A Machine Learning Approach to Identifying Blood-Based Biomarkers for Differential Diagnosis of Alzheimer’s Disease

Anjali Sreenivas | Nikola Tesla STEM High School, Redmond, WA | CBIO038

Introduction

What do we already know?
- Alzheimer’s Disease (AD) is a devastating age-related neurodegenerative disorder of the brain.
- Clinically characterized by cognitive deterioration and memory loss.
- Central pathological hallmarks: amyloid-β (Aβ) plaque and neurofibrillary tangles in brain parenchyma [10].
- AD accounts for 60-70% of the 5.5 million dementia cases across the globe [12].
- The public health crisis of the 21st century [12].

Current Situation & The Need

Validating a blood-based biomarker for AD would be groundbreaking since it would allow for routine & convenient monitoring of disease progression in patients with a simple blood test, shortening earlier diagnoses.

Methodology

Research Question & Goals

Gene ontology enrichment analysis was performed to identify which biological processes are altered in AD patients based on overrepresentation in the list of DEGs.
- Significantly enriched pathways altered in Alzheimer’s Disease:
  - The Blood-Brain Barrier Hypothesis (BBB):
    - Hypothesis: the BBB becomes increasingly permeable with aging.
    - Effects: modification of transporter cells located near the BBB, altered central nervous system (CNS) and bloodstream (e.g., immune cell trafficking, cytokine transport) [2].
  - **Significance**: changes that occur up in the brain in AD patients may be detectable in the blood.

Current Diagnosis Methods

- Can robust, reliable blood-based biomarkers unique to Alzheimer’s Disease be identified by implementing machine-learning techniques to analyze aggregated metabolic & transcriptional datasets?

Research Question & Goals

- Can robust, reliable blood-based biomarkers unique to Alzheimer’s Disease be identified by implementing machine-learning techniques to analyze aggregated metabolic & transcriptional datasets?
- Develop various machine learning classification models to differentiate between Alzheimer’s Disease and cognitively normal patients.
- Analyze the label performing machine learning models to identify candidate robust blood-based biomarkers.
- Perform gene ontology enrichment analysis to determine affected biological pathways based on the differentially expressed genes.

Data Aggregation

- Blood profiles of different study sites obtained from AD-related databases and by reaching out to study authors.
- Data pre-processing: metadata compilation, concatenation & merging, data imputation.
- Aggregated Dataset Dimensions:
  1. Transcriptional: 1358 samples, 17713 genes, 4 studies
  2. Serum Metabolomics: 1450 samples, 153 metabolites, 3 studies
  3. Plasma Metabolomics: 326 samples, 168 metabolites, 2 studies

Data Visualization

- **Purpose** to understand the features and limitations of the datasets & evaluate whether transformations are necessary.
- **t-SNE algorithm** used for dimensionality reduction & 2D visualization.

Data Analysis & Results

- Differential potential of each of the genes in the transcriptomics dataset was assessed via:
  - Mann-Whitney U Test with multiple test FDR corrections for calculation of p-value.
  - **Fold change = (mean blood transcript abundance of AD patients – mean of CN patients) / mean of CN patients**

- Highly Differentially Expressed Genes (DEGs)
  - **Criteria**: Fold change = 0.5, p-value < 0.1
  - 70 genes identified - upregulated in the blood of AD patients.

- **XGBoost Feature Importance Distribution**

- **Random Forest Feature Importance Distribution**

Discussion & Conclusions

- Successfully identified 87 genes whose blood transcript abundances show promising differentiating capability between Alzheimer’s and cognitively healthy patients across multiple study sites.
  - **Significance**: the genes represent candidate robust blood-based biomarkers that may take us a step closer to one day diagnosing Alzheimer’s with a simple blood test.
  - These genes were enriched with various biological pathways that are characteristic of AD pathogenesis.
  - **Significance**: corroborates the Blood-Brain Barrier Hypothesis and represents potential therapeutics targets to advance the quest for a cure.
  - Serum metabolites showed little to no potential as Alzheimer’s biomarkers.
  - **Limitations**: Control cohort was limited to cognitively healthy patients only.
  - Presence of co-morbidities were unaccounted for.

Future Opportunities

- Validate that a subset of the DEGs identified also have potential in differentiating AD from other types of dementias.
- Further investigate the novel DEGs to determine whether they are correlated with any comorbidities.
- Extend study to other multomics fields such as proteomics.
- Conduct further research on plasma metabolites.