Incorporating Uncertainty in Learning to Defer Algorithms for Safe Computer-Aided Diagnosis

Jessie Liu[1*], Blanca Gallego[1], Sebastiano Barbieri[1]

[1]Centre for Big Data Research in Health, University of New South Wales, Australia *Corresponding email: jessie.liu1@unsw.edu.au

UNSW SYDNEY

CENTRE FOR BIG DATA RESEARCH IN HEALTH

Background

Risks of using deep learning techniques in automated diagnoses:

- Errors committed by AI can be associated with very high costs for patients.
- AI/DL models can give uncertain answers:
  Lack of training regarding the domain knowledge -> Model uncertainty
  Noise in data -> Data-associated uncertainty

Introduction

We propose a learning to defer with uncertainty (LDU) algorithm which identifies patients for whom diagnostic uncertainty is high and defers them for evaluation by human experts.

The hypothesis to be tested in this study is that LDU results in higher diagnostic accuracy and fewer deferred patients when compared with learning to defer (without uncertainty) and direct triage by uncertainty algorithms, across different types of diagnostic tasks.

Three diagnostic tasks

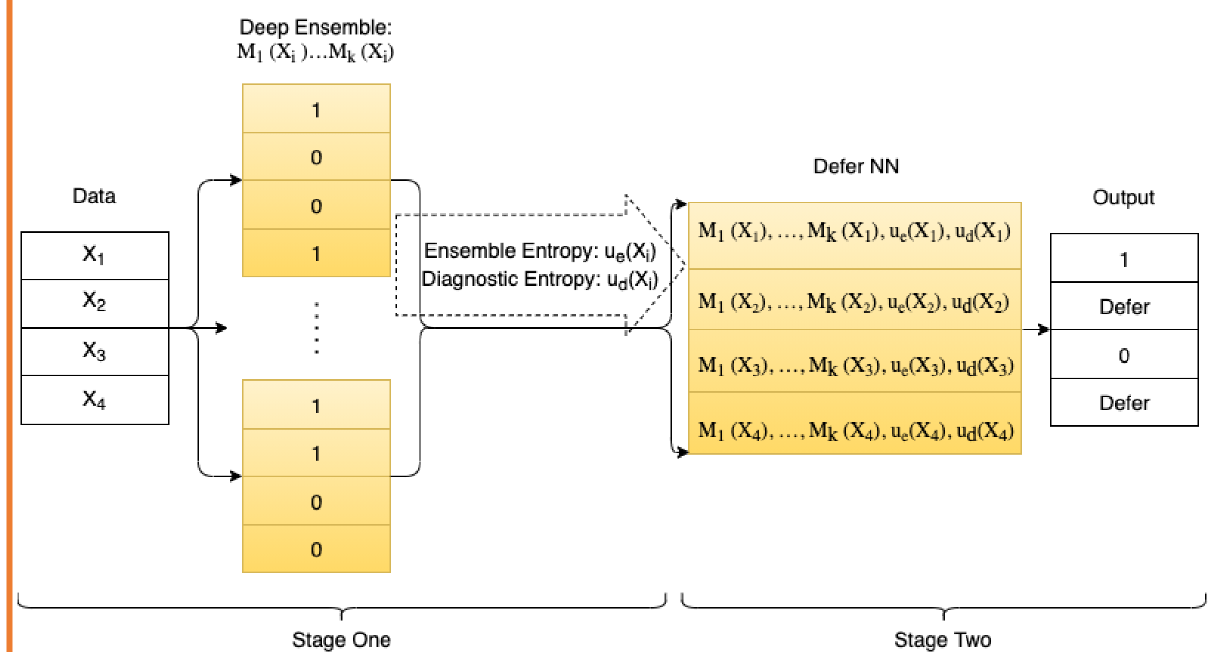
- Diagnosis of myocardial infarction (MIMIC-III, text notes)
- Diagnosis of any comorbidities (Heritage Health dataset, EHRs)
- Diagnosis of pleural effusion and pneumothorax (MIMIC-CXR, x-ray images)

Code availability:

References:

Method

The learning to defer with uncertainty (**LDU**) algorithm:



The **LDU** algorithm consists of two stages:

1. In stage one an ensemble of neural networks is trained for the diagnostic task, and diagnoses are determined for every patient with associated uncertainty measures.
2. In stage two a 'learning to defer' neural network takes as input thepredicted diagnoses and uncertainty measures from stage one and outputs either the patient's diagnosis or the decision to defer to a human expert. A parameter alpha is used to adjust the weight of samples that the model decides to defer.

Results and Key contributions

LDU Compared to the direct triage by uncertainty (DT) and the learning to defer without uncertainty information (LD):

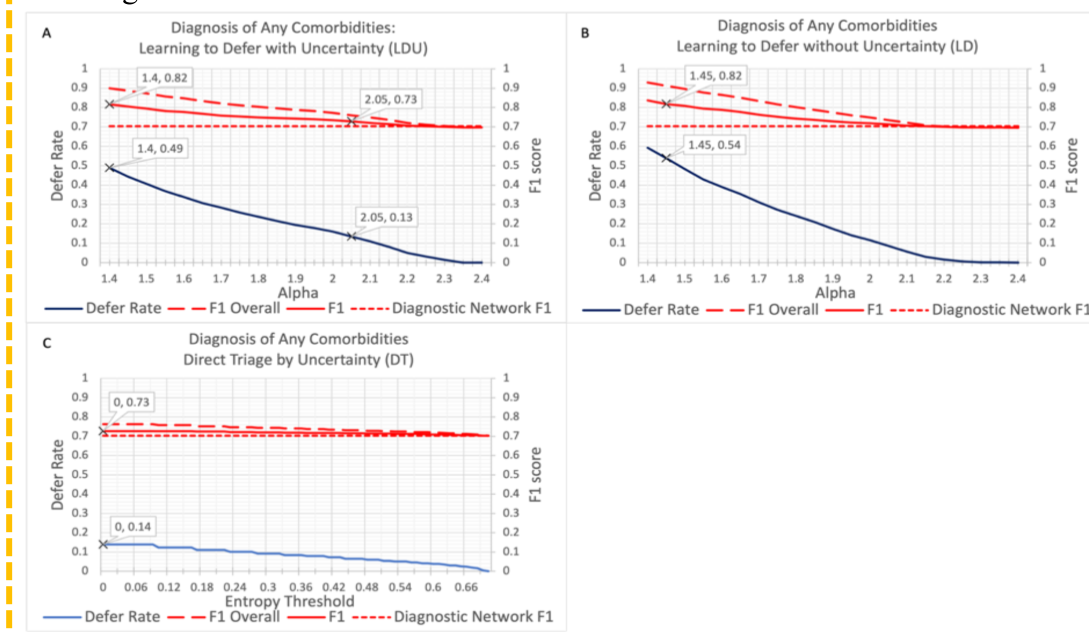
- LDU deferred considerably fewer patients for achieving the same performance.

Data Source	Diagnostic Task	Method	F1	Defer Rate	Accuracy	Sensitivity	Specificity
MIMIC-III	Myocardial Infarction	LD	0.88	55%	0.84	0.70	0.93
		LDU	0.88	38%	0.90	0.94	0.84
Heritage Health	Comorbidity	LD	0.82	54%	0.81	0.79	0.83
		LDU	0.82	49%	0.80	0.76	0.85
MIMIC-CXR	Pleural Effusion	LD	0.96	69%	0.96	0.97	0.96
		LDU	0.96	36%	0.96	0.98	0.94
	Pneumothorax	LD	0.84	75%	0.76	0.42	0.94
		LDU	0.84	51%	0.83	0.85	0.82

Results and Key contributions (continued)

- LDU algorithm can be used as a "wrapper" to any existing diagnostic model. **API** is provided at https://github.com/liu-res/Learning-to-Defer-with-Uncertainty.

- Effective when many patients were assigned the wrong diagnosis with high confidence.



- Trade-off between diagnostic accuracy and deferral rate can be obtained by adjusting the weight of the defer loss in LDU