Integrating Medication Recommendation and Lab Test Response Prediction for Enhanced Clinical Decision Support

Suman Bhoi[1], **Mong Li Lee**[1], **Wynne Hsu**[1], **Ngiap Chuan Tan**[2]
[1] National University of Singapore
[2] SingHealth Polyclinics
sumanbhoi@u.nus.edu, leeml@comp.nus.edu.sg, whsu@comp.nus.edu.sg, tan.ngiap.chuan@singhealth.com.sg

Abstract

The rich data available in Electronic Health Records has led to the development of numerous systems for disease inference, mortality prediction, and personalized medication recommendations. However, integrating these diverse systems to enhance clinical decision-making in disease management remains an under-explored territory. This work describes a unified system that melds two distinct clinical tasks, namely medication recommendation and lab test response prediction, to bolster clinical decision support. This system could assist clinicians in personalizing dosage titrations and add-on medications. We present empirical studies from real-world datasets to demonstrate the potential of such a system.

Introduction

The proliferation of Electronic Health Records (EHR) and advancements in machine learning techniques have propelled healthcare analytics to the forefront. There is a growing body of research leveraging EHR data for diverse applications such as disease inference (Ni et al. 2017), mortality prediction (Tan et al. 2019) and personalized medication recommendation (Shang et al. 2019; Bhoi et al. 2021). Yet, there is a gap in efforts to unify these separate systems to enhance clinical decision-making in disease management.

Our research investigates the potential of employing an integrated system that combines two distinct clinical tasks – medication recommendation and lab test response prediction – to augment clinical decision support. Medication recommendation systems generate a combination of drugs for patients based on their medical history extracted from EHRs and other clinical knowledge bases like drug interaction databases. At the same time, lab test response prediction systems predict the patient's response to a specific lab test, taking into account patient data from EHRs and other resources such as drug-lab interaction databases.

We envision a cohesive system capable of recommending various medication combinations for a patient and predicting the corresponding lab test responses for these combinations. Such a system would allow clinicians to understand and anticipate the effectiveness of different treatment regimes. This capability would provide enhanced decision support for clinicians to personalize dosage titrations and add-on medications, leading to improved disease management.

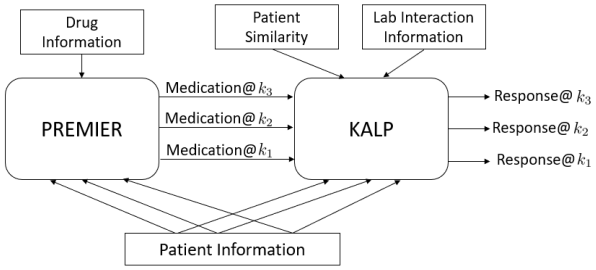

Figure 1: Overview of the combined system. Here k_1, k_2, k_3 depict the cut-off thresholds for PREMIER. Patient information consists of diagnosis, medications, and lab test responses for a patient over past visits.

In this work, we present evidence-based studies on two real-world datasets to demonstrate the utility of such a system. We discuss the challenges in developing these integrated systems, their current limitations, and assess the feasibility of implementing such systems in a real-world healthcare setting.

Related Work

Early works in medication recommendation learn a collection of rules from EHR. Solt and Tikk (Solt and Tikk 2009) extract rules from discharge summaries whereas Lakkaraju et al. (Lakkaraju and Rudin 2017) learn the mapping between the patient characteristics and treatments using Markov Decision Process. However, these approaches may introduce conflicting rules and are difficult to generalize and scale. Subsequent works employ recurrent neural networks (RNN) to model sequential dependency in patient's past visits and can be divided into two groups based on the inclusion or exclusion of drug interaction information while providing recommendations. The first group of works (Bajor and Lasko 2017; Le, Tran, and Venkatesh 2018) do not consider drug interactions to minimize adverse drug reactions. The second group of works (Shang et al. 2019; Wang et al. 2019) use Graph Convolutional networks (GCN) to learn and incorporate the drug interaction information to provide safe drug recommendations. These systems do not learn the rel-

evance of the information of the past visits to the current visit and are not able to generate any justifications for the recommendations. In contrast, PREMIER (Bhoi et al. 2021) uses neural attention based patient and graph attention based drug interaction modeling to learn the relevance of different information sources of past visits. Moreover, PREMIER is capable of providing justifications for the recommendations by using the attention weights obtained from the system.

Existing research in lab test response prediction uses patient-specific information such as demographics and past visit records to make predictions (Luo et al. 2016; Kang 2018). The work in (Luo et al. 2016) predicts Ferritin lab test result by using patient demographics and the results of other lab tests that have been ordered at the same time as the Ferritin lab test. In (Kang 2018), the authors examine the task of predicting HbA1c test results, and propose a recurrent neural network (RNN) based architecture to model the sequential dependencies across the past visits. However, the impact of medications or diagnosis on the target lab test result is not considered. For example, patients with high blood pressure are prescribed medications like Propranolol which is known to increase the HbA1c, a lab test typically performed to estimate the blood sugar levels (Dornhorst, Powell, and Pensky 1985). Similarly, HbA1c is increased by iron deficiency anemia and decreased by hemolytic anemia[1]. Recently, KALP (Bhoi et al. 2022) proposes a lab test response prediction system that considers patient history along with the impact of medications or diagnosis on a target lab test response to make predictions. They adapt transformer and graph attention networks to model sequential dependency in the diagnosis, medication dosages over multiple patient visits and impact of medications/diseases on a lab test response respectively. They further boost the performance of their system by leveraging on patient similarity analytics based on patient demographics and diagnosis.

All the works that we discuss above perform a single task i.e., recommend medications or predict lab test responses. However, no research exists to understand if we can use these systems in conjunction to supplement each other and enhance the decision support for clinicians to help in improving disease management.

Proposed System

Figure 1 gives an overview of the proposed system comprising of a medication recommendation system PREMIER (Bhoi et al. 2021) and a lab test response prediction system KALP (Bhoi et al. 2022).

PREMIER is a two-stage recommender system designed to generate personalized medications with minimal adverse drug interactions. The first stage utilizes a two-level neural attention mechanism to model patient-specific information such as diagnosis, procedures and previously prescribed medications as a query vector. This vector captures the relative importance of the sequential dependencies across multiple visits. In the second stage, information about potential drug interactions and drug co-occurrences are retrieved from the drug interaction repository and the EHR by using the query vector. This stage ensures that the recommended medication set has minimal drug interactions.

KALP provides a personalized prediction of patient's response to a target lab test while considering drug-lab interactions, diagnosis-lab interactions, medication dosages, as well as past lab test responses. It uses transformer encoder to capture the patient-specific information, while the information of similar patients is modeled using the modified graph attention network (GATv2) (Brody, Alon, and Yahav 2021). With this, KALP obtains a strong latent patient representation which incorporates fine-grained dosage information to accurately predict patient response to a target lab test, even amidst medication titrations. Further, KALP models the complex relationships between drugs, co-morbidities, and lab test results by augmenting the patient representation with the knowledge of drug-lab interactions and diagnosis-lab interactions.

Our goal is to take the medication combinations suggested by PREMIER as inputs to KALP, which in turn predicts the potential outcomes of specific lab tests for a patient upon taking these suggested medications. However, PREMIER recommends medications at the drug class level (e.g., class A10B which encompasses drugs like Metformin, Glipizide, etc.) whereas KALP requires fine-grained medication information i.e., drug names with dosage information, several modifications to PREMIER are required.

First, we align the patient data used in both systems by replacing the procedure information used in PREMIER with lab test information. Second, we refine PREMIER's vocabulary to specific drug names instead of the broader medication class. Third, we substitute the medication class-based drug interaction knowledge base TWOSIDES (Tatonetti et al. 2012) with the drug name-based knowledge base, DDInter (Xiong et al. 2022) to achieve greater specificity. Finally, we adjust the dimensions of the final layer in PREMIER to compute the probabilities of recommending the individual drugs in the dataset.

To enable PREMIER to produce different sets of recommended medications, we apply different cut-off values to the probabilities at its final layer. Medications with probability values above the cut-off are considered for recommendation. The sets of recommended medications, generated at different cut-offs, are then passed to KALP as shown in Figure 1.

For accurate lab test response prediction, KALP requires dosage information for the prescribed drugs. However, PREMIER only recommends drug names without providing the dosage. To bridge this gap between the two systems, we use the last recorded dosage in the EHR if the recommended medication has been prescribed to the patient before. If the recommended medication is new to the patient, we use the most commonly prescribed dosage for that medication in the EHR.

Evidence-based Study

We implemented the combined system in PyTorch and trained the models on two NVIDIA Titan RTX GPUs. We use the following two datasets for our analysis:

- MIMIC-III (Johnson et al. 2016) This is the largest pub-

[1]https://tinyurl.com/nhfzawse

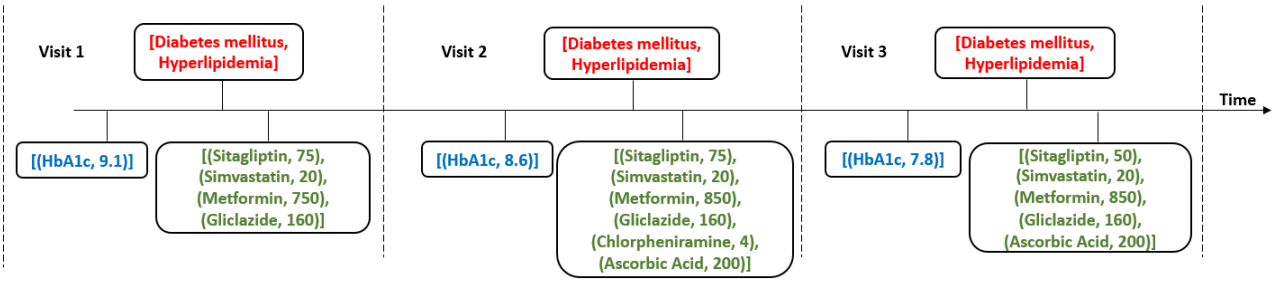

Figure 2: EHR of a patient with three visits. Information sources are color-coded. Green indicates medication with dosage information, red indicates diagnosis, and blue indicates lab test response.

Attribute	MIMIC-III	PRIVATE
Number of patients	6412	6312
Number of diagnosis	1919	139
Number of medications	2941	45
Average number of visits per patient	2.67	13.06
Average number of diagnosis per visit	13.47	5.90
Average number of medications per visit	32	4.07
Maximum number of diagnosis per visit	39	27
Maximum number of medications per visit	148	12

(a) Dataset statistics.

Variable	Type	MIMIC-III	PRIVATE
Age	Discrete	65.36 (13.27)	58.45 (15.67)
Gender	Categorical	4136 (M), 2276 (F)	3324 (M), 2988 (F)
Weight	Continuous	66.23 (15.25)	71.45 (14.12)
HbA1c	Continuous	9.2 (3.9)	10.5 (4.2)

(b) Patient characteristics.

Table 1: Summary of datasets.

licly available EHR dataset which contains clinical data for 7870 neonates and 38,597 adults admitted to ICU between 2001 and 2008.

- PRIVATE. This is a proprietary outpatient dataset that includes a wide array of patient information such as demographics, vital signs, results of blood tests, and details of prescribed medications over a period from 2010 to 2019.

Both MIMIC-III and PRIVATE use the ICD-9 coding system for diagnosis and the generic names for medications. Since the dosages are reported in different units, we standardize and convert all the dosage values to milligrams. Table 1(a) summarizes the statistics of these datasets. The characteristics of the patients in these datasets are given in Table 1(b). Here, we report the average age, gender, and HbA1c along with their standard deviation in parenthesis. For gender, we report the count of male (M) and female (F) patients. Compared to MIMIC III, the average number of diagnosis, and medications per visit in PRIVATE is fewer. We use HbA1c as the target lab test. Figure 2 shows a sample of the visit information extracted from a patient's EHR.

We separately train the models in PREMIER and KALP for the tasks of medication recommendation and lab test response prediction respectively and use the trained models in our combined system. The embedding size for the models is fixed at 128 and training is done using the Adam optimizer. The learning rate is 0.0002, and the best-performing model is chosen after 50 epochs. We apply dropout of 0.4 and 0.6 on the input embedding layer for MIMIC-III and PRIVATE respectively.

Sample Patient A from PRIVATE

We analyze the medical records of a sample Patient A from the PRIVATE dataset, who has been diagnosed with Diabetes Mellitus, Hyperlipidemia, and Hypertension over two visits. The actual HbA1c values of this patient fluctuate between 7.5 and 8.8 as depicted in Figure 3.

The recommended sets of medications at different cut-offs are shown together with the corresponding predicted HbA1c value. We see that cut-off values of 0.4 and 0.6 generally result in higher and lower number of prescribed medications respectively. KALP's predictions suggest that the HbA1c levels could be 8.9, 8.3, and 7.7, respectively, for the medication sets recommended by PREMIER at cut-offs of 0.4, 0.5, and 0.6.

From these predictions, the clinician might opt for the medication set at the 0.6 cut-off. Notably, this selection aligns with the actual medication regimen prescribed in the next visit, leading to an actual HbA1c value of 7.5.

This case highlights the potential of our integrated approach to assist clinicians to predict patients' lab test re-

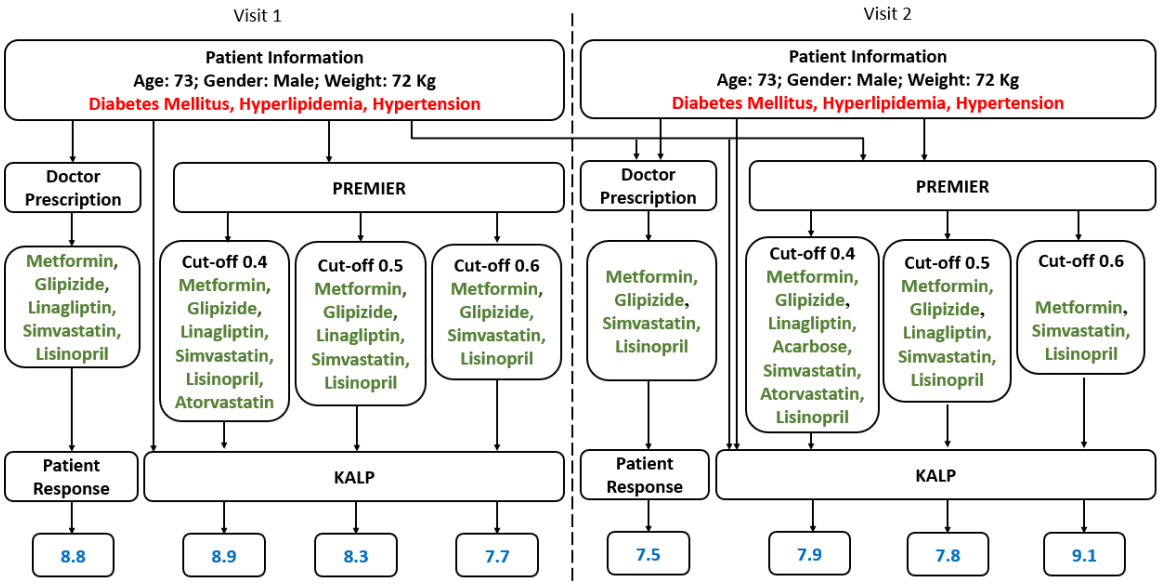

Figure 3: Predicted and actual HbA1c response for Patient A with medication input generated from PREMIER. Cut-off 0.4 denotes the medication set generated by PREMIER at a threshold of 0.4. The information sources are color coded as red, green, blue for diagnosis, medication, lab test response respectively.

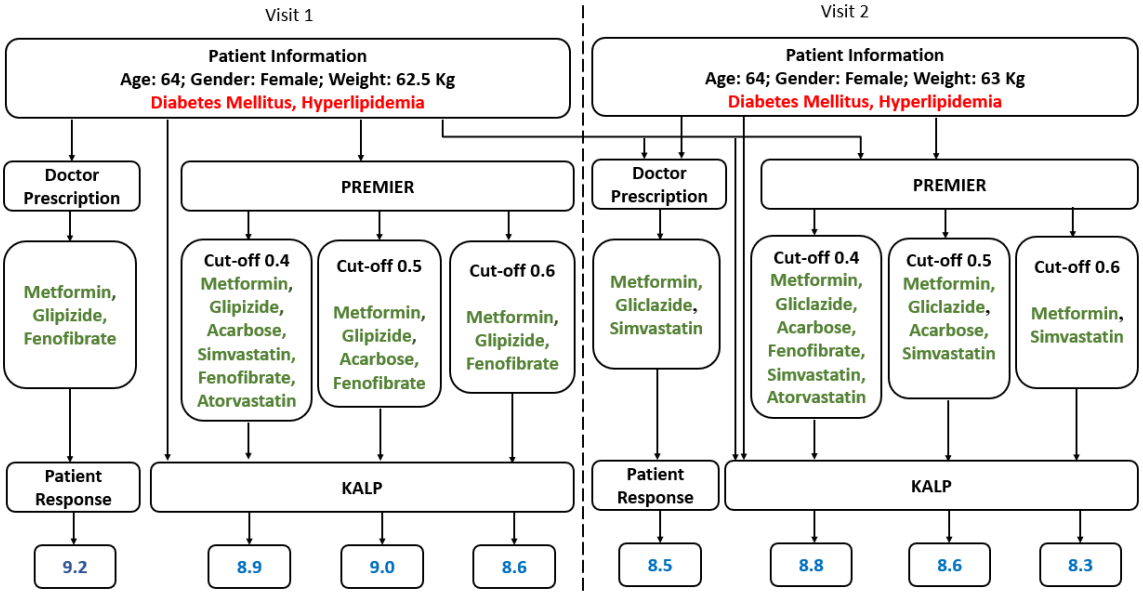

Figure 4: Predicted and actual HbA1c response for Patient B with medication input generated from PREMIER. Cut-off 0.4 denotes the medication set generated by PREMIER at a threshold of 0.4. The information sources are color coded as red, green, blue for diagnosis, medication, lab test response respectively.

sponses for different medication regimens. This in turn enables more personalized treatment strategies to improve the management of various health conditions.

Sample Patient B from PRIVATE

Sample Patient B has received diagnoses of Diabetes Mellitus and Hyperlipidemia across two visits and has an HbA1c of 8.7 at baseline. Figure 4 shows the actual prescribed medications across two visits, as well as the sets of medications recommended by PREMIER at two cut-off values. The corresponding actual and predicted HbA1c is also shown.

We observe that when the patient was actually administered a mixture of Metformin, Glipizide, and Fenofibrate, the HbA1c level increases to 9.2. However, KALP predicts it to be 8.6, a slight decrease from the baseline HbA1c. A closer inspection of the prescribed and recommended medications

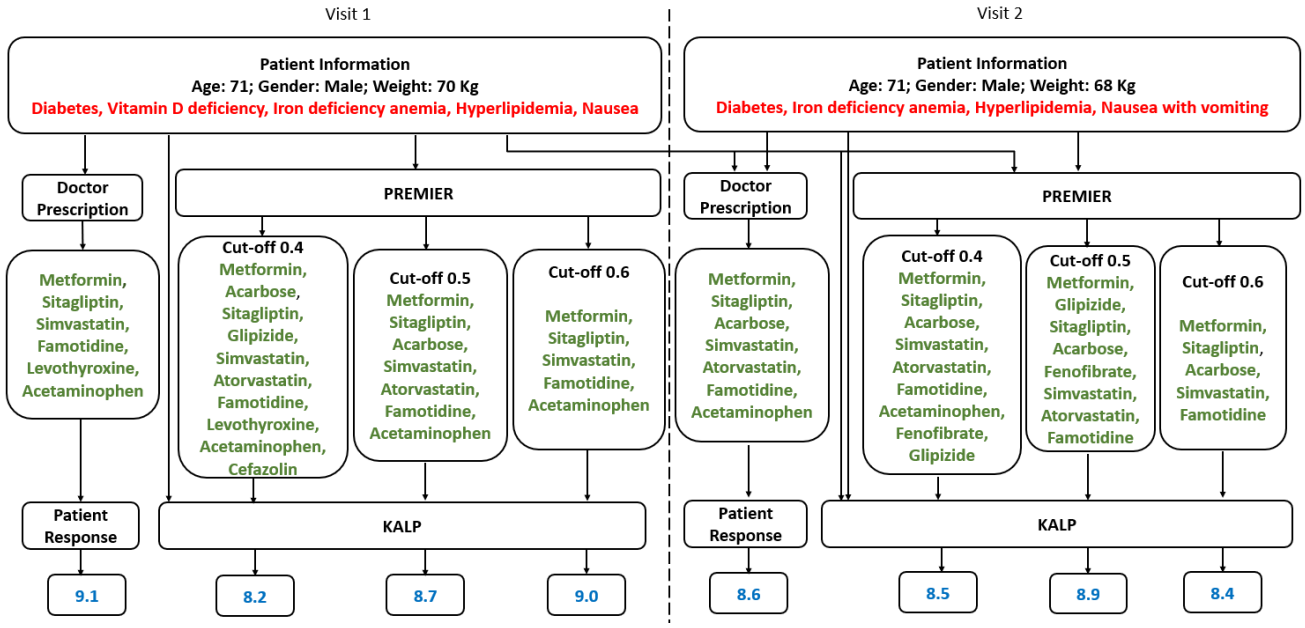

Figure 5: Predicted and actual HbA1c response for Patient C with medication input generated from PREMIER. Cut-off 0.4 denotes the medication set generated by PREMIER at a threshold of 0.4. The information sources are color coded as red, green, blue for diagnosis, medication, lab test response respectively.

reveals that the dosages of the prescribed medications were of a lower value as compared to the dosages of the recommended medications. This discrepancy arises because PREMIER does not include dosage information in the recommendations, and we have assigned the most frequently prescribed dosages for the medications, which were higher than the actual administered dosages. This indicates the need for more fine-grained medication recommenders that includes dosage along with the medications.

Sample Patient C from MIMIC-III

Finally, we examine the lab test response generated by our integrated system for a patient extracted from the MIMIC-III dataset. As shown in Figure 5, this patient has been diagnosed with Diabetes, Hyperlipidemia, Nausea, Vitamin D and Iron deficiency. The HbA1c levels for this patient vary from 8.6 to 9.1.

We observe that the predicted HbA1c response to the recommended medication set at a cut-off threshold of 0.5 is 8.7 on the first visit. It is interesting to see that a medication combination similar to that recommended in the first visit at a threshold of 0.5 is actually prescribed on the next visit leading to an HbA1c response of 8.6 which is similar to the response predicted by our integrated system. This shows that the integrated system discussed in this work can help clinicians to anticipate the personalized response of a patient to a combination of medications.

Conclusion

In this work, we have investigated the potential of an integrated system consisting of a medication recommendation system and a lab test response prediction system to enhance clinical decision support for managing the health conditions of patients. While the initial results are promising, there is still room for improvement. Currently, KALP employs the set of medications suggested by PREMIER as input, defaulting to the most recently prescribed or commonly administered dosage of these medications. This may not be the most effective approach to assigning dosage. A more fine-grained recommender system, suggesting a personalized set of medications along with their corresponding dosages, is needed.

Future work includes broadening the system by incorporating data from a range of sources such as clinician notes and physiological data to build a more comprehensive clinical decision support system. Furthermore, providing clear justifications for the generated recommendations is vital, as it enables deeper insights and fosters trust among clinicians using the system.

Acknowledgments

This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-GC-2019-001-2A).

References

Bajor, J. M.; and Lasko, T. A. 2017. Predicting medications from diagnostic codes with recurrent neural networks. In *ICLR*.

Bhoi, S.; Lee, M. L.; Hsu, W.; Fang, H. S. A.; and Chuan Tan, N. 2022. Chronic Disease Management with Personalized Lab Test Response Prediction. In *Proceedings of the Thirty-First International Joint Conference on Artifi-*

cial Intelligence, IJCAI-22, 5038–5044. International Joint Conferences on Artificial Intelligence Organization.

Bhoi, S.; Lee, M. L.; Hsu, W.; Fang, H. S. A.; and Tan, N. C. 2021. Personalizing Medication Recommendation with a Graph-Based Approach. *TOIS*, 40(3): 1–23.

Brody, S.; Alon, U.; and Yahav, E. 2021. How Attentive are Graph Attention Networks? *arXiv:2105.14491*.

Dornhorst, A.; Powell, S.; and Pensky, J. 1985. Aggravation by propranolol of hyperglycaemic effect of hydrochlorothiazide in type II diabetics without alteration of insulin secretion. *The Lancet*, 325(8421).

Johnson, A. E.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1): 1–9.

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

Lakkaraju, H.; and Rudin, C. 2017. Learning Cost-Effective and Interpretable Treatment Regimes. In Singh, A.; and Zhu, J., eds., *AISTATS*, volume 54 of *Proceedings of Machine Learning Research*, 166–175. PMLR.

Le, H.; Tran, T.; and Venkatesh, S. 2018. Dual memory neural computer for asynchronous two-view sequential learning. In *SIGKDD*, 1637–1645. ACM.

Luo, Y.; Szolovits, P.; Dighe, A. S.; and Baron, J. M. 2016. Using machine learning to predict laboratory test results. *AJCP*, 145(6).

Ni, J.; Fei, H.; Fan, W.; and Zhang, X. 2017. Cross-network clustering and cluster ranking for medical diagnosis. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 163–166. IEEE.

Shang, J.; Xiao, C.; Ma, T.; Li, H.; and Sun, J. 2019. Gamenet: Graph augmented memory networks for recommending medication combination. In *AAAI*, volume 33, 1126–1133.

Solt, I.; and Tikk, D. 2009. Yet another rule-based approach for extracting medication information from discharge summaries. In *Proceedings of the third i2b2 workshop on challenges in natural language processing for clinical data*.

Tan, Q.; Ma, A. J.; Ye, M.; Yang, B.; Deng, H.; Wong, V. W.-S.; Tse, Y.-K.; Yip, T. C.-F.; Wong, G. L.-H.; and Ching, J. Y.-L. 2019. Ua-crnn: Uncertainty-aware convolutional recurrent neural network for mortality risk prediction. In *ACM CIKM*, 109–118.

Tatonetti, N. P.; Patrick, P. Y.; Daneshjou, R.; and Altman, R. B. 2012. Data-driven prediction of drug effects and interactions. *Science translational medicine*, 4(125): 125ra31–125ra31.

Wang, S.; Ren, P.; Chen, Z.; Ren, Z.; Ma, J.; and de Rijke, M. 2019. Order-free Medicine Combination Prediction With Graph Convolutional Reinforcement Learning. In *CIKM 2019*.

Xiong, G.; Yang, Z.; Yi, J.; Wang, N.; Wang, L.; Zhu, H.; Wu, C.; Lu, A.; Chen, X.; Liu, S.; et al. 2022. DDInter: an online drug–drug interaction database towards improving clinical decision-making and patient safety. *Nucleic acids research*, 50(D1): D1200–D1207.