Title

Author 1
Department of YYY, University of XXX
and
Author 2
Department of ZZZ, University of WWW

July 22, 2021

Contents

1 Abstract

2 Introduction

2.1 Background

2.2 Related Work

3 Methodology

3.1 Acquisition of Data

Our gold-standard dataset consisted of 16K sampled patients with APOE gene ¡insert info about Biobank¿. Each patient's EHR record was reviewed by an expert clinician (neurologist, psychiatrist, or geriatric psychiatrist) to annotate patients with presence, absence, or inconclusiveness of cognitive impairment.

3.2 Preprocessing

We first created a list of keywords related to cognitive impairment (full list with frequency matches in Appendix A). For each patient in the BioBank data repository (mentioned above), we extracted all sequences containing any of the keywords. Several preprocessing steps were completed to increase the sequences to 800 characters minimum (more details in Appendix B) to allow for annotators to have more context when evaluating sequences.The preprocessed sequences were used as input for all models.

3.3 Machine Learning Models

3.4 Deep Learning Model

4 Results

4.1 Model Results and Comparison

5 Conclusion

6 Future Work

7 References

8 Appendix A

9 Appendix B

Rank	Keyword	Match Count
1	**Memory**	109218
2	**Cognition**	87655
3	**Dementia**	51034
4	**Cerebral**	45886
5	**Cerebrovascular**	36370
6	**Cerebellar**	26863
7	**Cognitive Impairment**	20267
8	**Alzheimer**	20581
9	**MOCA**	9767
10	**Neurocognitive**	7711
11	**MCI**	3889
12	**Amnesia**	3695
13	**AD**	2673
14	**Lewy**	2561
15	**MMSE**	2134
16	**LBD**	224
17	**Corticobasal**	147
18	**Pick's**	41

Table 1: Keywords indicative of Cognitive Impairment