BIS 687 Group 4 Proposal

Baijia Xu, Jinxuan Bian, Huangrui Chu

2024-02-20

Background

Alzheimer's Disease (AD) stands as a formidable challenge in the landscape of global health, affecting an estimated 50 million people worldwide. Characterized by progressive neurodegeneration leading to cognitive decline, AD is the most common cause of dementia among older adults, casting a shadow over the golden years of an increasing segment of the population. Despite significant research efforts, the pathophysiology of AD remains complex and multifactorial, involving genetic, molecular, and environmental components that interplay over time. The Alzheimer's Disease Neuroimaging Initiative (ADNI) has been instrumental in elucidating this complexity, providing an extensive database of clinical, cognitive, and biomarker information. However, the trajectory of AD progression varies greatly among individuals, and current predictive models often fail to capture the nuances of this variability. The incorporation of time-varying factors into these models is a burgeoning field of inquiry that promises to refine predictive accuracy. This background underscores the urgent need for innovative research approaches that can leverage such dynamic data to enhance our understanding and prediction of AD progression, ultimately leading to improved patient outcomes and more effective allocation of healthcare resources.

Significance

Enhanced predictive models incorporating time-varying factors could drastically improve early detection of AD, a critical factor since early intervention can slow disease progression and significantly improve patient quality of life. Beyond individual patient care, the implications of more accurate predictions extend to healthcare systems and resource allocation, enabling more efficient and targeted use of resources for patient support and research initiatives. Understanding the dynamics of AD progression through this research could also shed light on the underlying biological and environmental mechanisms of the disease, potentially opening new avenues for therapeutic intervention and preventative strategies. Furthermore, the impact on caregivers and families could be profound, providing them with a more precise prognosis and better preparation for future care requirements.

Innovation

This study introduces an approach to Alzheimer's Disease (AD) progression modeling by integrating landmark analysis with random survival forests—a method not conventionally applied in this domain. This innovation lies in the dynamic incorporation of time-varying factors,

such as biomarker changes or treatment responses, which are crucial for capturing the real-time progression of AD. By benchmarking these advanced machine learning techniques against traditional Cox proportional hazards models, the research promises to not only improve prediction accuracy but also to provide a more granular understanding of AD dynamics.

Research Plan

The research plan for evaluating and improving the predictive power of clinical and cognitive measures on the progression of Alzheimer's Disease (AD) by incorporating a time-varying factor includes several steps: 1) Data Collection 2) Identification of Time-Varying Factors 3) Development of Predictive Model 4) Model Validation and Benchmarking 5) Analysis and Interpretation 6) Continual Improvement and Future Research. Details of each step are included as below.

Research Strategy

Specific Aim

To Evaluate and improve the Predictive Power of Clinical and Cognitive Measures at Multiple Time Points on the Progression of Alzheimer's Disease by Incorporating an Additional Time-Varying Factor

Hypothesis

We hypothesize that incorporating a time-varying factor (such as Mini Mental State Exam (MMSE), Magnetic Resonance Imaging (MRI), and etc.) at multiple time points into predictive models with clinical and cognitive measures will significantly improve the predictions of Alzheimer’s Disease (AD) progression, providing a more comprehensive assessment.

Rationale

Time-varying factors in the ADNI dataset refer to those variables that change over time as they are measured across multiple visits or time points for the study participants.

- Cognitive Scores: Mini-Mental State Examination (MMSE), the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog), and others.
- Biomarkers: Biomarkers in cerebrospinal fluid (CSF), blood, and other bodily fluids, such as amyloid-beta levels, tau protein levels, and neurofilament light chain, and others.
- Imaging Data: Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans, including brain volume, cortical thickness, amyloid and tau PET imaging.
- Clinical Assessments: Clinical Dementia Rating (CDR) and the conversion from Mild Cognitive Impairment (MCI) to Alzheimer's Disease, and others.

These factors have a dynamic relationship with cognitive health and can change significantly over time. Traditional models often utilize static snapshots of disease status and risk factors, potentially overlooking insights into how changes in the factors impact AD progression. By

incorporating time-varying data may capture the complexity of disease progression more accurately.

Experimental Approach

- **Data Collection and Preparation**: Utilize data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), focusing on patients with detailed records of clinical and cognitive assessments, alongside longitudinally collected factor data.
- **Selection of Time-Varying Factors**: Identify key factors that vary over time and are likely to influence AD progression and that have been recorded at multiple time points. Begin with literature review and meta-analyses to figure out which cognitive scores, biomarkers, genetic factors, and clinical assessments have shown significant correlations with disease progression. Pearson or Spearman correlation coefficients may also be used to identify the most relevant key factors.
- **Statistical Modeling**: Employ Landmark Analysis to periodically select cohorts of patients who have reached certain points in the disease progression. Then, for each "landmark" time, we build random survival forests to predict survival probabilities based on the data, accounting for time-dependent nature of survival data and potentially more dynamic and accurate predictions over time.
- **Analysis**: Conduct feature selection using Principal component analysis (PCA), identifying the most predictive variables from clinical scores, cognitive test outcomes, and time-varying factors such as medication use, lifestyle alterations, or biomarker levels. Then, the statistical approach of landmark analysis with random survival forests will be employed. These forests, adept at capturing complex interactions, will incorporate time-varying factors as dynamic inputs, allowing for nuanced predictions that reflect changes over time.
- **Competing method**: Cox proportional hazards models with time-varying covariates would serve as a competing benchmark.
- **Model Validation**: Use cross-validation techniques to assess the predictive performance of the model. Performance metrics, such as concordance index (C-index) would be used to evaluate the model's predictive capabilities. Sensitivity analyses would also be conducted to further explore the robustness of the model, helping to understand how changes in time-varying factors or assumptions about the disease progression impact the predictions and providing a deeper insight into the dynamics of AD progression and enhancing the model's applicability in clinical settings.

Interpretation of Results

The results will provide insights into the progression of Alzheimer's Disease (AD). Should the hypothesis be confirmed, the successful incorporation of time-varying factors—potentially biomarkers—into the predictive models would signify a notable improvement over the traditional method. This could lead to a more nuanced understanding of AD progression and

potentially inform more personalized treatment approaches. The interpretation will focus on the magnitude and direction of the impact that the time-varying factor has on disease progression. A significant change in predictive power, as reflected by an increased C-index in the model including the time-varying factor, would support the hypothesis that incorporating such dynamic factors offers a more accurate and holistic assessment of AD risk over time. Additionally, the landmark analysis would provide snapshots of the disease progression at various stages, which could reveal critical windows for therapeutic intervention.

Potential Problems and Alternative Approaches

The complexity of AD progression may mean that even sophisticated models like random survival forests may not fully capture the disease's dynamics. If the models do not perform as expected, alternative machine learning approaches, such as deep learning methods that can handle complex interactions in high-dimensional data, could be considered. Secondly, the accuracy of the time-varying factors is contingent on the quality and granularity of the data collected. Poor data quality or missing data could lead to inaccurate predictions. In such cases, robust imputation methods or sensitivity analyses that account for missing data should be employed. Finally, there's the challenge of ensuring that the models do not overfit the training data, which would compromise their generalizability. To mitigate this, a rigorous cross-validation framework is essential, and alternative approaches such as penalized regression models could be explored if overfitting is detected. These alternative approaches can provide a more parsimonious model, reducing the risk of overfitting while still allowing for the inclusion of time-varying effects.

Reference

Li, Kan, and Sheng Luo. “Dynamic prediction of Alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data.” Statistics in medicine vol. 38,24 (2019): 4804-4818. doi:10.1002/sim.8334

Shu Jiang, Yijun Xie, Graham A. Colditz, Functional Ensemble Survival Tree: Dynamic Prediction of Alzheimer’s Disease Progression Accommodating Multiple Time-Varying Covariates, Journal of the Royal Statistical Society Series C: Applied Statistics, Volume 70, Issue 1, January 2021, Pages 66–79, https://doi.org/10.1111/rssc.12449