Explainable Survival Analysis for Dementia Prediction

Jacob Thrasher, Prashnna Gyawali

***Abstract*—This study explores different machine learning-based survival analysis approaches to predict the probability of Alzheimer's Disease (AD) dementia progression. We utilize the Alzheimer's Disease Neuroimaging Initiative (ADNI) data and analyze different features to explain their importance in disease progression.**

***Clinical Relevance*—The study's findings can help us understand mechanism of AD Dementia, predict the patients' AD shift efficiently and recommend personalized treatment to mitigate or postpone the effects of AD.**

I. Introduction

Alzheimer's Disease (AD) and related dementia are one of the most common neurodegenerative diseases, and their early prediction can significantly aid in preventing or delaying the development of symptoms or allow timely access to treatments to manage the symptoms [3]. Although numerous machine learning (ML) methods have been explored to predict the probability of AD or dementia, the ability to accurately predict the next stage of progression of dementia for a patient over a given period is clinically more applicable as it would help physicians make informed clinical decisions on treatment strategies. Toward this, ML-based survival analysis can be useful in providing a time-to-event prediction for AD data [1]. However, little to no similar research has been conducted to explain the importance of features in such time-to-event prediction allowing trust and understanding of the model, which are required for real-world clinical adoption. As such, in this study, we focus on exploring different survival analysis approaches and use permutation importance [2] to analyze the effect of various characteristics such as age, gender, or clinical results on the predicted outcome of an event.

II. Methods

We consider the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which contains biological and clinical assessment data for classifying patients to different stages of dementia. This study focuses only on the clinical assessment and demographic data. V_p represents all recorded visits for a given patient p and $V_{p,i}$ corresponds to visit i for patient p. For each patient, we select $V_{p,0}$ as our input and $V_{p,D}$ as the label, where D is the first visit p received either a positive dementia diagnosis or the final visit if right-censored.

We selected three models for baseline analysis: Cox Regression (cox), penalized Cox Regression (coxnet), and Random Survival Forest (RSF) from the SciKit Survival library [4]. We analyzed model performance on multiple time frames from 6 months to 5 years, increasing by 6 month intervals. The models were then evaluated using time-dependent AUC (td-AUC) metric. We then use permutation importance to measure the effect each feature has on the performance of the models.

III. Results

We observed that the td-AUC increased steadily as the time frame expanded with means of $0.891, 0.897$, and 0.926 for cox, coxnet, and RSF, respectively. These results are expected since more information becomes available as time frames increase. The permutation importance results across the three models were aggregated by averaging each feature as shown in **Fig. 1**. Short-term auditory-verbal retention (RAVLT), AD Assessment Scale (ADAS), and the Functional Activities Questionnaire (FAQ) were found to be the three most important features, consistent with the clinical findings in [3].

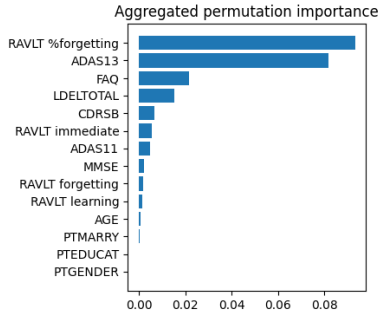

Fig. 1. Aggregated scores obtained from selected models

IV. Discussion and Future Work

This work presents the survival analysis for dementia progression prediction and explains the role of input features for such outcome. Our future work aims to expand this to a robust multimodal model that considers both clinical and biological data for survival analysis.

References

[1] G. Mirabnahrazam et al., "Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis," Neurobiology of Aging (2023).

[2] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

[3] Porsteinsson, A.P., Isaacson, R.S., Knox, S. et al. Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021. J Prev Alzheimers Dis 8, 371–386 (2021). https://doi.org/10.14283/jpad.2021.23

[4] S. Pölsterl, "scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn," Journal of Machine Learning Research, vol. 21, no. 212, pp. 1–6, 2020.