ARIA
APPLIED RESEARCH IN ACTION

Using Multi-Modal Data and Self-Supervised Approaches for Deep Learning in Healthcare

Transformers that simultaneously learn from time series and text data via self-supervised learning, thereby improving clinical predictions

Shixuan Li

Bo Wang
ACADEMIC SUPERVISOR

Felipe Perez
INDUSTRY SUPERVISOR

Model	Mortality		Phenotyping		Transfer to ICU	
	ROC-AUC	PR-AUC	ROC-AUC	PR-AUC	ROC-AUC	PR-AUC
XGBoost	0.886	0.593	0.829	0.589	0.833	0.446
LSTM	0.881	0.533	0.756	0.447	0.777	0.327
mTAND	0.864	0.540	0.812	0.553	0.817	0.398
Raindrop	0.878	0.546	0.824	0.577	0.821	0.413
STraTS	0.882	0.552	0.820	0.565	0.789	0.329
DuETT	**0.912**	0.627	0.838	0.604	0.841	0.467
DuETT (reproduced)	0.895	0.611	0.839	0.591	0.837	0.454
RadBERT-MLP	0.744	0.284	0.764	0.493	0.614	0.242
With Early Fusion	0.898	**0.631**	**0.852**	**0.618**	**0.851**	**0.505**
With Late Fusion	0.898	0.618	0.842	0.596	0.848	0.497

Table 1. Performance on tasks using the MIMIC-IV dataset. The baseline scores are cited from the original DuETT paper.

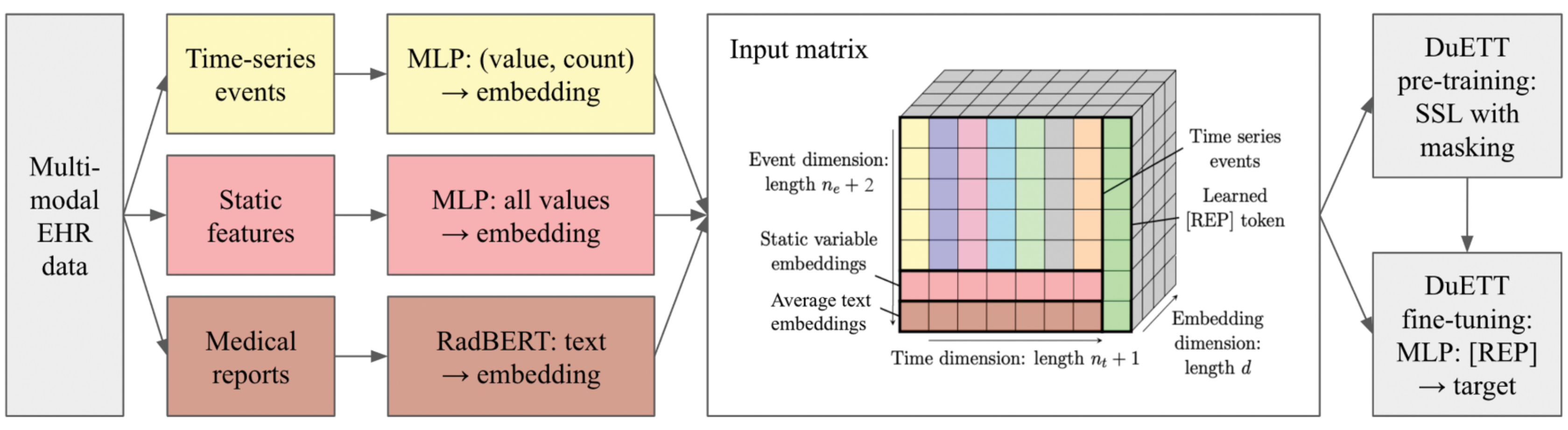
Multi-modal EHR data
Time-series events
Static features
Medical reports
MLP: (value, count) → embedding
MLP: all values → embedding
RadBERT: text → embedding
Input matrix
Event dimension: length $n_e + 2$
Static variable embeddings
Average text embeddings
Time dimension: length $n_t + 1$
Time series events
Learned [REP] token
Embedding dimension: length d
DuETT pre-training: SSL with masking
DuETT fine-tuning: MLP: [REP] → target

Figure 1. The multi-modal learning pipeline. A MLP is trained to create a *d*-dimensional embedding from (value, count) pairs, which are the last value and the number of occurrences of an event in a time bin. A second MLP is trained to create a *d*-dimensional embedding using all static features of a patient, and the same embedding is used across all time bins. The embeddings created by RadBERT from all medical reports are averaged and used across the time bins. The constructed matrix is used in the pre-training stage which applies masking and in the fine-tuning stage which uses learned [REP] embeddings to predict targets.

PROJECT SUMMARY

With the recent advances in deep learning, this study aims to effectively leverage multi-modal health data that includes static features, time series records, and medical notes. This study uses state-of-the-art transformer-based models and self-supervised approaches to learn latent patient representations from the abundant unlabeled data. These representations are then fine-tuned for specific downstream tasks. This study extends and improves a state-of-the-art model named DuETT (Dual Event Time Transformer) [1] by including text embeddings generated by RadBERT [2]. We develop different fusion techniques to create a pipeline that ingests both tabular and text data. Using the popular MIMIC-IV benchmark [3], we show that by incorporating text data, the model's performance can be significantly improved in different tasks. Our pipeline achieves 0.85 ROC-AUC and 0.62 PR-AUC in the phenotyping task and 0.85 ROC-AUC and 0.51 PR-AUC in the transfer to ICU prediction task, outperforming a number of baseline models. This study demonstrates the effectiveness of multi-modality learning and the application of deep learning in healthcare. Additionally, the DuETT pipeline has been trained and validated using data from St. Michael's Hospital. It outperforms the baseline model, showing its applicability to real data.

REFERENCES

[1] Alex Labach, Aslesha Pokhrel, Xiao Shi Huang, Saba Zuberi, Seung Eun Yi, Maksims Volkovs, Tomi Poutanen, and Rahul G Krishnan. Duett: Dual event time transformer for electronic health records. arXiv preprint arXiv:2304.13017, 2023

[2] An Yan, Julian McAuley, Xing Lu, Jiang Du, Eric Y Chang, Amilcare Gentili, and Chun-Nan Hsu. Radbert: Adapting transformer-based language models to radiology. Radiology: Artificial Intelligence, 4(4):e210258, 2022.

[3] Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. Mimic-iv (version 2.2). PhysioNet, 2023.

SIGNAL 1

Computer Science
UNIVERSITY OF TORONTO

Master of Science in
Applied Computing