

Feature Discovery And Diagnostic Modeling Of Cognitive Disease Using Explainable Artificial Intelligence

Emir Basic¹, Jakob Svensson¹, Anders Wallin², Petronella Kettunen^{2, 1}
¹ Memory Clinic, Department of Neuropsychiatry, Sahlgrenska University Hospital, Mölndal, Sweden
² Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, University of Gothenburg, Gothenburg, Sweden



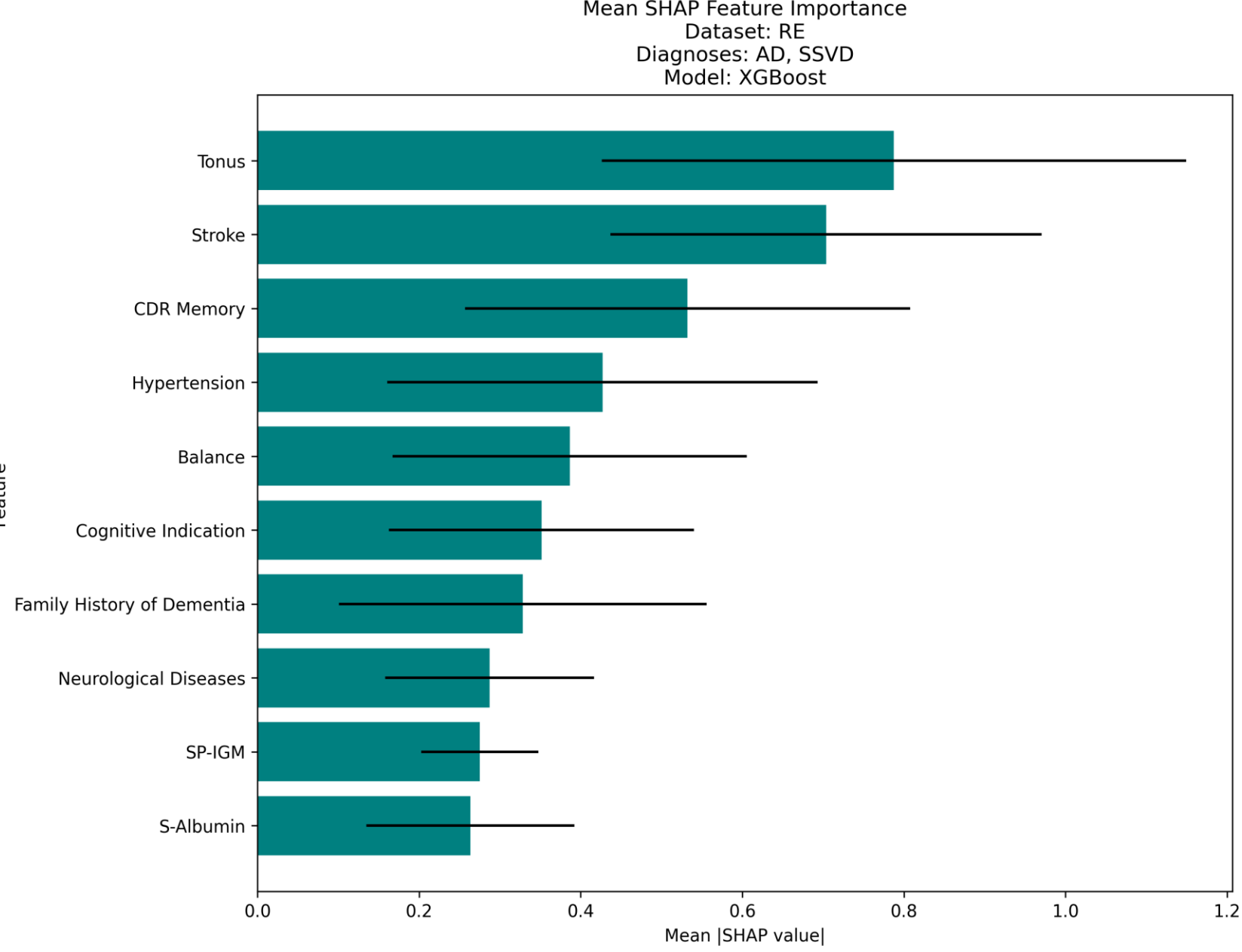
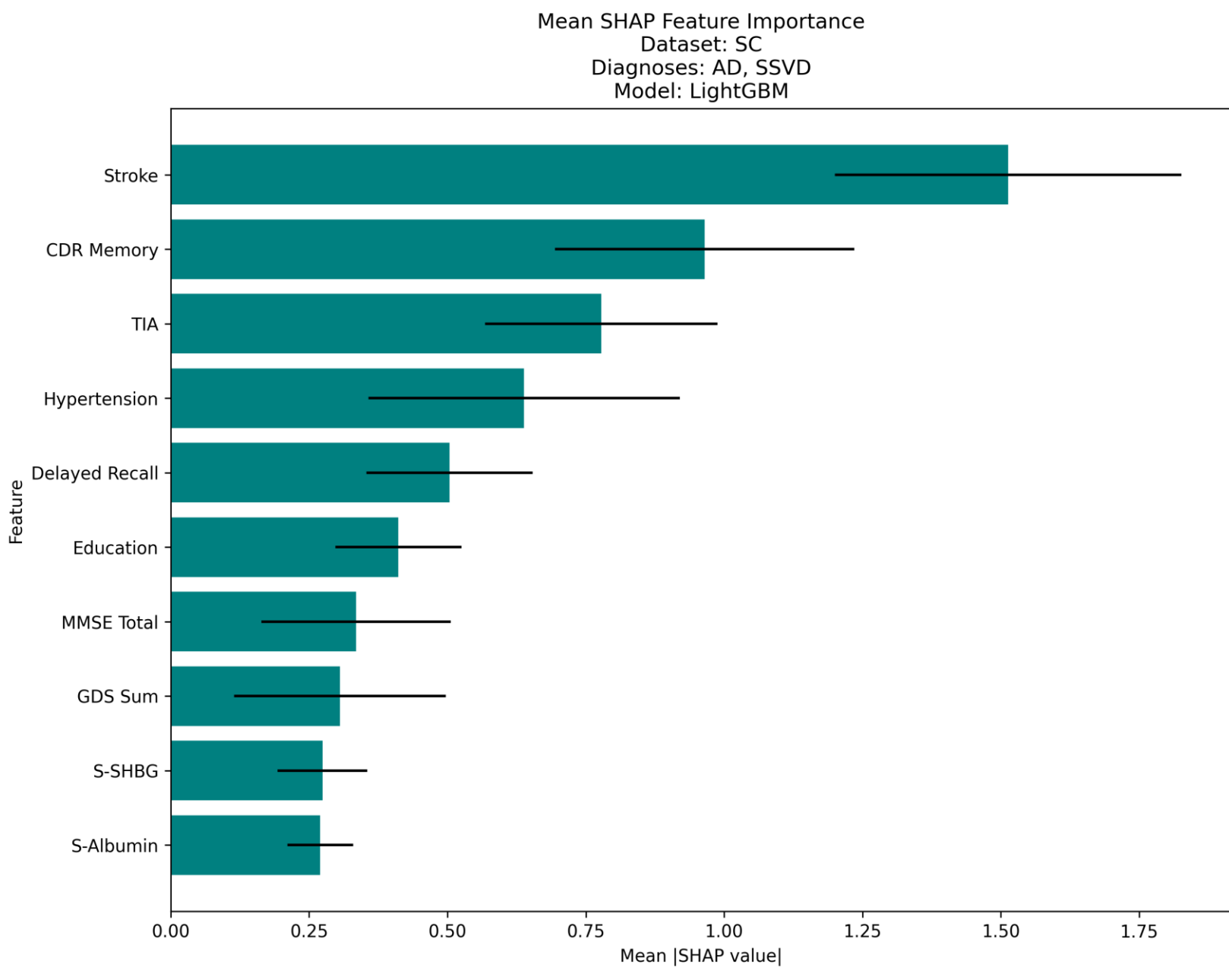
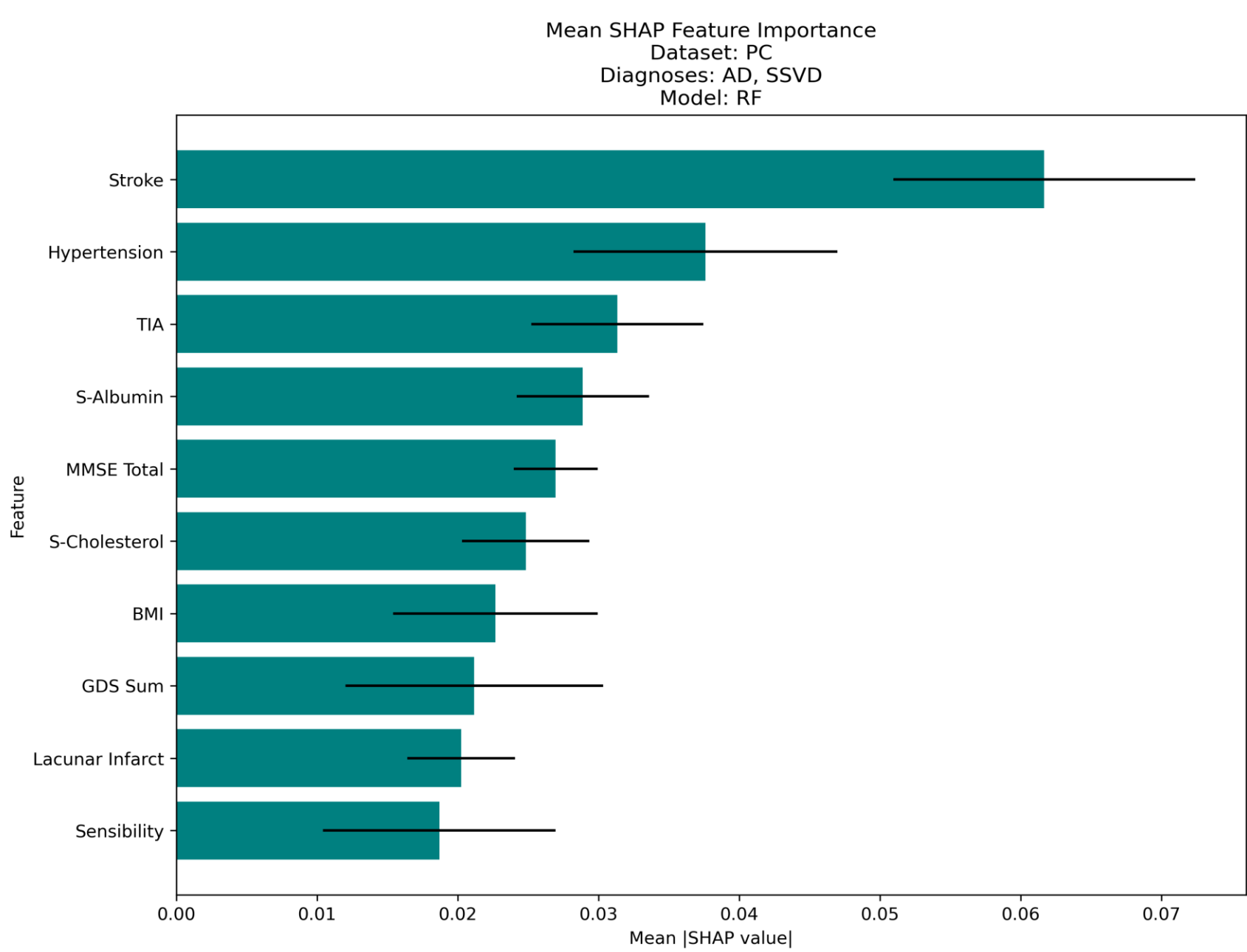
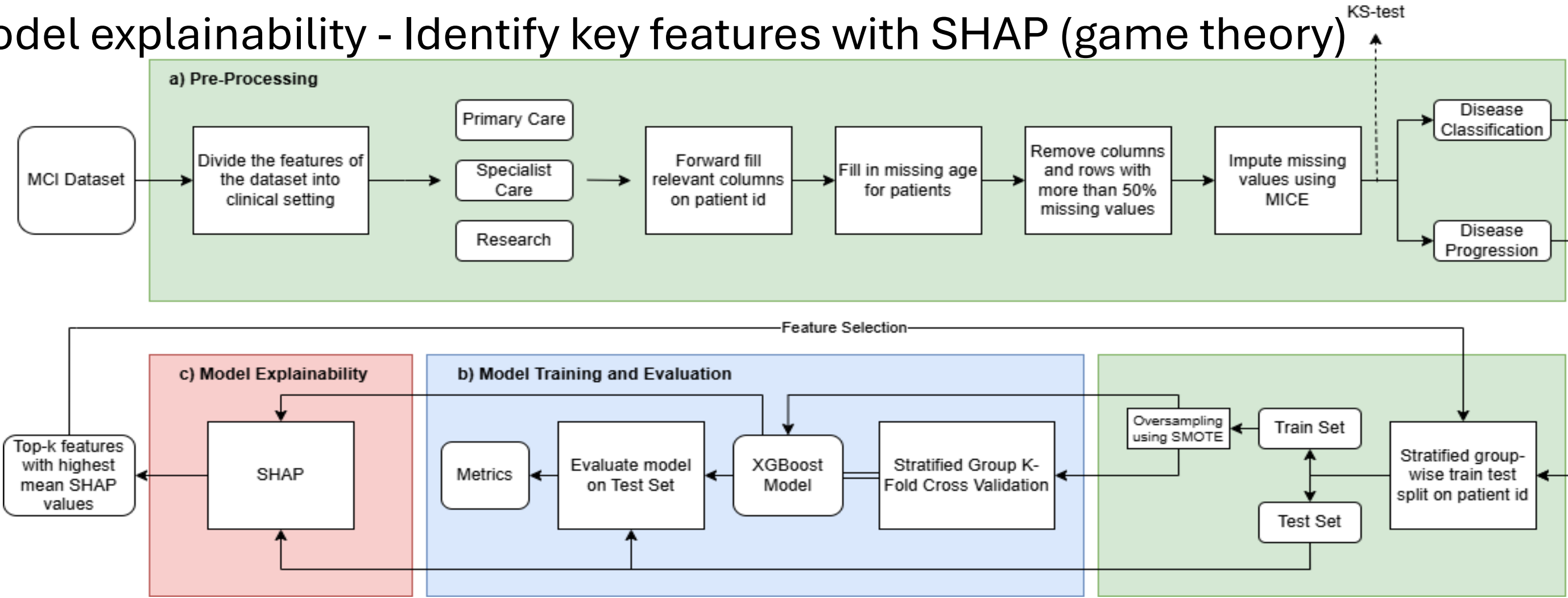
Introduction

Vascular Cognitive Disease (VCD), including subcortical small vessel disease (SSVD), remains one of the most underdiagnosed cases of age-related cognitive diseases. Given the clinical need for early, accurate, and explainable diagnosis, this study explored machine learning (ML) techniques trained on real-world data leveraged from the Gothenburg Mild Cognitive Impairment (MCI) study, to uncover key variables in SSVD diagnosis. Data was available from nearly 1000 patients with cognitive diseases and controls followed for 10 years at the Memory Clinic in Gothenburg.

The aims were to identify the best ML models for the available dataset and to identify the most important patient variables separating patients with SSVD from those with Alzheimer’s disease (AD).

Methods

- a) Data pre-processing - Imputation for missing data, division into three clinical settings
- b) Model training and evaluation - Class imbalance, ML models, evaluation metrics
- c) Model explainability - Identify key features with SHAP (game theory)



Results

- Highest performing models:**
- Random Forest (RF) in primary care (PC)
 - LightGBM in specialist care (SC)
 - XGBoost in research (RE)

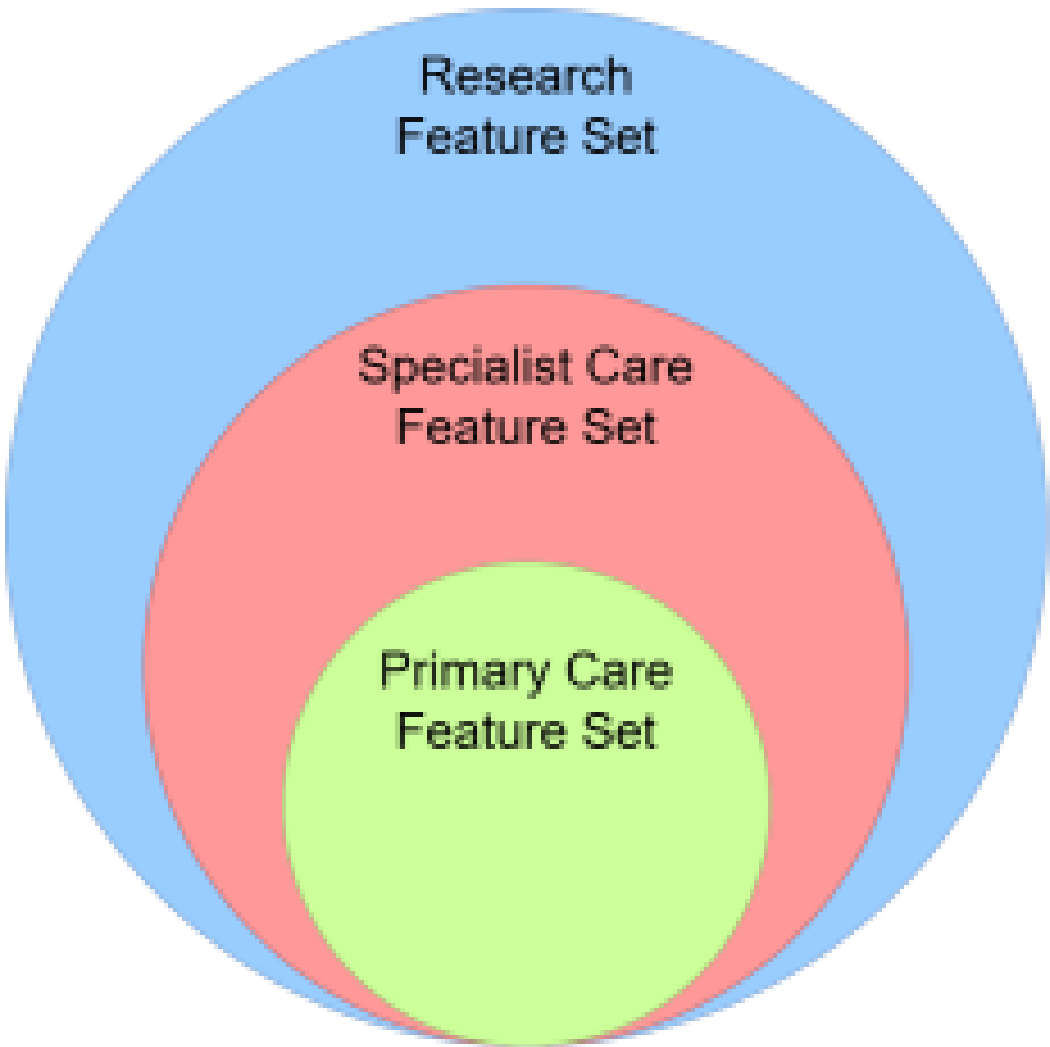
- Most prominent features:**
- Stroke
 - Hypertension
 - Transient ischemic attack (TIA)

Cohort

Data was partitioned into three subsets to reflect distinct clinical settings:

- Research data (606 variables)
- Specialist care (137 variables)
- Primary care (71 variables)

Each dataset had varying levels of feature availability



Conclusions

- We demonstrated the utility of ML techniques in developing precise and interpretable diagnostic models using real-world clinical data.
- We found an effective approach to optimize data pre-processing, a functioning classification of SSVD from AD, and identified the most important variables for SSVD.
- Our results show the potential for ML and explainable AI to enhance diagnostic accuracy in the clinic.