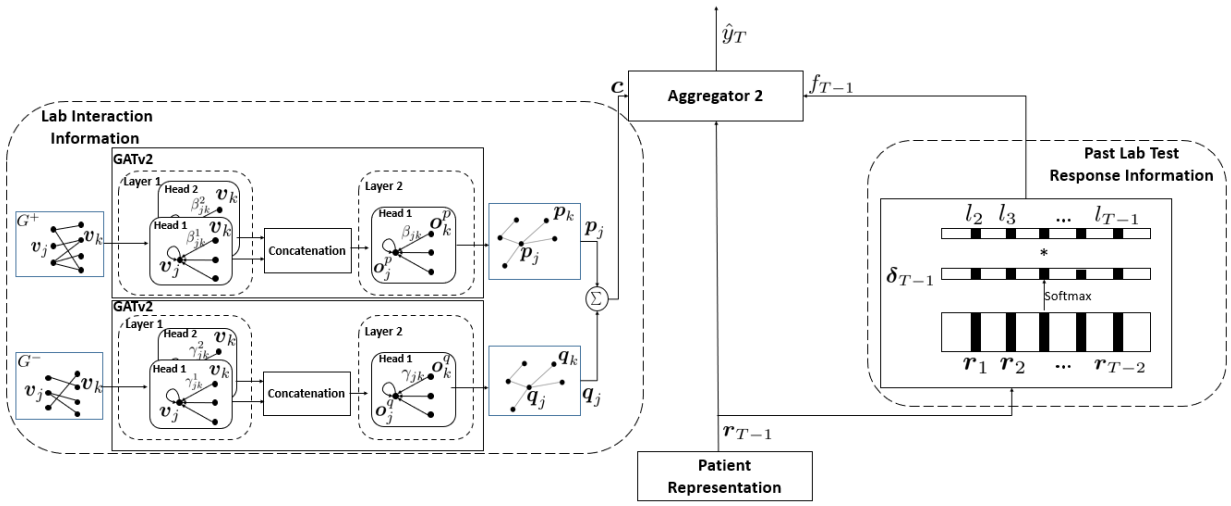

Figure 3: Details of the knowledge augmentation component.

MEDI [Wei *et al.*, 2013]. Incorporating this information makes the system aware of the impact of these interactions on the lab test values. Since the interactions can be positive (increases the lab test value) or negative (decreases the lab test value), we use two bipartite graphs to represent them: a positive lab interaction graph $G^+ = (X, Y^+)$ and a negative lab interaction graph $G^- = (X, Y^-)$ where each node $v \in X$ denotes either a lab test, medication or diagnosis and a labelled edge $(v_i, v_j) \in Y^+/Y^-$ denotes positive/negative interaction between nodes v_i and v_j. The lab tests constitute one type of nodes, and medications, diagnosis together constitute another type of nodes and there is no edge between nodes of the same type.

Similar to what we have done for the weighted patient similarity graph, we use GATv2 with two layers and two attention heads in the first layer to learn the node representations from the positive and negative lab interaction graphs (see Figure 3). The node representations of G^+ and G^- are mapped to a low dimensional space by applying the embedding matrices $\boldsymbol{E}^{b+}$ and $\boldsymbol{E}^{b-}$ respectively in the first layer for the b^{th} attention head where $1 \leq b \leq 2$. With this, we obtain the node representations $\boldsymbol{p}$ and $\boldsymbol{q}$ of the target lab test from G^+ and G^- respectively. Then the lab interaction vector $\boldsymbol{c}$ is given by:

$$\boldsymbol{c} = \boldsymbol{p} + \lambda \boldsymbol{q} \tag{7}$$

where λ regulates the fusion of positive and negative lab interaction information.

Past Lab Test Response Information. Since patients respond to treatments differently, incorporating a patient's own past lab test responses will guide KALP to generate personalized lab responses. We have a repository of patients' past lab test responses stored in the form of key-value pairs where the key is the patient representation and the value is the lab test response. Given the patient representation $\boldsymbol{r}_{T-1}$, we compute the attention on the previous representations $\boldsymbol{r}_i$, $1 \leq i \leq T-2$, stored in the repository as follows:

$$\boldsymbol{\delta}_{T-1} = softmax(\boldsymbol{r}_1 \cdot \boldsymbol{r}_{T-1},\ \boldsymbol{r}_2 \cdot \boldsymbol{r}_{T-1}, ..., \boldsymbol{r}_{T-2} \cdot \boldsymbol{r}_{T-1}) \tag{8}$$

With this attention $\boldsymbol{\delta}_{T-1} \in \mathbb{R}^{T-2}$, we obtain the weighted past visit lab response $\boldsymbol{f}_{T-1}$ as follows:

$$\boldsymbol{f}_{T-1} = \sum_{i=1}^{T-2} \delta_{T-1}[i] \cdot l_{i+1} \tag{9}$$

where $\delta_{T-1}[i]$ depicts the i^{th} entry in the attention vector $\boldsymbol{\delta}_{T-1}$, and l_{i+1} is the lab test result at the $(i+1)^{th}$ visit. By concatenating $\boldsymbol{r}_{T-1}$, $\boldsymbol{c}$, and $\boldsymbol{f}_{T-1}$, we obtain the final output vector which is then passed through a linear layer to predict the result of the target lab test:

$$\hat{y}_T = \boldsymbol{w} \times (\boldsymbol{r}_{T-1} \parallel \boldsymbol{c} \parallel f_{T-1}) \tag{10}$$

where $\boldsymbol{w}$ is the gradient vector of the transformation function in the linear layer. Since the prediction of lab test result is a regression task, our objective function is to minimize the square of the error defined as follows:

$$\mathcal{L}_{\text{mse}} = \frac{1}{T-1} \sum_{i=2}^{T} (y_i - \hat{y}_i)^2 \tag{11}$$

where T is the total number of visits, $\hat{y}_i$ and y_i are the predicted and ground-truth lab test result for the i^{th} visit.

3.3 Influential Factors

The design of our system allows us to identify the factors that have influenced the predicted value $\hat{y}_T$ of the target lab test. Let $\boldsymbol{w} = \boldsymbol{w}^r \parallel \boldsymbol{w}^c \parallel \boldsymbol{w}^f$ where $\boldsymbol{w}^r$, $\boldsymbol{w}^c$, and $\boldsymbol{w}^f$ are the sub-vectors of the linear transformation $\boldsymbol{w}$ in Equation 10 corresponding to patient representation $\boldsymbol{r}_{T-1}$, lab interaction $\boldsymbol{c}$, and past lab test results $\boldsymbol{f}_{T-1}$ respectively.

Then the influence of each diagnosis for the past visits is given by:

$$\boldsymbol{\eta}^d = \|_{i=1}^{T-2} a^d_{T-1}[i] \times (\boldsymbol{w}^r \cdot \boldsymbol{E}^d[:, k]) \times \boldsymbol{d}_i[k] \tag{12}$$

where $\boldsymbol{a}^d_{T-1}$ is the average of attention weights across all the heads and layers of the encoders in the transformer for the

diagnosis information for time $T-1$, $a^d_{T-1}[i]$ is the i^{th} entry in $\boldsymbol{a}^d_{T-1}$, $\boldsymbol{E}^d$ is the embedding matrix for diagnosis, and $\boldsymbol{d}_i[k]$ is the k^{th} diagnosis in the i^{th} visit.

Similarly, the influence of a patient's prescribed medication on the predicted lab test result is:

$$\boldsymbol{\eta}^m = \|_{i=1}^{T-2} a^m_{T-1}[i] \times (\boldsymbol{w}^r \cdot \boldsymbol{E}^m[:,k]) \times \boldsymbol{m}_i[k] \quad (13)$$

where $\boldsymbol{a}^m_{T-1}$ is the average of attention weights across all the heads and layers of the encoders in the transformer for the medication information for time $T-1$, $a^m_{T-1}[i]$ is the i^{th} entry in $\boldsymbol{a}^m_{T-1}$, $\boldsymbol{E}^m$ is the embedding matrix for medication, $\boldsymbol{m}_i[k]$ is the k^{th} medication in the i^{th} visit.

We obtain the influence of a patient's similarity with other patients as follows:

$$\boldsymbol{\eta}^s = \|_{k \in \mathcal{S}_j} a^s_j[k] \times \boldsymbol{w}^r \cdot (\boldsymbol{E}^s \cdot \boldsymbol{n}_k) \quad (14)$$

where $\boldsymbol{a}^s_j$ is the average of attention weights across all heads and layers of GATv2 for node j, $a^s_j[k]$ is the k^{th} entry in $\boldsymbol{a}^s_j$, $\boldsymbol{E}^s$ is the average of embedding matrix $\boldsymbol{E}^b$ over all heads for the nodes in the weighted patient similarity graph, and $\boldsymbol{n}_k$ is the one-hot vector of the node for the k^{th} patient.

The influence of both positive and negative lab interactions are given by:

$$\boldsymbol{\eta}^p = \|_{k \in \mathcal{S}_j} a^p_j[k] \times \boldsymbol{w}^c \cdot (\boldsymbol{E}^p \cdot \boldsymbol{v}_k) \quad (15)$$

$$\boldsymbol{\eta}^q = \|_{k \in \mathcal{S}_j} \lambda \times a^q_j[k] \times \boldsymbol{w}^c \cdot (\boldsymbol{E}^q \cdot \boldsymbol{v}_k) \quad (16)$$

where $\boldsymbol{a}^p_j$ and $\boldsymbol{a}^q_j$ are the averages of attention weights across all heads and layers of the GATv2 for the lab interaction graphs G^+ and G^- respectively for node j, $a^p_j[k]$ is the k^{th} entry in $\boldsymbol{a}^p_j$ and $a^q_j[k]$ is the k^{th} entry in $\boldsymbol{a}^q_j$, $\boldsymbol{E}^p$ and $\boldsymbol{E}^q$ are the average of embedding matrices $\boldsymbol{E}^{b+}$ and $\boldsymbol{E}^{b-}$ over all heads in G^+ and G^- respectively, and $\boldsymbol{v}_k$ is the one-hot vector of the k^{th} node in the graphs.

Finally, the influence of a patient's past lab test responses is given by:

$$\boldsymbol{\eta}^l = \|_{i=1}^{T-2} \boldsymbol{w}^f \times \delta_{T-1}[i] \times l_{i+1} \quad (17)$$

We normalize all these influences and rank them to obtain the top factors that could have influenced the prediction of the target lab test response.

4 Performance Study

We implemented KALP in PyTorch and trained the models on two NVIDIA Titan RTX GPU. We adopt the widely used metrics for evaluation namely, root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The following datasets are used:

MIMIC-III [Johnson *et al.*, 2016]. This is the largest publicly available EHR dataset which contains clinical data for 7870 neonates (infants) and 38,597 adults admitted to ICU between 2001 and 2008, and captures attributes such as lab reports, medications, etc.

PRIVATE. This is a 10-year outpatient proprietary dataset. Compared to the inpatient MIMIC III, the number of diagnosis per patient is fewer in this dataset.

In our experiments, the target lab test is HbA1c. We filter out the patients who have less than two HbA1c results and perform 10 fold cross validation. We compare KALP with the following baselines:

- **Previous Value (PV).** The predicted lab response for the current visit is given by the previous visit.
- **Linear Regression (LR).** This is the least square method based linear regression.
- **Nearest Neighbour (NN).** The lab test response of the most similar patient is used.
- **RNN [Kang, 2018].** This method employs gated recurrent neural networks to predict lab test results.
- **DMNC [Le *et al.*, 2018].** We adapt this dual memory neural computer to predict lab test response taking into account demographics, medication, and diagnosis.
- **HiTANet [Luo *et al.*, 2020].** We adapt HiTANet to use demographics, medication, and diagnosis to predict lab response.
- **BEHRT [Li *et al.*, 2020].** We adapt BEHRT to use medication in addition to diagnosis and demographics for lab response prediction.

Table 1 shows that KALP outperforms all baselines on both MIMIC-III and PRIVATE. The one-way ANOVA [Fisher, 1992] test shows that the improvements are statistically significant with p-values < 0.05. Compared to DMNC and BEHRT which model only sequential dependency of a patient's visits over time, our knowledge augmented approach dramatically widens the performance gap. This suggests that incorporating similar patients, lab interactions and past lab test responses is effective in reducing the prediction errors.

Table 2 shows the performance after medication titration. Here, we use a subset of the patients whose medications have dosage changes between visits. All the methods show an increase in the prediction errors compared to Table 1 as it is difficult to predict the lab test results for these patients which may not follow the general trend after titration. Despite this, KALP has the lowest prediction errors demonstrating its applicability in the real world for chronic disease management.

Additional ablation study shows that not using any medication or patient similarity information leads to the highest RMSE, MAE, and MAPE on both datasets.

5 Case Study

Finally, we present a case study from PRIVATE to show how KALP is able to identify the top influential factors that led to its lab result predictions. Figure 4 shows a chronic diabetic patient with HbA1c ranging between 5.4 and 6.5. KALP predicts the HbA1c with an error margin of 0.2 while the state-of-the-art BEHRT has an error margin as large as 0.7 (see Visit 4). On Visit 3 where the ground truth HbA1c is 5.4, KALP predicts the HbA1c to be 5.5 while BEHRT predicts it to be 5.1. The top influential factor for this prediction is the past lab response, highlighting the need to take into consideration patients' past responses to personalize the prediction. For Visit 4, we again see that KALP is able to predict the rise in the HbA1c value to within 0.1 of the ground truth value

Methods	MIMIC-III			PRIVATE		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
PV	2.39 ± 0.21	1.91 ± 0.19	14.47 ± 2.51	2.08 ± 0.17	1.95 ± 0.19	13.22 ± 2.18
LR	2.91 ± 0.25	2.52 ± 0.27	18.26 ± 2.76	2.38 ± 0.21	2.05 ± 0.20	15.34 ± 2.45
NN	2.61 ± 0.20	2.18 ± 0.21	16.58 ± 2.57	2.11 ± 0.19	1.94 ± 0.17	14.48 ± 2.33
RNN	2.37 ± 0.18	1.88 ± 0.19	14.04 ± 2.21	2.04 ± 0.17	1.72 ± 0.15	12.53 ± 2.13
DMNC	1.95 ± 0.16	1.67 ± 0.15	13.19 ± 2.08	1.83 ± 0.15	1.55 ± 0.12	10.64 ± 2.01
HiTANet	1.89 ± 0.17	1.51 ± 0.15	12.37 ± 2.11	1.72 ± 0.14	1.43 ± 0.11	9.52 ± 1.93
BEHRT	1.57 ± 0.14	1.38 ± 0.13	11.28 ± 1.95	1.41 ± 0.11	1.24 ± 0.13	8.49 ± 1.82
KALP	**1.15 ± 0.11***	**0.80 ± 0.10***	**6.87 ± 1.63***	**0.85 ± 0.09***	**0.69 ± 0.07***	**3.42 ± 1.47***

* indicates that the result is statistically significant when compared to the second best with p-value < 0.05.

Table 1: Results for comparative study.

Methods	MIMIC-III			PRIVATE		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
PV	4.16 ± 0.23	3.73 ± 0.21	23.11 ± 2.61	3.91 ± 0.24	3.27 ± 0.21	20.52 ± 2.32
LR	4.56 ± 0.31	3.98 ± 0.28	24.03 ± 2.89	4.05 ± 0.29	3.43 ± 0.26	21.04 ± 2.67
NN	4.12 ± 0.28	3.69 ± 0.27	22.45 ± 2.77	3.87 ± 0.24	3.22 ± 0.22	20.26 ± 2.51
RNN	3.62 ± 0.25	3.26 ± 0.26	20.64 ± 2.64	3.34 ± 0.23	3.01 ± 0.21	19.16 ± 2.39
DMNC	3.18 ± 0.22	2.76 ± 0.23	19.01 ± 2.55	2.89 ± 0.21	2.33 ± 0.20	17.02 ± 2.25
HiTANet	3.08 ± 0.19	2.75 ± 0.15	18.41 ± 2.41	2.65 ± 0.18	2.19 ± 0.17	16.18 ± 2.21
BEHRT	2.65 ± 0.18	2.39 ± 0.17	17.70 ± 2.43	2.26 ± 0.16	2.02 ± 0.15	15.02 ± 2.23
KALP	**1.85 ± 0.13***	**1.34 ± 0.12***	**10.28 ± 2.13***	**1.27 ± 0.10***	**1.10 ± 0.11***	**8.18 ± 1.96***

* indicates that the result is statistically significant when compared to the second best with p-value < 0.05.

Table 2: Results when there are changes in medication dosage.

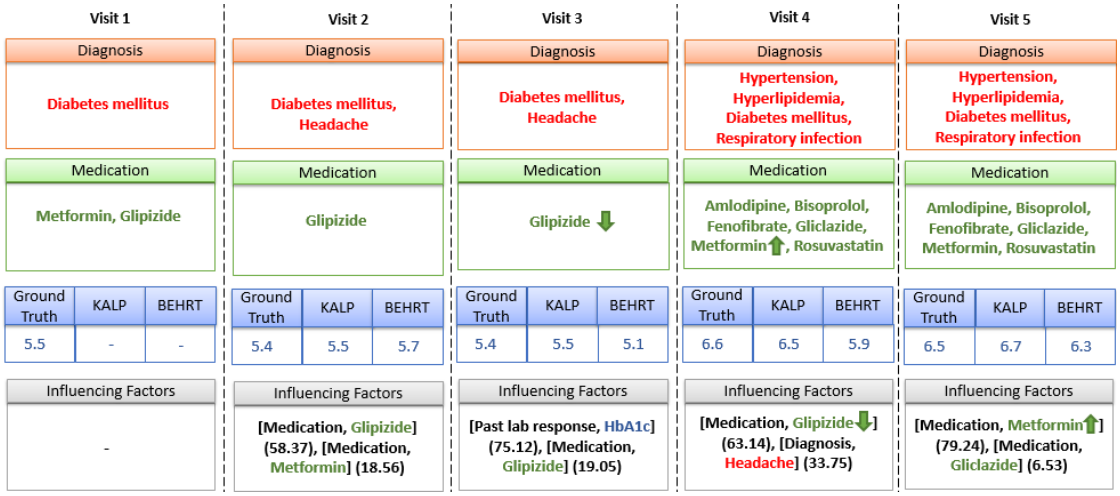

Figure 4: Predicted and ground truth HbA1c. Influencing factors are obtained from KALP. Blue depicts HbA1c values, red depicts diagnosis, and green depicts medication. ⇓ and ⇑ depict decrease and increase of dosage respectively.

compared to BEHRT. A closer examination reveals that there is a reduction in the dosage of Glipizide on Visit 3 and KALP has correctly attributed the top influential factor for this prediction to be the reduction in Glipizide dosage.

6 Conclusion

In this work, we have described a personalized lab test result prediction approach that learns a strong patient representation incorporating both patient information accumulated over the visits as well as information from similar patients. This representation captures fine-grained dosage information enabling us to adjust the prediction in response to changes in treatment regime, which is often needed in the management of chronic patient care. To the best of our knowledge, KALP is the only system that takes into account dosage information when making lab test predictions. Experimental results on two real-world datasets demonstrate the effectiveness of KALP in providing predictions that are close to the actual lab results. Future work includes extending KALP to incorporate the severity and frequency of lab interactions, while jointly modeling the plethora of multi-modal EHR information.

Acknowledgements

This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-GC-2019-001-2A). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

References

[Baytas *et al.*, 2017] I. Baytas, C. Xiao, X. Zhang, F. Wang, A. K. Jain, and J. Zhou. Patient subtyping via time-aware lstm networks. In *ACM SIGKDD*, 2017.

[Bhoi *et al.*, 2021] Suman Bhoi, M. L. Lee, W. Hsu, H. S. A. Fang, and N. C. Tan. Personalizing medication recommendation with a graph-based approach. *TOIS*, 40(3):1–23, 2021.

[Brody *et al.*, 2021] S. Brody, U. Alon, and E. Yahav. How attentive are graph attention networks? *arXiv:2105.14491*, 2021.

[Choi *et al.*, 2016] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *NeurIPS*, 2016.

[Darabi *et al.*, 2020] Sajad Darabi, Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. Taper: Time-aware patient ehr representation. *JBHI*, 24(11):3268–3275, 2020.

[Dornhorst *et al.*, 1985] Anne Dornhorst, StephenH Powell, and Jack Pensky. Aggravation by propranolol of hyperglycaemic effect of hydrochlorothiazide in type ii diabetics without alteration of insulin secretion. *The Lancet*, 325(8421):123–126, 1985.

[Fisher, 1992] Ronald Aylmer Fisher. Statistical methods for research workers. In *Breakthroughs in statistics*, pages 66–70. Springer, 1992.

[Jia *et al.*, 2020] Zheng Jia, Xian Zeng, Huilong Duan, Xudong Lu, and Haomin Li. A patient-similarity-based model for diagnostic prediction. *IJMI*, 135:104073, 2020.

[Johnson *et al.*, 2016] Alistair EW Johnson, Tom J Pollard, et al. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[Kang *et al.*, 2015] Seokho Kang, Pilsung Kang, Taehoon Ko, Sungzoon Cho, Su-jin Rhee, and Kyung-Sang Yu. An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction. *ESA*, 42(9):4265–4273, 2015.

[Kang *et al.*, 2017] Seokho Kang, Sungzoon Cho, Su-jin Rhee, and Kyung-Sang Yu. Reliable prediction of anti-diabetic drug failure using a reject option. *PAA*, 20(3):883–891, 2017.

[Kang, 2018] Seokho Kang. Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks. *AIM*, 85:1–6, 2018.

[Kenton and Toutanova, 2019] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186, 2019.

[Kuhn *et al.*, 2016] Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. The sider database of drugs and side effects. *Nucleic acids research*, 44(D1):D1075–D1079, 2016.

[Le *et al.*, 2018] Hung Le, Truyen Tran, and Svetha Venkatesh. Dual memory neural computer for asynchronous two-view sequential learning. In *SIGKDD*, pages 1637–1645, 2018.

[Li *et al.*, 2020] Yikuan Li, Shishir Rao, Jose Roberto Ayala Solares, et al. Behrt: transformer for electronic health records. *Scientific reports*, 10(1):1–12, 2020.

[Lu *et al.*, 2021] Chang Lu, Chandan K. Reddy, Prithwish Chakraborty, Samantha Kleinberg, and Yue Ning. Collaborative graph learning with auxiliary text for temporal event prediction in healthcare. In *IJCAI*, pages 3529–3535, 2021.

[Luo *et al.*, 2016] Yuan Luo, Peter Szolovits, Anand S Dighe, and Jason M Baron. Using machine learning to predict laboratory test results. *AJCP*, 145(6):778–788, 2016.

[Luo *et al.*, 2020] Junyu Luo, Muchao Ye, et al. Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In *SIGKDD*, pages 647–656, 2020.

[Ma *et al.*, 2017] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *SIGKDD*, pages 1903–1911, 2017.

[MacDonald *et al.*, 2010] Michael R MacDonald, Dean T Eurich, et al. Treatment of type 2 diabetes and outcomes in patients with heart failure: a nested case–control study from the uk general practice research database. *Diabetes care*, 33(6):1213–1218, 2010.

[Ni *et al.*, 2017] Jingchao Ni, Hongliang Fei, Wei Fan, and Xiang Zhang. Cross-network clustering and cluster ranking for medical diagnosis. In *ICDE*, pages 163–166. IEEE, 2017.

[Oei *et al.*, 2021] R. W. Oei, A. Fang, W. Y. Tan, , W. Hsu, M. L. Lee, and N. C. Tan. Using domain knowledge and data-driven insights for patient similarity analytics. *JPM*, 11(8), 2021.

[Prakash *et al.*, 2021] PKS Prakash, S. Chilukuri, N. Ranade, and S. Viswanathan. Rarebert: Transformer architecture for rare disease patient identification using administrative claims. In *AAAI*, volume 35, pages 453–460, 2021.

[Shang *et al.*, 2019] Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. Gamenet: Graph augmented memory networks for recommending medication combination. In *AAAI*, volume 33, pages 1126–1133, 2019.

[Sukhbaatar *et al.*, 2015] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. End-to-end memory networks. *NeurIPS*, 2015:2440–2448, 2015.

[Tan *et al.*, 2019] Qingxiong Tan, Andy Jinhua Ma, et al. Uacrnn: Uncertainty-aware convolutional recurrent neural network for mortality risk prediction. In *CIKM*, pages 109–118, 2019.

[Unnikrishnan *et al.*, 2012] Ranjit Unnikrishnan, Ranjit Mohan Anjana, and Viswanathan Mohan. Drugs affecting hba1c levels. *IJEM*, 16(4):528, 2012.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, et al. Attention is all you need. In *NeurIPS*, pages 6000–6010, 2017.

[Wang *et al.*, 2019] Shanshan Wang, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. Order-free medicine combination prediction with graph convolutional reinforcement learning. In *CIKM*, pages 1623–1632, 2019.

[Wei *et al.*, 2013] Wei-Qi Wei, Robert M Cronin, Hua Xu, Thomas A Lasko, Lisa Bastarache, and Joshua C Denny. Development and evaluation of an ensemble resource linking medications to their indications. *JAMIA*, 20(5):954–961, 2013.

[Wirbka *et al.*, 2020] Lucas Wirbka, Walter E Haefeli, and Andreas D Meid. A framework to build similarity-based cohorts for personalized treatment advice–a standardized, but flexible workflow with the r package simbaco. *PloS one*, 15(5):e0233686, 2020.