Predicting Outcomes for Cancer Patients with Transformer-based Multi-task Learning

Leah Gerrard[1,2], Xueping Peng[1], Allison Clarke[2], Clement Schlegel[2] and Jing Jiang[1]

1. Australian Artificial Intelligence Institute, Faculty of Engineering and IT, University of Technology Sydney.
2. Australian Government Department of Health.

INTRODUCTION

Predicting outcomes for cancer patients is an ongoing challenge due to the complexity of individuals with caner and growing treatment options[1]. Cancer patients often experience frequent hospital admissions due to cancer symptoms and treatment, resulting in possible delays in therapy and reduced quality of life[2]. Accurately predicting outcomes for cancer patients is therefore crucial in providing personalised care and improving patient outcomes.

Existing deep learning models with Electronic Health Record (EHR) data have primarily focused on Recurrent Neural Network (RNN) models. However there is growing interest in the use of Transformer models which may be able to produce better predictions than RNNs. Furthermore, existing models are primarily based on single-task predictions, and increasing evidence suggests that multi-task learning, where multiple tasks are predicted simultaneously, may offer performance improvements for related tasks. To address these limitations, this research proposes a Transformer-based multi-task model to predict multiple outcomes for hospitalised cancer patients.

AIM

To develop a multi-task Transformer model for EHR data and predict the following two outcomes for cancer patients: 1) **Hospital readmission** – defined as readmission within 30 days of the previous admission, and 2) **Future diagnosis** – defined as the principal diagnosis or primary reason for hospitalisation in the next visit.

METHODS

Model Overview

Each patient's EHR is made up of a series of visits, which contain one or more diagnosis codes (Fig. 1). An embedding layer encodes categorical diagnosis codes to dense numerical vectors. Then, an attention pooling layer compresses a set of diagnosis code embeddings from the visit into a single context-aware vector representation. The position embeddings are added to the learned visit vectors, and normalised outputs are fed into the prediction tasks. The structure of the two prediction tasks are identical, using a Transformer to learn the visit relationships in the patient record. Lastly, a prediction model is used to predict the outcomes in the final hospital visit.

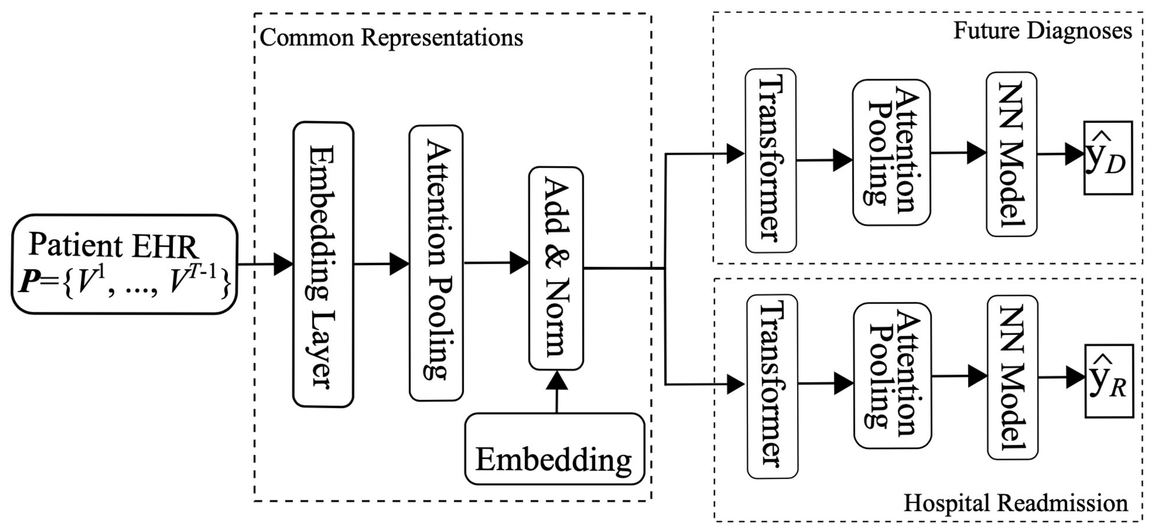

Fig. 1. The Proposed TransMT Model.

Transformer Model

The Transformer is identical to that of BERT[3] (Bidirectional Encoder Representations from Transformers).

Multi-task Learning

Multi-task learning is employed to jointly learn the two prediction tasks for the final patient visit. Given a patient's visit, we simultaneously predict future diagnosis and hospital readmission. To allow shared information between the two prediction tasks, we jointly train the tasks using a weighted loss function.

References

1. Gupta, S. et al. (2014), Machine-learning prediction of cancer survival: a retrospective study using elec- tronic administrative records and a cancer registry. BMJ Open 4(3).
2. Fadol, A. et al. (2019), A quality improvement approach to reducing hospital readmissions in patients with cancer and heart failure. Cardio-Oncology 5, 5.
3. Devlin,J. et al. (2018), Bert:Pre-trainingofdeepbidirec- tional transformers for language understanding. arXiv preprint arXiv:1810.04805.

EXPERIMENTS

Datasets

- Medical Information Mart for Intensive Care (MIMIC)-III and MIMIC-IV.
- Publicly available EHR data.

Cancer Patient Identification

- At least one cancer-related International Classification of Disease (ICD)-9 diagnosis code of 140 – 239.

CCS Mapping

- Clinical Classifications Software (CCS) mapping was used to reduce the number of diagnosis code categories for model training to 285 CCS categories.

Table 2. Statistics of the MIMIC datasets for cancer patients.

Dataset	MIMIC-III	MIMIC-IV
# of patients	2,070	20,953
# of visits	5,552	79,177
Avg. # of visits per patient	2.68	3.78
# of unique ICD9 codes	3,228	6,939
Avg. # of ICD9 codes per visit	13.67	13.47
Max # of ICD9 codes per visit	39	57
# of principal category	166	270

RESULTS

To validate the proposed Transformer-based multi-task model (TransMT), we compare the performance with two common RNN baselines (RETAIN and Dipole), and with a single-task learning (STL) Transformer model.

TransMT outperforms all baseline models on the prediction tasks for both datasets based on average F1 score and Area Under the Receiver Operating Characteristic (ROC) Curve (Table 3. and Fig. 2). This indicates the benefit of using multi-task learning to jointly learn the prediction tasks. The single-task learning Transformer model also outperforms the RNN baselines, supporting the use of Transformers over RNNs for EHR prediction problems.

Table 3. Performance comparison of prediction tasks.

Dataset	Model	F1 Score (%)	
		Future diagnosis	Hospital readmission
MIMIC-III	RETAIN	12.48 ± 2.35	9.65 ± 5.67
	Dipole	10.39 ± 0.49	19.14 ± 3.85
	STL	16.19 ± 1.05	24.27 ± 4.46
	TransMT	**16.56** ± 1.34	**24.70** ± 3.95
MIMIC-IV	RETAIN	7.60 ± 0.54	35.12 ± 1.64
	Dipole	6.99 ± 0.23	38.64 ± 6.03
	STL	8.97 ± 0.43	34.81 ± 1.24
	TransMT	**9.04** ± 0.31	**40.34** ± 2.32

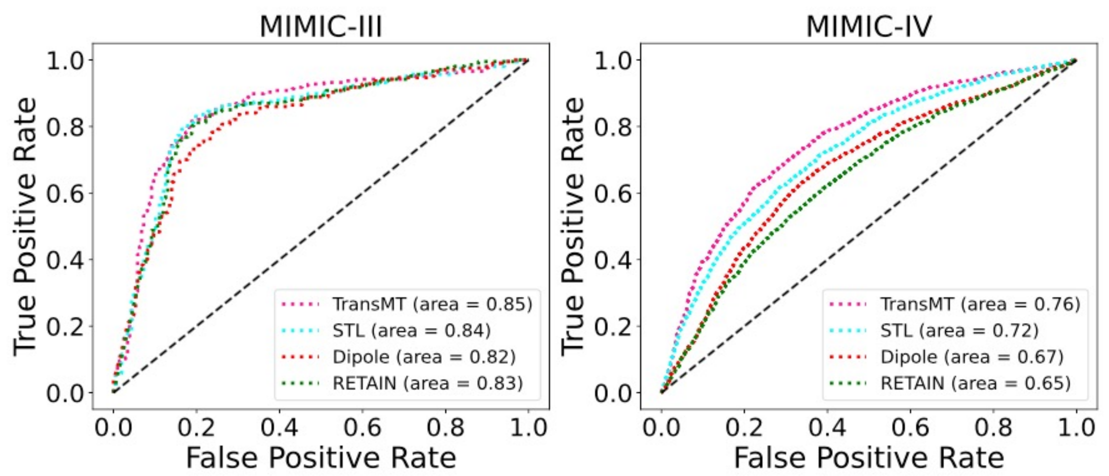

Fig. 2. ROC of hospital readmission on two datasets.

CONCLUSIONS

We proposed TransMT to capture the relationship between patient visits in EHR to predict future diagnosis and hospital readmission for cancer patients. As demonstrated by the experiments, TransMT produces better predictions than single-task Transformer and RNN baselines.

The use of other cancer datasets, additional patient data (e.g. medications, demographics), and use of pre-trained models are potential future avenues to further improve predictions from EHR data and for cancer-related outcomes.

Acknowledgments

- **Australian Government Department of Health** – we thank the department for supporting this work through an industry-based PhD program.
- This research is also supported by an Australian Government Research Training Program Scholarship.