High-throughput DeepPRM-Stellar proteomics coupled with machine learning enables precise quantification of atherosclerosis-stroke progression biomarkers and risk prediction

Ye Liu[#a], Ouyang Hu[#b], Jingyi Wang[b] ,Yijie Qiu[b] ,Jin Xiao[b] , Xin Cheng[a], Pengyuan Yang[‡ a] ,Ningshao Xia[b] , Yueting Xiong[b]*, Quan Yuan[b]*

a. Department of Neurology, National Center for Neurological Disorders, National Clinical Research Centre for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai 200040, China.

b. State Key Laboratory of Vaccines for Infectious Diseases, Xiang An Biomedicine Laboratory, School of Public Health, National Institute for Data Science in Health and Medicine, Xiamen University, Fujian, 361102, China.

#:These authors contributed equally.

‡:This study is dedicated to the memory of Professor Pengyuan Yang (1949-2021)

†: Electronic Supplementary Information (ESI) available. See DOI:

10.1039/x0xx00000x

Table of Content:

Figure S1

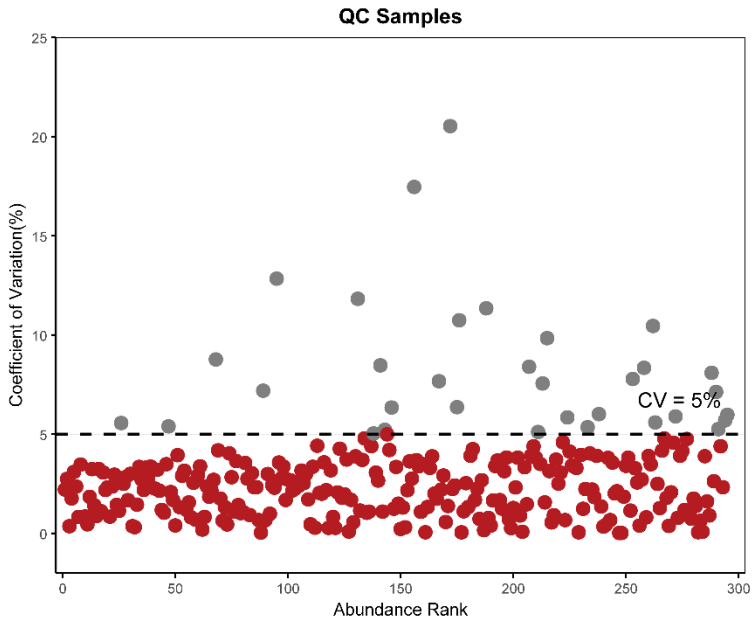


Figure S1. Quality control assessment of peptide measurements in PRM assays. Scatter plot showing coefficient of variation (CV) versus abundance rank for all QC sample measurements (n = 3 technical replicates per peptide). Each point represents an individual peptide measurement, with red points indicating high reproducibility (CV < 5%) and gray points showing greater variability (CV ≥ 5%). The dashed line at CV = 5% marks the acceptable variation threshold, with 90% of measurements demonstrating high reproducibility (red points clustered below the threshold line). The stable CV distribution across the full abundance range (0-300) confirms consistent instrument performance.

Figure S2

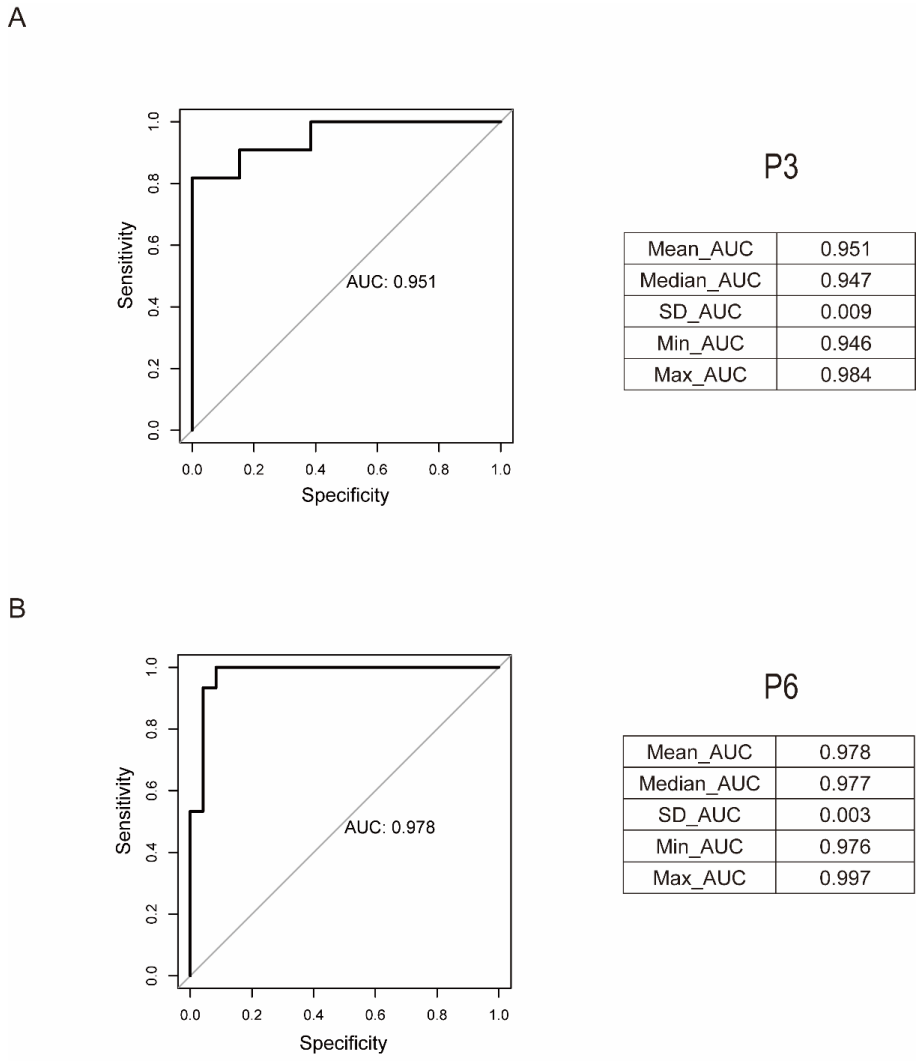


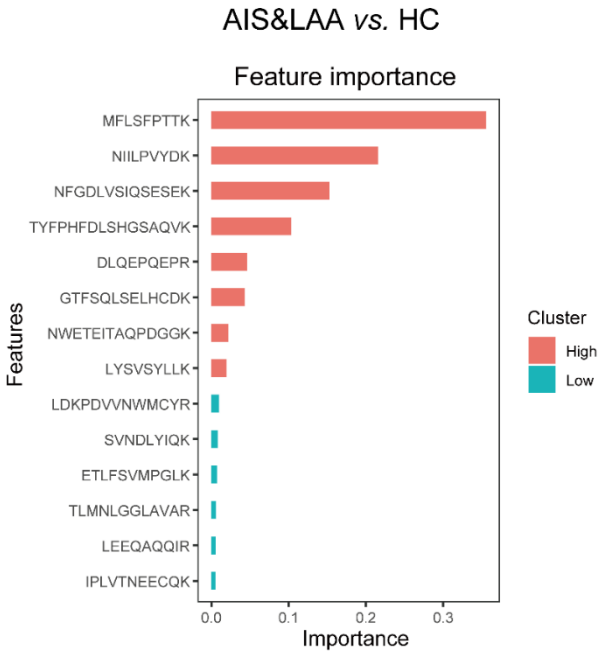
Figure S2. Model performance evaluation using Leave-One-Out Cross-Validation (LOOCV).

A). ROC curve for AIS vs LAA classification (*n* = 24). The black solid line shows the model performance (AUC = 0.951), with the diagonal representing random guessing (AUC = 0.5). The right-side table summarizes AUC statistics from LOOCV iterations: Mean_AUC = 0.951, Median_AUC = 0.947, SD_AUC = 0.009, Min_AUC = 0.946, Max_AUC = 0.984.

B). ROC curve for AIS&LAA vs HC classification (*n* = 39, AUC = 0.978). The table lists LOOCV results: Mean_AUC = 0.978, Median_AUC = 0.977, SD_AUC = 0.003, Min_AUC = 0.976, Max_AUC = 0.997. All models were trained using XGBoost .

Figure S3

A



B

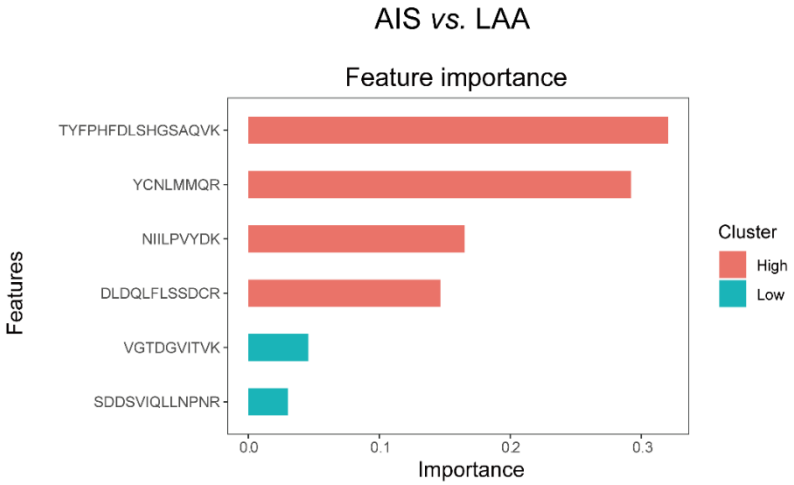


Figure S3. Feature importance analysis of candidate biomarkers.

A). Feature importance ranking for distinguishing AIS&LAA from healthy controls (HC). Proteins are sorted by their Cover-based importance scores from XGBoost.

B). Feature importance ranking for distinguishing AIS from LAA subtypes. In both panels, features are color-coded by their relative importance (red: high; blue-green: low).