Using Deep Learning to Identify Patients with Cognitive Impairment in Electronic Health Records

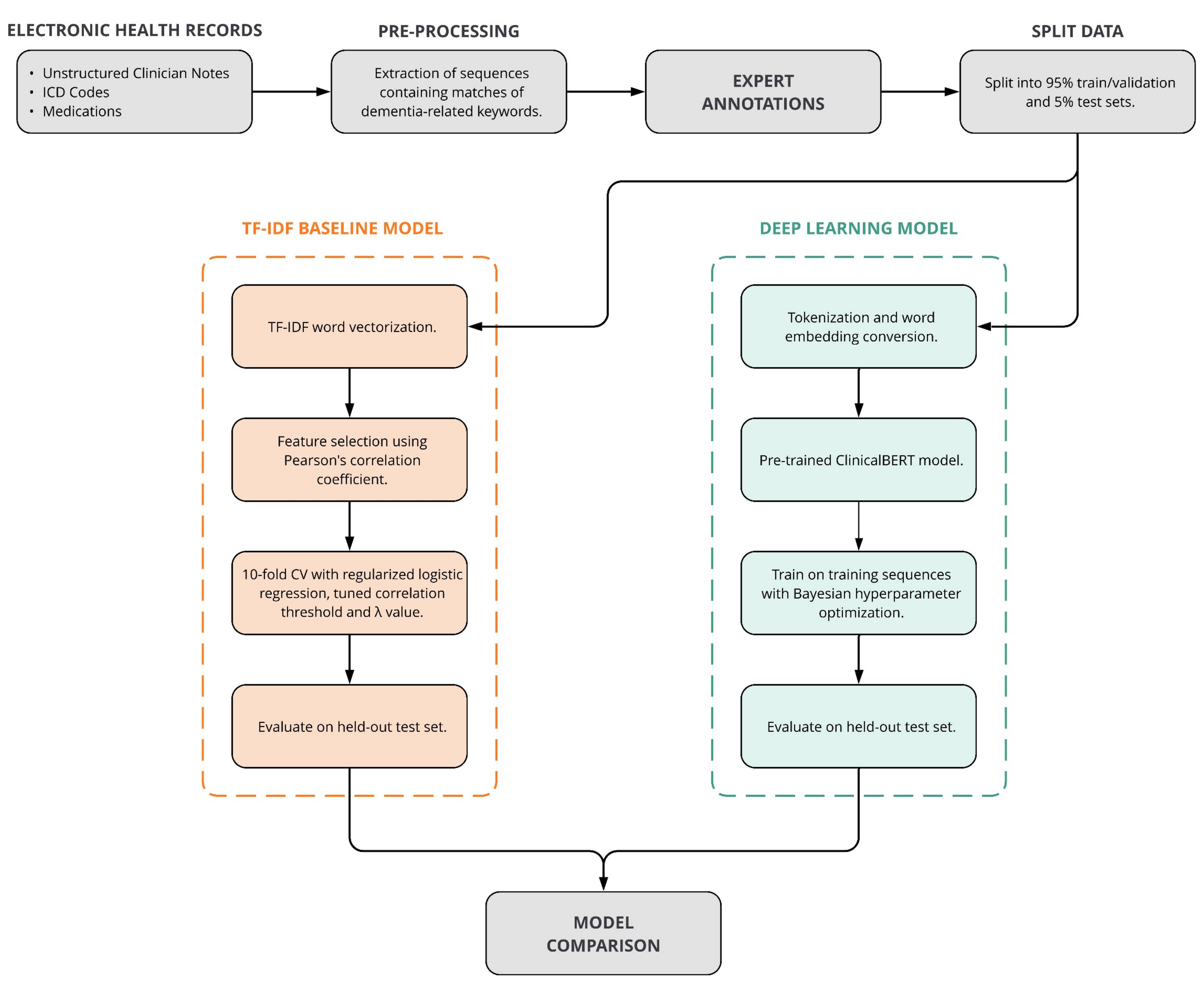
Tanish Tyagi*, Colin G. Magdamo*, Ayush Noori*, Zhaozhi Li*, Xiao Liu*, Mayuresh Deodhar, Zhuoqiao Hong, Wendong Ge, Elissa M. Ye, Yi-han Sheu, Haitham Alabsi, Laura Brenner, Gregory K. Robbins, Sahar Zafar, Nicole Benson, Lidia Moura, John Hsu, Alberto Serrano-Pozo, Dimitry Prokopenko, Rudolph E. Tanzi, Bradley T. Hyman, Deborah Blacker, Shibani S. Mukerji, M. Brandon Westover, Sudeshna Das

NEURAL INFORMATION PROCESSING SYSTEMS

Abstract

Dementia is a neurodegenerative disorder that causes cognitive decline and affects more than 50 million people worldwide. Dementia is underdiagnosed by healthcare professionals – only one in four people who suffer from dementia receive a diagnosis – and successful diagnoses may not be entered as a structured International Classification of Diseases (ICD) diagnosis code in a patient's chart. Indeed, information relevant to cognitive impairment (CI) is often found within electronic health records (EHR) but manual review of clinician notes by experts is both time consuming and often prone to errors. Automated mining of these notes presents an opportunity to label patients with cognitive impairment in EHR data.

We developed natural language processing (NLP) tools to identify patients with cognitive impairment and demonstrate that linguistic context enhances performance for the classification task. We fine-tuned our attention based deep learning model, which can learn from complex language structures, and substantially improved accuracy (0.93) relative to a baseline TF-IDF (term frequency-inverse document frequency) NLP model (0.84). Further, we show that deep learning NLP can successfully identify dementia patients without dementia-related ICD codes or medications.



Method Overview

Cohort Details

Characteristic	(N = 16428)
Age (years) mean (SD)	73.01 (7.96)
Gender Male, n (%)	8740 (53.2)
Race, n (%)	
White	14896 (90.7)
Other/Not Recorded	608 (3.7)
Black	570 (3.5)
Hispanic	170 (1.0)
Asian	168 (1.0)
Indigenous	16 (0.01)
APOE Genotype, n (%)	
APOE $\varepsilon 2$	2028 (12.3)
APOE $\varepsilon 3$	10177 (62.0)
APOE $\varepsilon 4$	4223 (25.7)
Average Speciality Visits (SD)	1.67 (4.6)
Average PCP Encounters (SD)	5.25 (5.63)

Our dataset consisted of N = 16,428 patients from the Mass General Brigham HealthCare system who were older than 60 years, had *APOE* genotypes (the biggest genetic risk factor for Alzheimer's), and at least one note with a dementia related keyword.

Classification Task

Experts annotated sequences using a web-based annotation tool and labeled them as **(1) Yes**, i.e., patient has CI; **(2) No** i.e., Patient does not have CI; and **(3) Neither** i.e., sequence has no information on cognition.
We used a dataset of 8,656 annotated sequences from N = 2,487 unique patients split between train (90%) and holdout test (10%) sets.

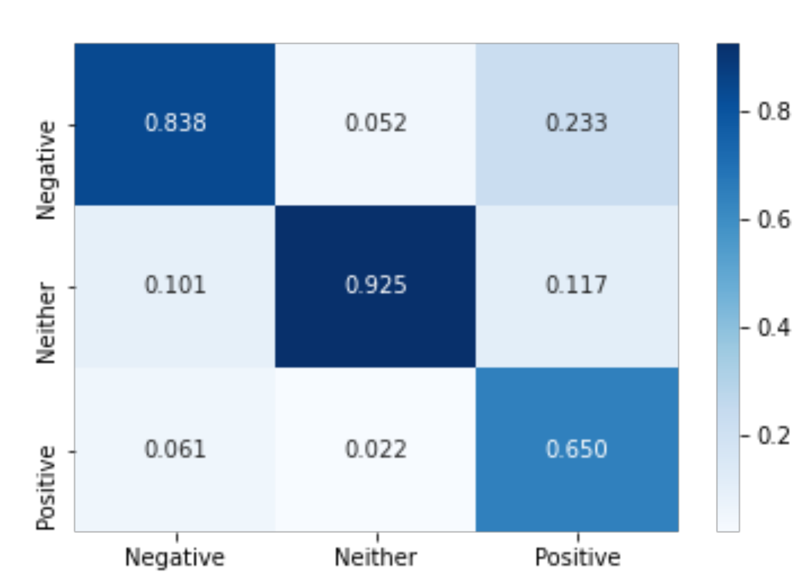
Model Performance

We developed and compared two NLP models for the classification task:

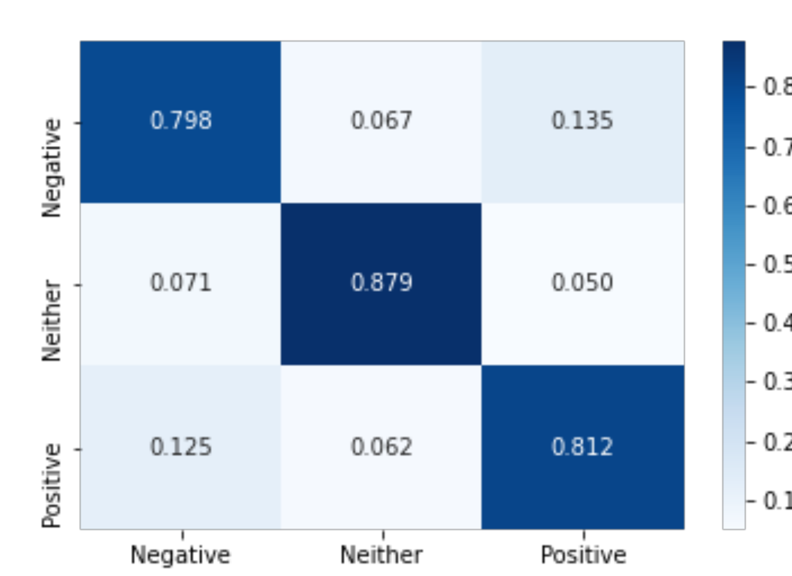
BASELINE: *Logistic Regression with TF-IDF Vectors*
We performed TF-IDF (term frequency-inverse document frequency) vectorization on the annotated sequences and selected features based on a term's Pearson correlation with the cognitive impairment outcome.

DEEP LEARNING: *Transformer Based Sequence Classification Model*
We utilized a pre-trained language model called ClinicalBERT. After text preprocessing, input texts were tokenized and converted to embeddings. Optuna was used to optimize the hyperparameters over 20 trials.

Model	AUC	Accuracy	Sensitivity	Specificity	Micro F1	Macro F1	Weighted F1
TF-IDF	0.95	0.84	0.83	0.92	0.84	0.81	0.84
ClincialBERT	0.98	0.93	0.91	0.96	0.93	0.92	0.93

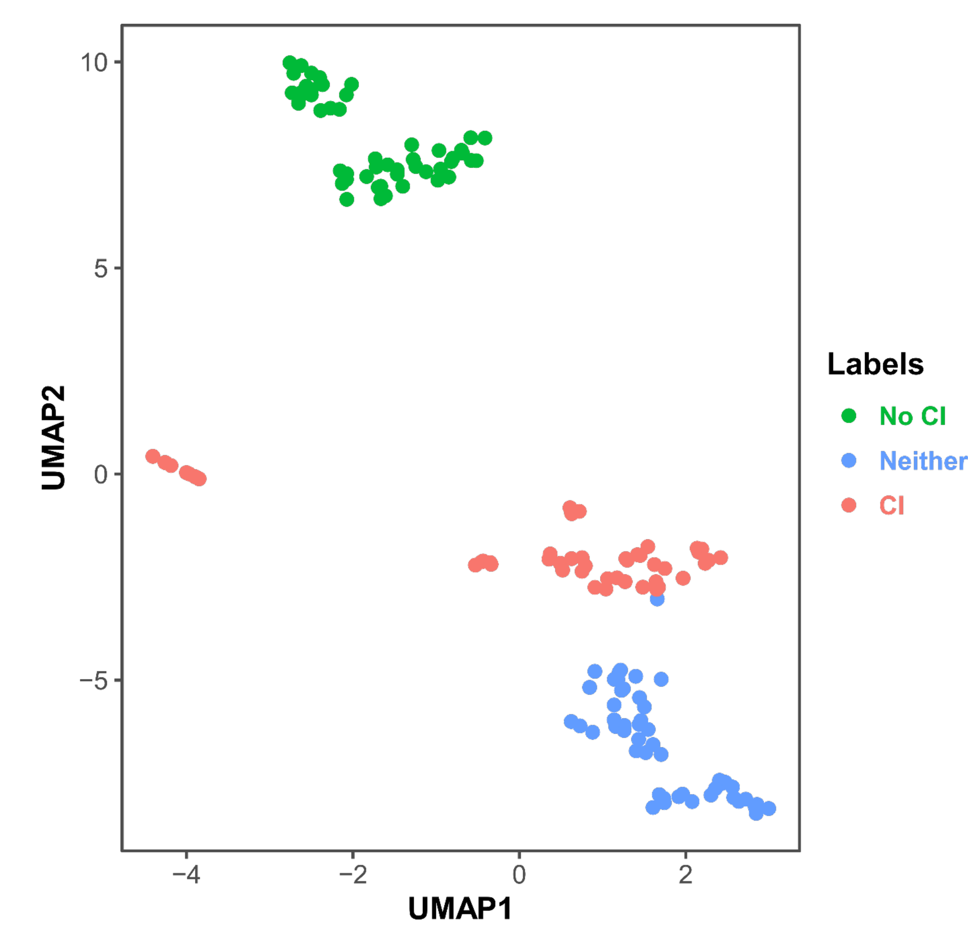


Precision Matrix



Recall Matrix

Advantages of ClinicalBERT vs. TF-IDF



ClinicalBERT, with its more complex architecture, was able to leverage the context of the keyword matches within the sequences and overcome these issues. Indeed, using a small dataset of manually annotated sequences (n=150) which did not match an always-pattern, the ClinicalBERT embeddings were able to able accurately discriminate between all three classes.

Example Sequences

Examples of sequences with the keyword "memory" used in different contexts:

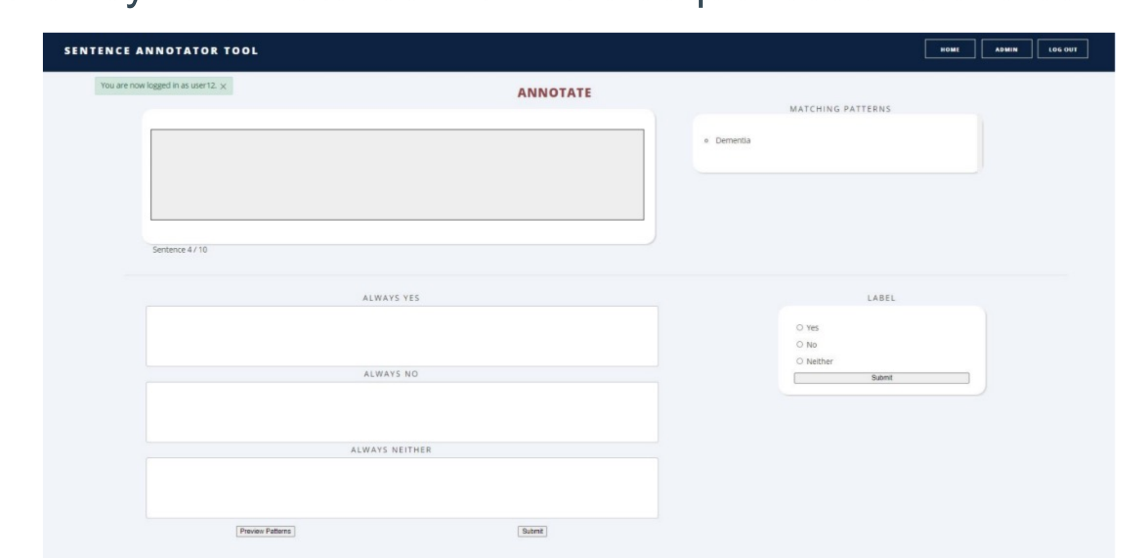
(1) Yes: "increased short-term memory loss and confusion"

(2) No: "fund of knowledge and memory were normal"

(3) Neither: "mother had memory problems in her 70s"

Sentence Level Annotation Tool

ClinicalBERT, with its more complex architecture, was able to leverage the context of the keyword matches within the sequences and overcome these



Annotation Tool User Interface

References

E. Alsentzer *et al.,* (2019). "Publicly Available Clinical BERT Embeddings." In: Proceedings of the 2nd Clinical Natural Language Processing Workshop, pp. 72–78. doi: 10.18653/v1/W19-1909.

Thomas Wolf *et al.,* (2019). "Huggingface's transformers: State-of-the-art natural language processing." In: ArXiv, abs/1910.03771,2019.

Leland McInnes, John Healy, and James Melville. (2018) "Umap: Uniform manifold approximation and projection for dimension reduction". In: arXiv:1802.03426, 2018.

MASSACHUSETTS GENERAL HOSPITAL