Predicting COVID-19 Patient Shielding: A Comprehensive Study

Vithya Yogarajan[1], Jacob Montiel[2], Tony Smith[2] and Bernhard Pfahringer[2]

[1] School of Computer Science, The University of Auckland
[2] Department of Computer Science, The University of Waikato

COVID-19 Patient Shielding

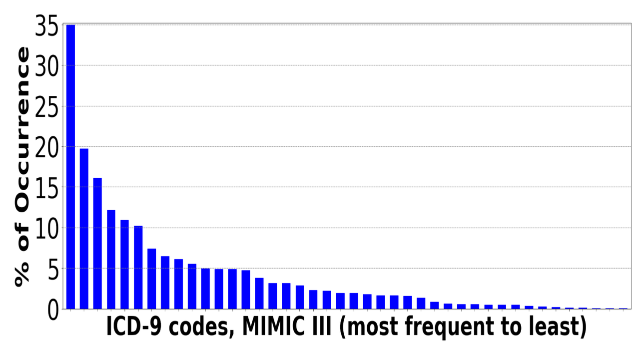
- Identify and protect patients who are clinically extremely vulnerable.
- Patients in this category:
  - identified before the pandemic as being clinically extremely vulnerable.
  - identified through the COVID-19 Population Risk Assessment model.
  - present with co-morbidity (hence, multi-label classification).
  - organ transplant, cancer, heart conditions while pregnant.

Data

MIMIC-III

- Free-form medical text.
- 35,458 patients with medical text data and 42 medical codes.
- ICD-9 code examples:
  - Code 285 (35% of the instances) Other and unspecified anemias
  - Code 996 (16%) Complications peculiar to certain specified procedures

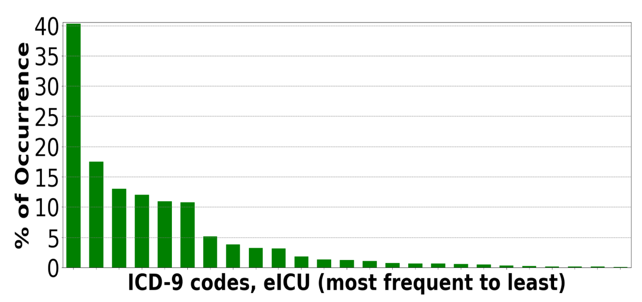
MIMIC III (sample text)
"82 yo M with h/o CHF, COPD on 5 L oxygen at baseline, tracheobronchomalacia s/p stent, presents with acute dyspnea over several days, and lethargy. This morning patient developed an acute worsening in dyspnea, & called EMS."



eICU

- Semi-structured medical text.
- 34,387 patients with medical text data and 25 medical codes.
- ICD-9 code examples:
  - Code 491 (40.3% of the instances) Chronic bronchitis
  - Code 288 (17.5%) Diseases of white blood cells

eICU (sample text)
"infectious diseases | medications | therapeutic antibacterials | cardio | inotropic agent | norepinephrine"



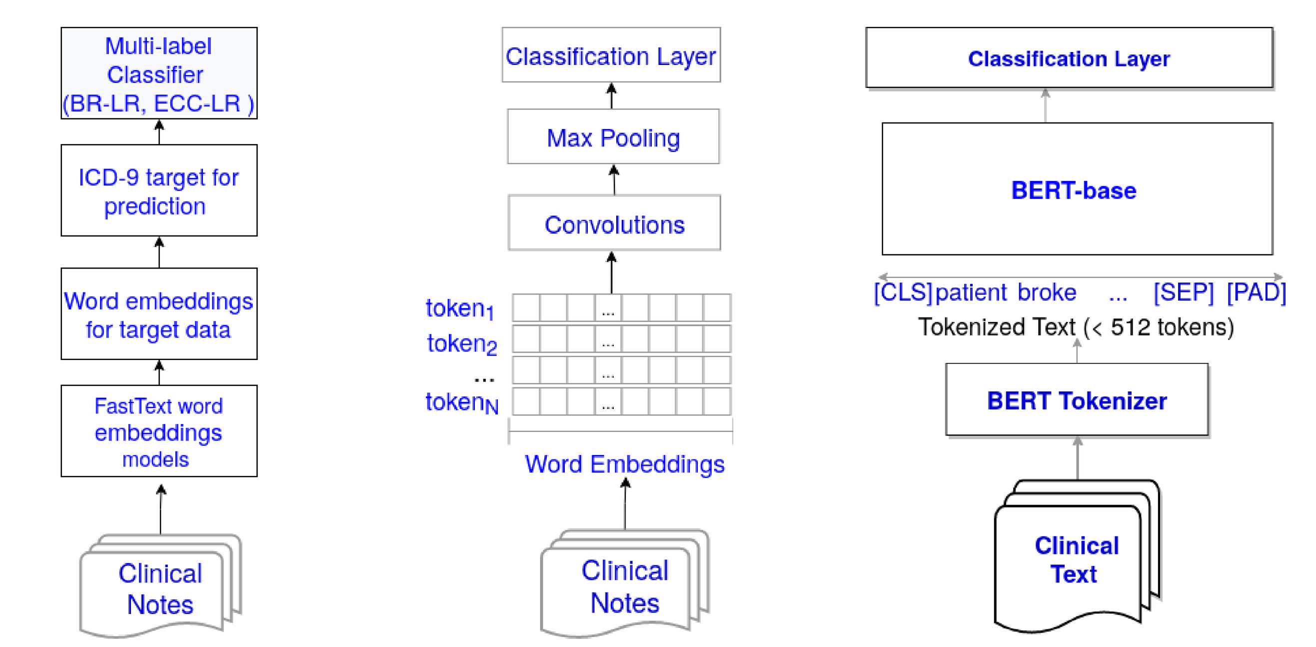


Fig 1: Traditional classifiers (left); CNNText (middle); and BERT-base for multi-label classification (right).

Multi-label Classifiers

- Multi-label classification assigns a set of labels to an instance. From a collection of labels, each record will be assigned relevant medical codes/labels.
- Traditional multi-label classifiers: binary relevance (BR) and ensembles of classifier chains (ECC), with base classifier logistic regression (LR).
- Neural networks using fastText pre-trained domain specific embeddings: CNNText, BiGRU, Convolutional Attention for Multi-Label classification (CAML) and Description Regularized-CAML (DR-CAML).
- Transformers that can handle a maximum sequence length of 512 tokens: BERT-base, ClinicalBERT, PubMedBERT, RoBERTa-base and BioMed-RoBERTa.
- Transformers that can handle long sequences of text: Longformer and TransformerXL.

Novel Multi-label Classification Approach

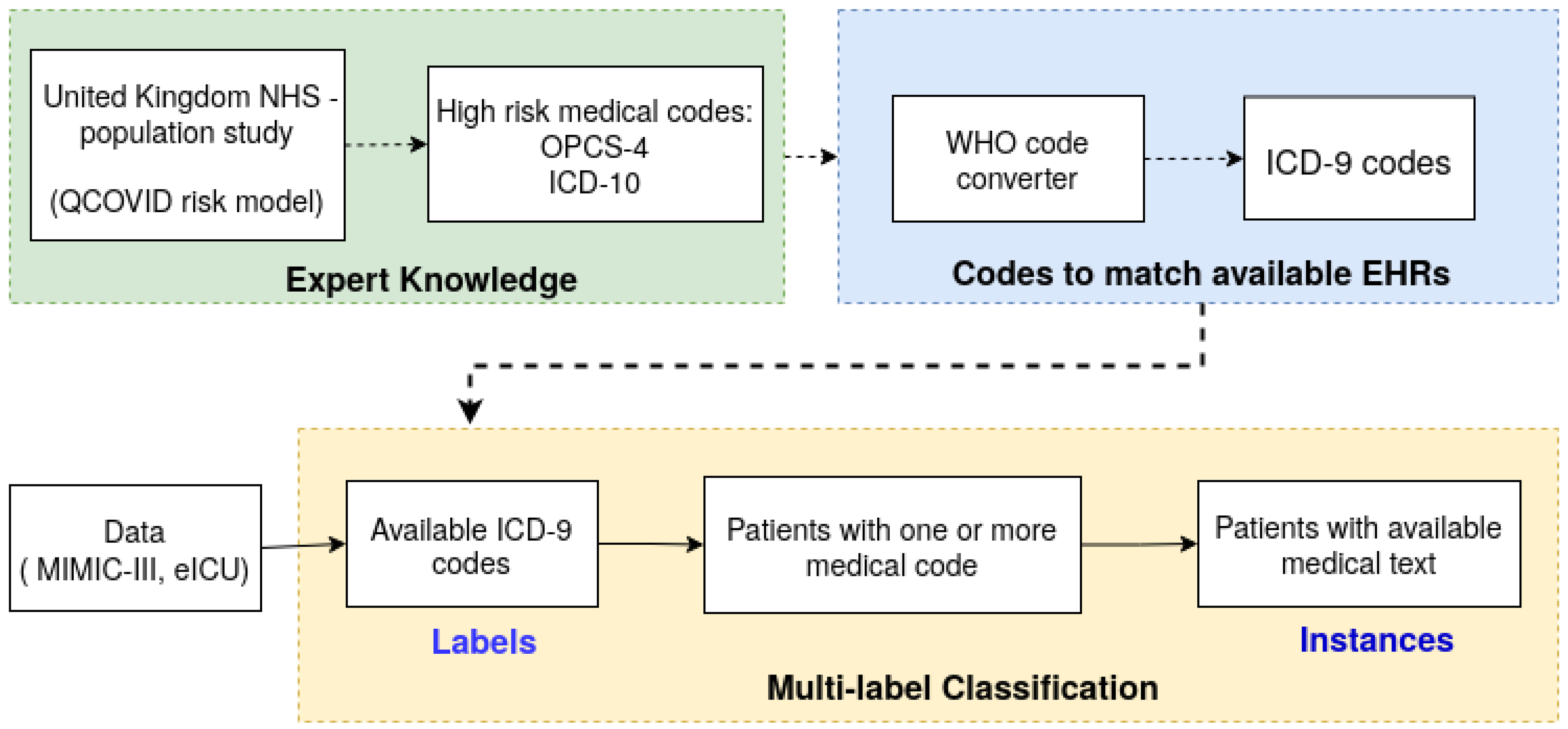


Fig 2: Flow chat forming labels and instances for multi-label classification of predicting COVID-19 patient shielding

Results

Classifiers	MIMIC-III Micro -F1	MIMIC-III Macro-F1	MIMIC-III Total Time (per run)	eICU Micro -F1	eICU Macro-F1	eICU Total Time (per run)
BR-LR	0.39	0.26	**12 min**	**0.54**	**0.28**	**7 min**
ECC-LR	**0.45**	**0.27**	38 min	0.51	**0.28**	34 min
CNNText	0.58	**0.42**	46 min	0.59	**0.36**	45 min
BiGRU	0.59	0.31	216 min	0.59	0.35	210 min
CAML	**0.61**	0.40	49 min	0.60	0.32	48 min
DRCAML	0.60	0.39	64 min	**0.61**	0.32	60 min
BERT-base	0.50	0.44	10 hr	0.60	0.36	11 hr
ClinicalBERT	0.51	0.45	16 hr	0.60	0.36	11 hr
BioMed-RoBERTa	0.53	0.45	12 hr	0.61	0.37	11 hr
PubMedBERT	0.54	0.48	16 hr	0.64	0.39	14 hr
Longformer	0.58	0.50	82 hr	0.61	0.40	49 hr
TransformerXL	0.65	0.51	206 hr	0.63	0.40	53 hr

Tab. 1: Micro-F1 and macro-F1 of ICD-9 codes for various multi-label classifiers for both MIMIC-III and eICU data. Time required per run is also presented. Experiments for neural networks and transformers were run on a 12 core Intel(R) Xeon(R) W-2133 CPU @ 3.60GHz, and a GPU device GV100GL. Experiments for BR and ECC were run on a 4 core Intel i7-6700K CPU @ 4.00GHz with 64GB of RAM. Results are averaged over three runs for neural networks and transformers. 10-fold cross validation is used for BR-LR and ECC-LR.

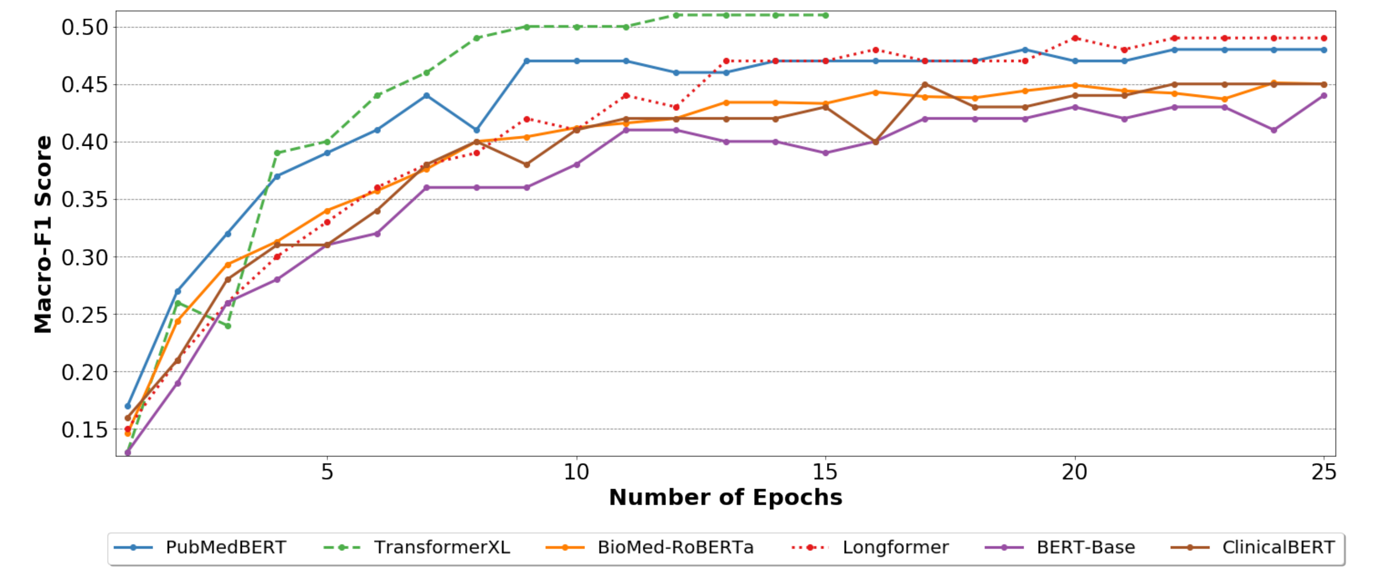


Fig 3: Macro F1 scores of transformer models for COVID-19 patient shielding over number of epochs for MIMIC-III data. Due to resource restrictions, TransformerXL was only run for 15 epochs.

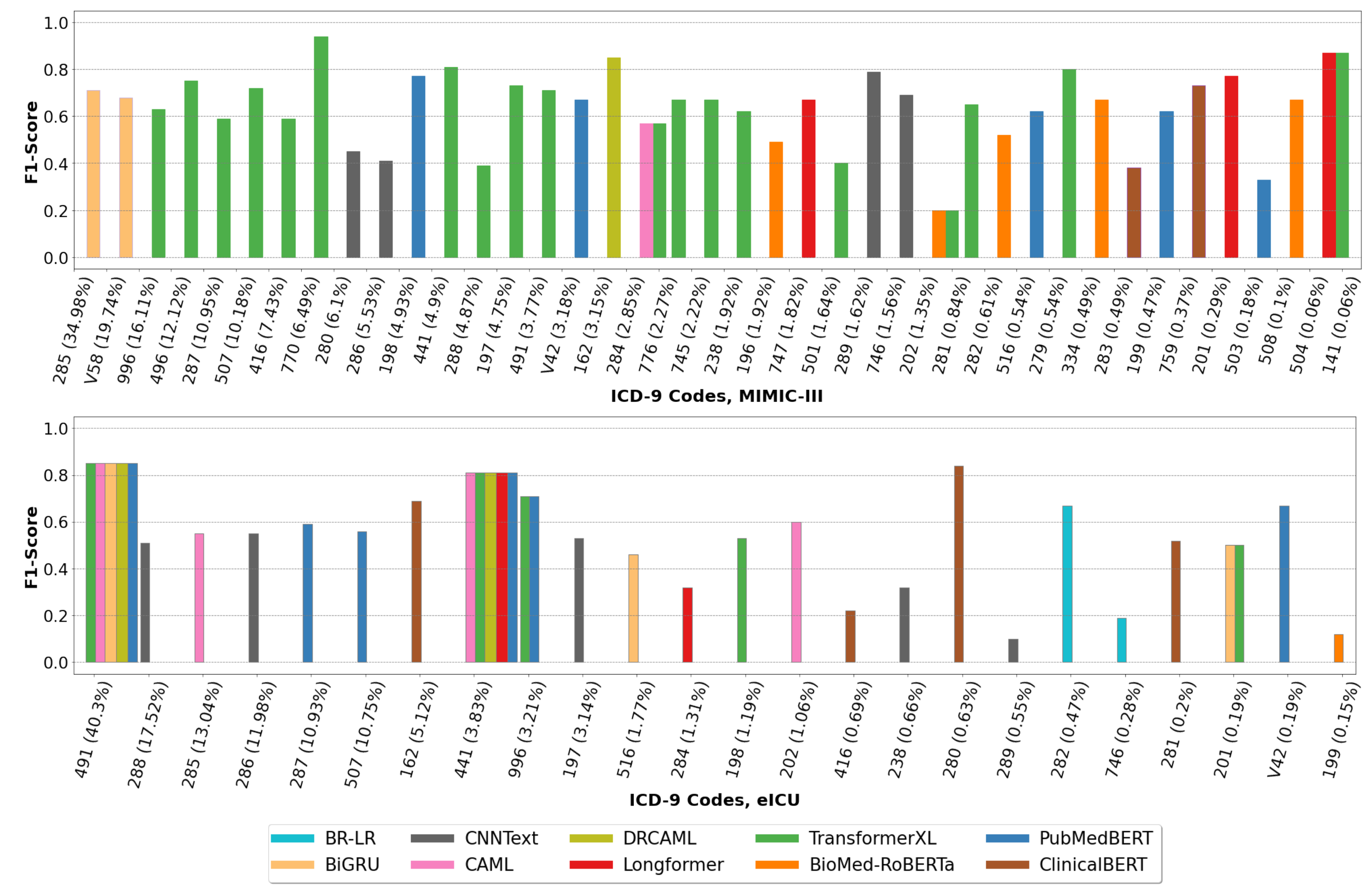


Fig 4: Best F1 scores and corresponding multi-label classifiers for individual predicted labels.

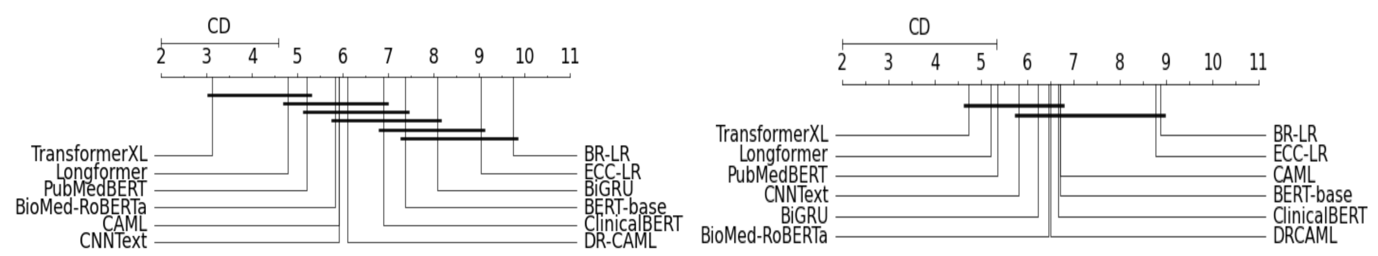


Fig 5: Critical difference plots of label F1 scores for COVID-19 patient shielding: MIMIC-III (left) and eICU (right).

Discussion

- Research in the use of machine learning approaches in the fight against COVID-19 is fast growing.
- We propose a novel multi-label approach to predicting COVID-19 patient shielding from medical text.
- We use publicly available information on COVID-19 and publicly available data.
- We present an extensive study on 12 different multi-label approaches.
- If overall predictive accuracy is the only deciding factor TransformerXL is the best option.
- However, given pandemic situations where time is a significant factor, the predictive accuracy of models is not the only factor to consider in selecting an approach.

Yogarajan, V., Montiel J., Smith T., & Pfahringer B. (2022) Predicting COVID-19 Patient Shielding: A Comprehensive Study. In Proc 34th Australasian Joint Conference on Artificial Intelligence (AJCAI), Sydney, Australia. Springer.
https://github.com/vithyayogarajan/COVID-19-Patient-Shielding