Deep Neural Learning for Automated Diagnostic Code Group Prediction Using Unstructured Nursing Notes

Aditya Jayasimha*, Tushaar Gangavarapu*, Sowmya Kamath S, and Gokul S Krishnan

HALE Lab

* Equal contribution

ICD-9 Code Group Prediction

ICD-9 codes are a widely employed medical coding ontology by healthcare professionals.

Risk Assessment

Manifold nature: Mapping each patient record (nursing note) to multiple ICD-9 code groups.

- Manual coding vs. automated prediction?
- Risk stratification as a multi-label problem.
- Use: epidemiology and insurance policies.

Granularity of Codes

Unique ICD-9 code per symptom vs. category-level prediction vs. group-level prediction.

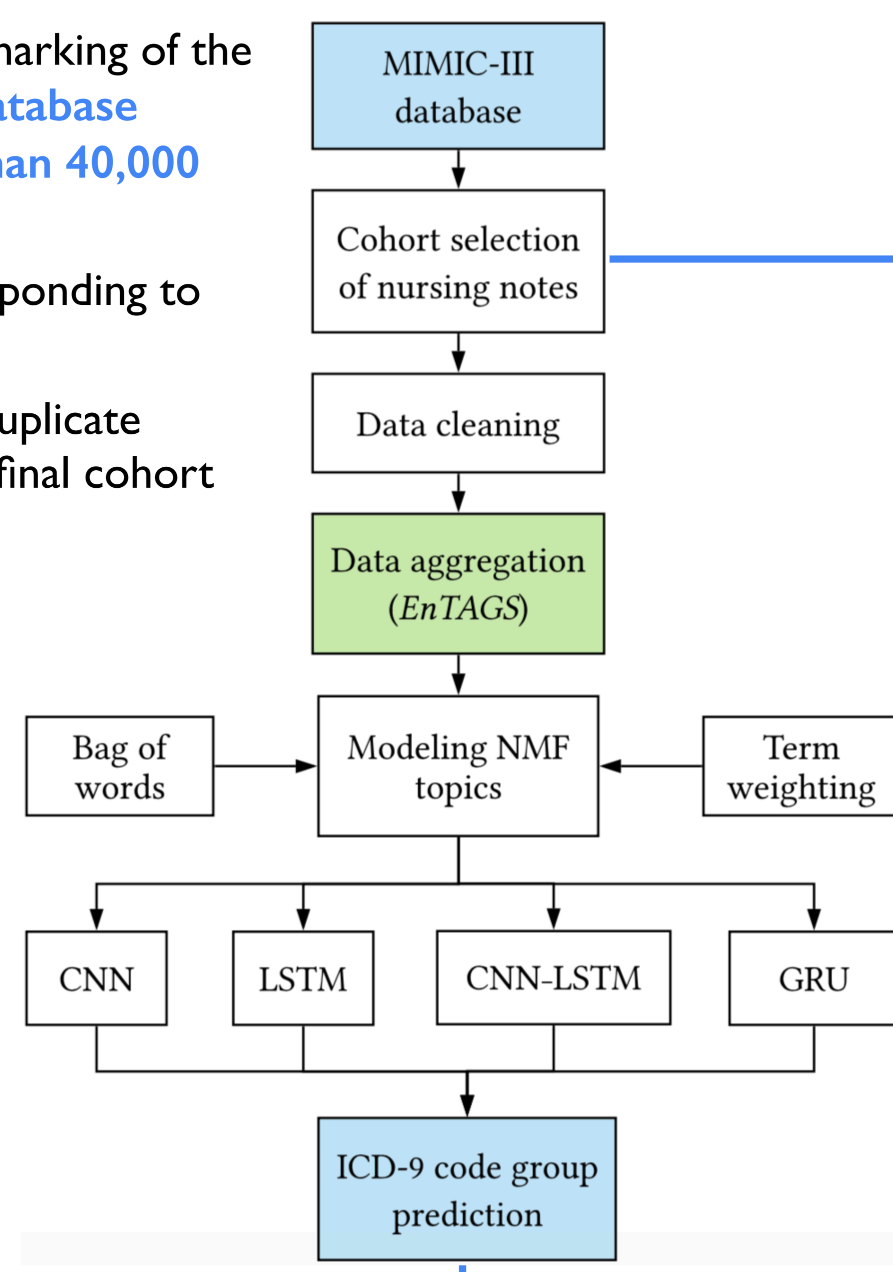
- ICD-9 taxonomy: over 13,000 codes!
- Computational complexity vs. granularity.
- Group-level: state-of-the-art works.

We focus on risk stratification as a multi-label code group prediction problem

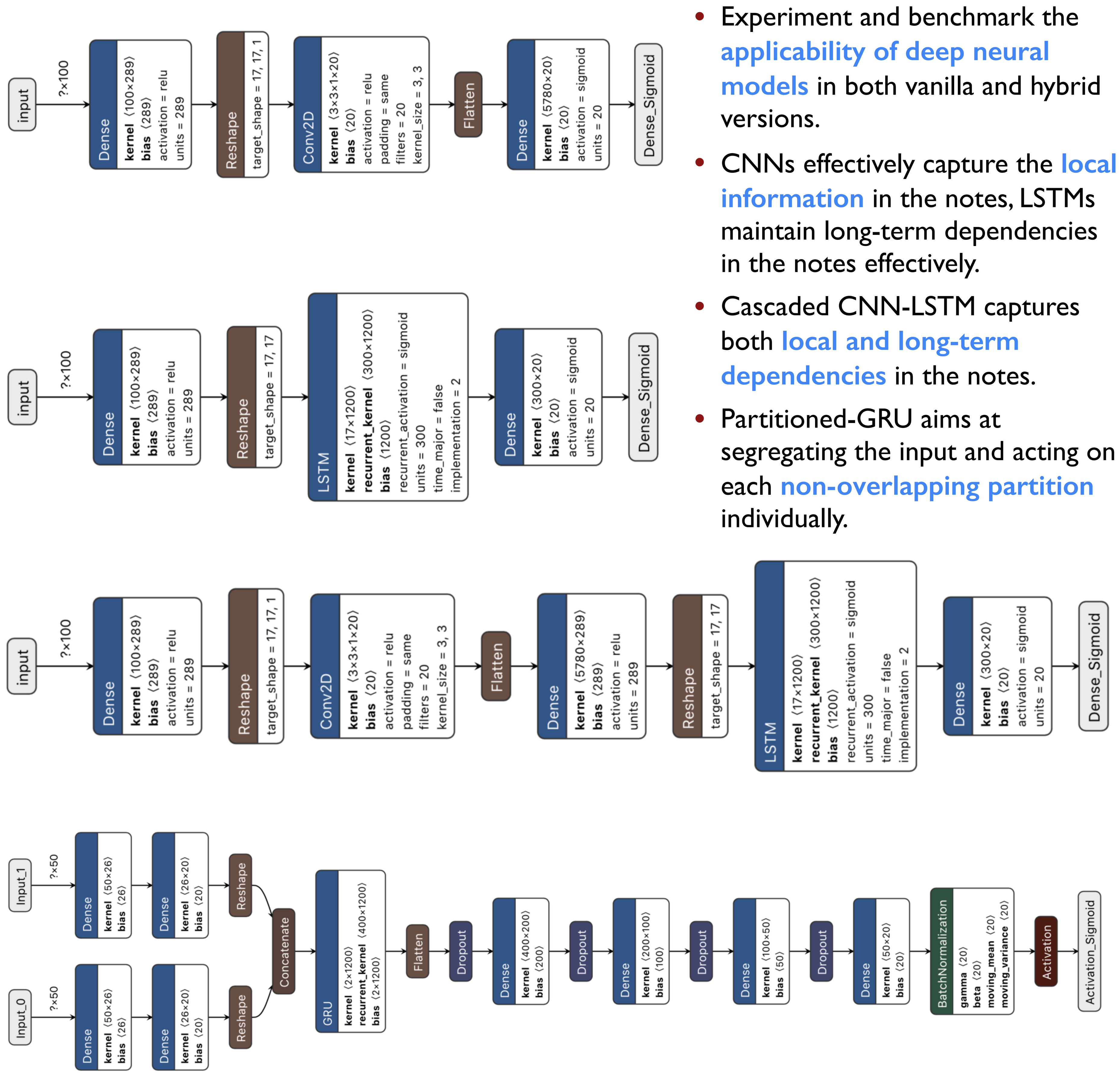
EnTAGS: Deep Neural Learning

- To facilitate experimental validation and benchmarking of the proposed strategy, we utilized the MIMIC-III database comprising of of diverse health data of more than 40,000 Intensive Care Unit (ICU) patients.
- Patient cohort: consists of nursing notes corresponding to 7,638 patients, with a median age of 66 years.
- Data cleaning: faulty and inconsistent records, duplicate entries, incorrect code mapping, and others — final cohort of 6,532 patients.
- EnTAGS for aggregation by independently modeling the nursing notes of a patient recorded over time, followed by NMF-TW modeling.
- Deep neural learning for automated diagnostic code group prediction: CNN, LSTM, cascaded CNN-LSTM, and partitioned GRU.

Parameter	Total
Clinical nursing notes	223, 556
Sentences in the clinical nursing notes	5, 244, 541
Words in the clinical nursing notes	79, 988, 065
Unique words in the clinical nursing notes	715, 821



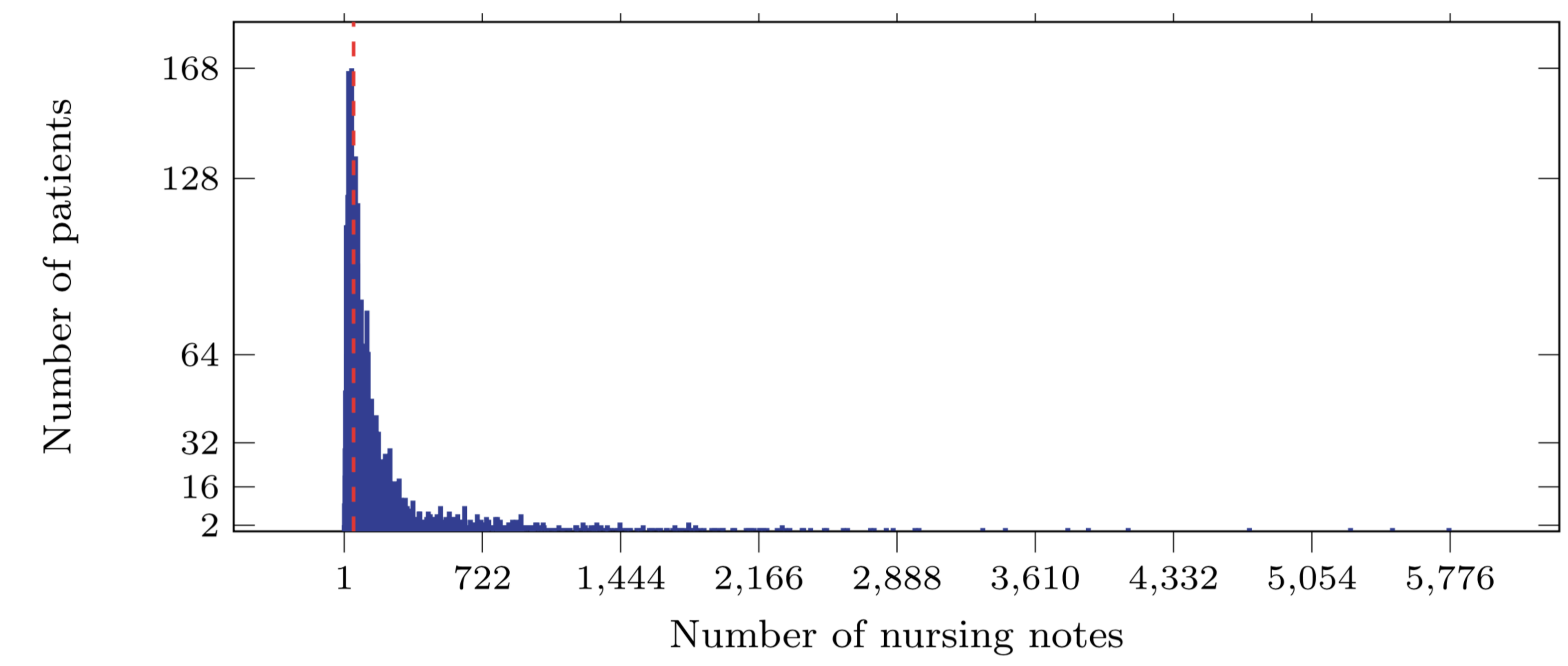
Neural Architectures for Diagnostic Code Group Prediction



- Experiment and benchmark the applicability of deep neural models in both vanilla and hybrid versions.
- CNNs effectively capture the local information in the notes, LSTMs maintain long-term dependencies in the notes effectively.
- Cascaded CNN-LSTM captures both local and long-term dependencies in the notes.
- Partitioned-GRU aims at segregating the input and acting on each non-overlapping partition individually.

Feature modeling using NMF eliminates the structure, so why CNN-LSTM?

Modeling Nursing Notes for Intelligent Clinical Decision Support



- Clinical nursing notes maintain objective and subjective assessments of a patient's condition — can be utilized to uncover hidden clues about the mental state of a patient.
- Modeling nursing notes is challenging due to their high-dimensionality, rawness, sparsity, complex linguistic and temporal nature, inconsistent abbreviations, and occurrence of rich medical jargon.

Eliminating dependency on structured EHRs is essential in developing nations

Experiments: Patient Cohort

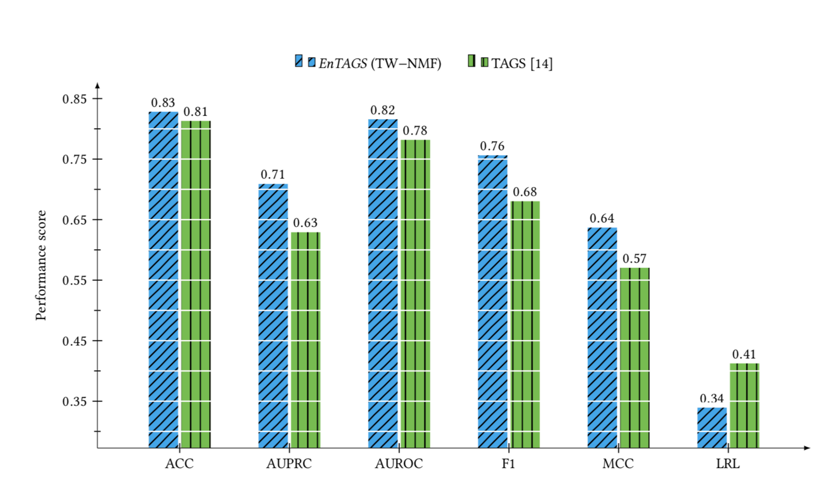
Dataset: Medical Information Mart for Intensive Care (MIMIC-III)

- Nursing notes of the database with the same selection criteria as employed in the state-of-the-art[Gangavarapu'19].
  - Records of neonates (age < 15) were eliminated from the chosen cohort.
  - Only notes corresponding to the first admission of a patient to a hospital.
- EnTAGS: the nursing notes of a patient are not aggregated by the note content.
- Independent modeling of notes reduces false negatives — false alarms vs. severity of not predicting a possible disease?

[Gangavarapu'19] Gangavarapu, Tushaar, *et al.* "TAGS: Towards Automated Classification of Unstructured Clinical Nursing Notes." *International Conference on Applications of Natural Language to Information Systems.* Springer, Cham, 2019.

Experimental Results

	Classifier	ACC	AUPRC	AUROC	CE	F1	LRL	MCC
BoW–NMF	CNN	0.7965 ± 0.0007	0.6688 ± 0.0018	0.7860 ± 0.0020	18.5327 ± 0.1412	0.7187 ± 0.0041	0.3898 ± 0.0045	0.5750 ± 0.0013
	LSTM	0.7921 ± 0.0005	0.6638 ± 0.0018	0.7794 ± 0.0017	18.6699 ± 0.0822	0.7093 ± 0.0030	0.4024 ± 0.0040	0.5652 ± 0.0016
	CNN–LSTM	0.8048 ± 0.0021	0.6806 ± 0.0031	0.7911 ± 0.0024	18.3760 ± 0.0919	0.7240 ± 0.0034	0.3825 ± 0.0046	0.5897 ± 0.0042
	GRU	0.7945 ± 0.0063	0.6698 ± 0.0057	0.7772 ± 0.0080	18.9530 ± 0.3058	0.7039 ± 0.0114	0.4081 ± 0.0143	0.5666 ± 0.0137
TW–NMF	CNN	0.8174 ± 0.0006	0.6948 ± 0.0014	0.8091 ± 0.0014	17.9663 ± 0.0562	0.7489 ± 0.0016	0.3499 ± 0.0032	0.6181 ± 0.0008
	LSTM	0.8129 ± 0.0015	0.6908 ± 0.0024	0.7992 ± 0.0012	18.2588 ± 0.0398	0.7347 ± 0.0020	0.3694 ± 0.0021	0.6062 ± 0.0028
	CNN–LSTM	**0.8282 ± 0.0023**	**0.7089 ± 0.0046**	**0.8157 ± 0.0019**	**17.6853 ± 0.0566**	**0.7562 ± 0.0021**	**0.3392 ± 0.0036**	**0.6368 ± 0.0042**
	GRU	0.82486 ± 0.0021	0.7089 ± 0.0019	0.8073 ± 0.0040	18.3412 ± 0.2126	0.7434 ± 0.0057	0.3569 ± 0.0079	0.6273 ± 0.0050



- EnTAGS aggregated, TW-NMF modeled, and classified using cascaded CNN-LSTM
- Seven evaluation metrics to assess the reliability of the proposed system
- Outperforms the state-of-the-art: 1.87% in accuracy, 12.68% in AUPRC, and 11.64% in MCC.

Discussion: TW-NMF EnTAGS Model

- Despite the use of NMF-modeled data with deep neural classifiers, we observe improved performance.
- NMF facilitates disentanglement of the hidden structure of the underlying data by learning features that exhibit sparse part-based representations.
- NMF forces data encoding to be nonnegative — additive representations of data.

NMF modeling is particularity well-suited to train deep neural models

Concluding Remarks

- Patient-specific information in the nursing notes to facilitate multi-label ICD-9 code group prediction.
- Promising results of 82.82% in accuracy, 70.89% in AUPRC, and 63.68% in MCC score; and outperformed the state-of-the-art model by 1.87% in accuracy, 12.68% in AUPRC, and 11.64% in MCC score.

Acknowledgements:
Government of India's DST-SERB Early Career Research Grant (ECR/2017/001056) to Sowmya Kamath S.