Supplementary material for "Interpretable multimodal deep learning for diagnosis and symptom-onset prediction in genetic frontotemporal dementia"

Wenjie Kang[a,*], Tine Swartenbroekx[b], Myrthe F. van Haaften[a], Lize C. Jiskoot[b], Stefan Klein[a], Harro Seelaar[b], Esther E. Bron[a]

[a]*Biomedical Imaging Group Rotterdam, Department of Radiology & Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands*
[b]*Department of Neurology, Erasmus MC, Rotterdam, The Netherlands*

1. Coding of genetic variables

The multimodal models included two genetic variables: the gene affected in the family and the specific pathogenic variant recorded for that family. The observed categories are listed in Table S1. The two sets of categories are presented independently and should not be interpreted as row-wise gene–variant pairs.

2. Group-level feature importance in the testing sets

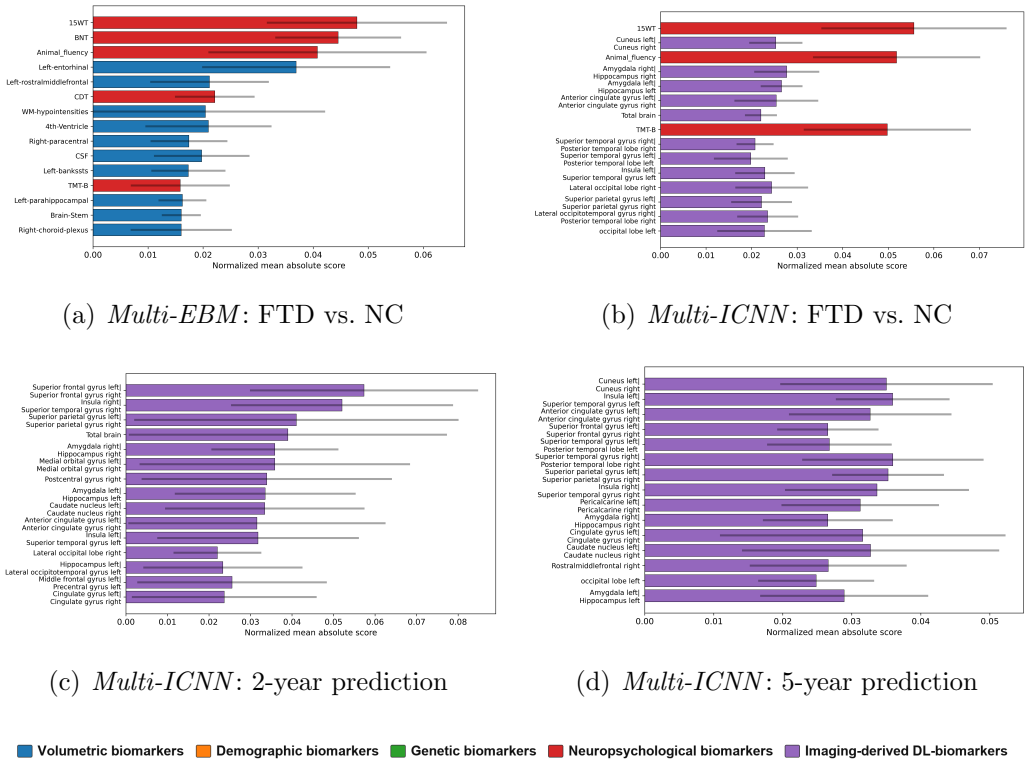
To assess whether the feature importance patterns identified in the training sets were also observed in unseen data, group-level feature importance was calculated separately in the corresponding testing sets. For each model and task, the 15 highest-ranked training-set features are displayed in the same order as in the main manuscript, allowing direct comparison of their relative

*Corresponding author: w.kang@erasmusmc.nl

Table S1: Pathogenic variants observed for each gene in the study population.

Gene	**Observed pathogenic variants**
GRN	Ser82ValfsX174; Gln125X; Q249X; Val411SerfsX1; Val8Glu; Arg493Ter; c.1179+104_1536delinsCTGA; Asp254fs
MAPT	P301L; G272V; R406W; L315R; S320F
C9orf72	GGGGCC hexanucleotide repeat expansion
TARDBP	Ile383Val

importance between the training and testing sets. Feature importance scores were averaged across the 10 repeated train–validation–test splits.



(a) *Multi-EBM*: FTD vs. NC

(b) *Multi-ICNN*: FTD vs. NC

(c) *Multi-ICNN*: 2-year prediction

(d) *Multi-ICNN*: 5-year prediction

Figure S1: Group-level feature importance estimated from the test sets. For each model and task, the 15 features with the highest mean importance in the corresponding training sets are displayed using the same feature order as in the main manuscript. The normalized feature importance scores were averaged across the 10 repeated train–validation–test splits, with error bars representing the standard deviation across splits. Colors indicate the biomarker categories shown in the shared legend.

3. Complete group-level feature importance rankings

To complement the top-15 results presented in the main manuscript, Figs. S2–S5 show the complete training-set group-level feature importance rankings. Figure S2 presents all features used by *Multi-EBM* for FTD vs. NC classification. Neuropsychological biomarkers had the largest contributions, whereas demographic biomarkers contributed comparatively little.

Figure S3 presents all features used by *Multi-ICNN* for FTD vs. NC classification. Two image patches that did not substantially overlap predefined anatomical regions are labelled "Background left" and "Background right". Figures S4 and S5 show the corresponding rankings for the 2-year and 5-year phenoconversion-prediction tasks. Neuropsychological biomarkers contributed less to these prediction tasks, whereas imaging-derived *DL-biomarkers* dominated the rankings.

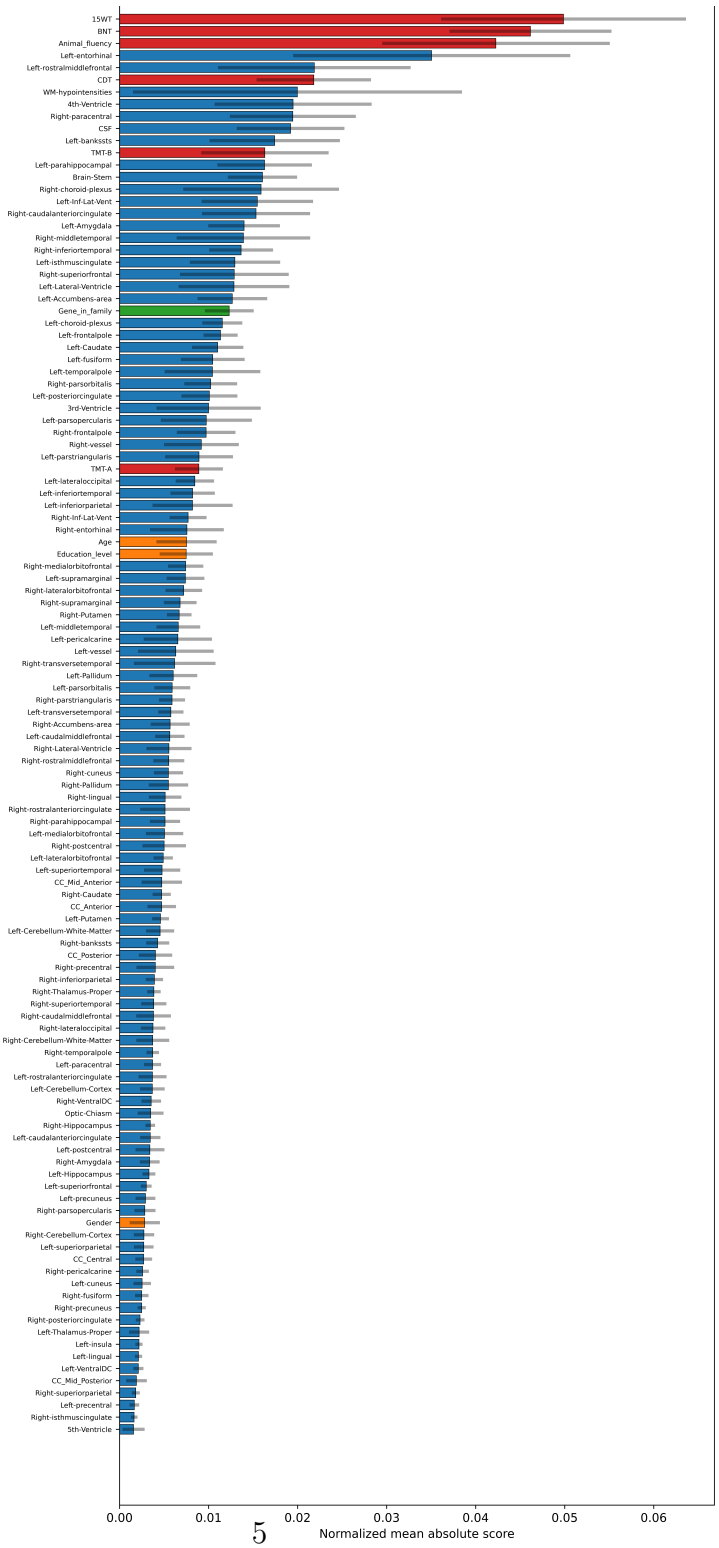


Figure S2: Complete training-set group-level feature importance ranking of *Multi-EBM* for FTD vs. NC classification. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations. The model includes 117 features.

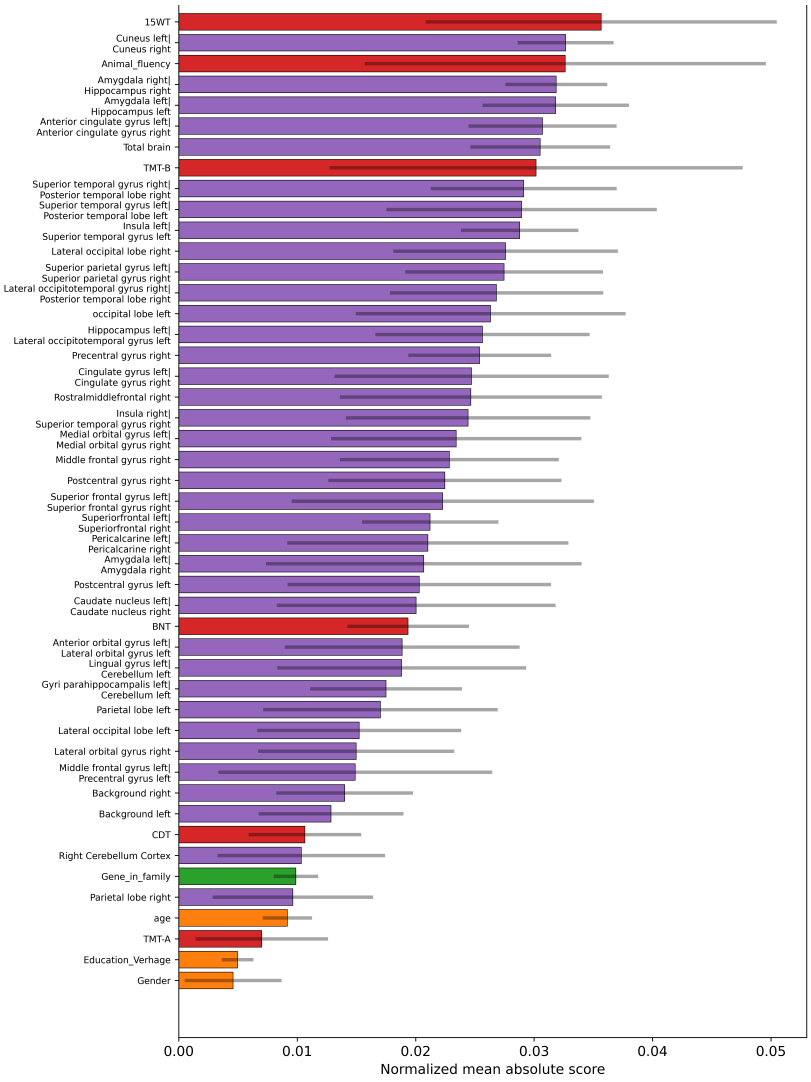


Figure S3: Complete training-set group-level feature importance ranking of *Multi-ICNN* for FTD vs. NC classification. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations. The model includes 47 features.

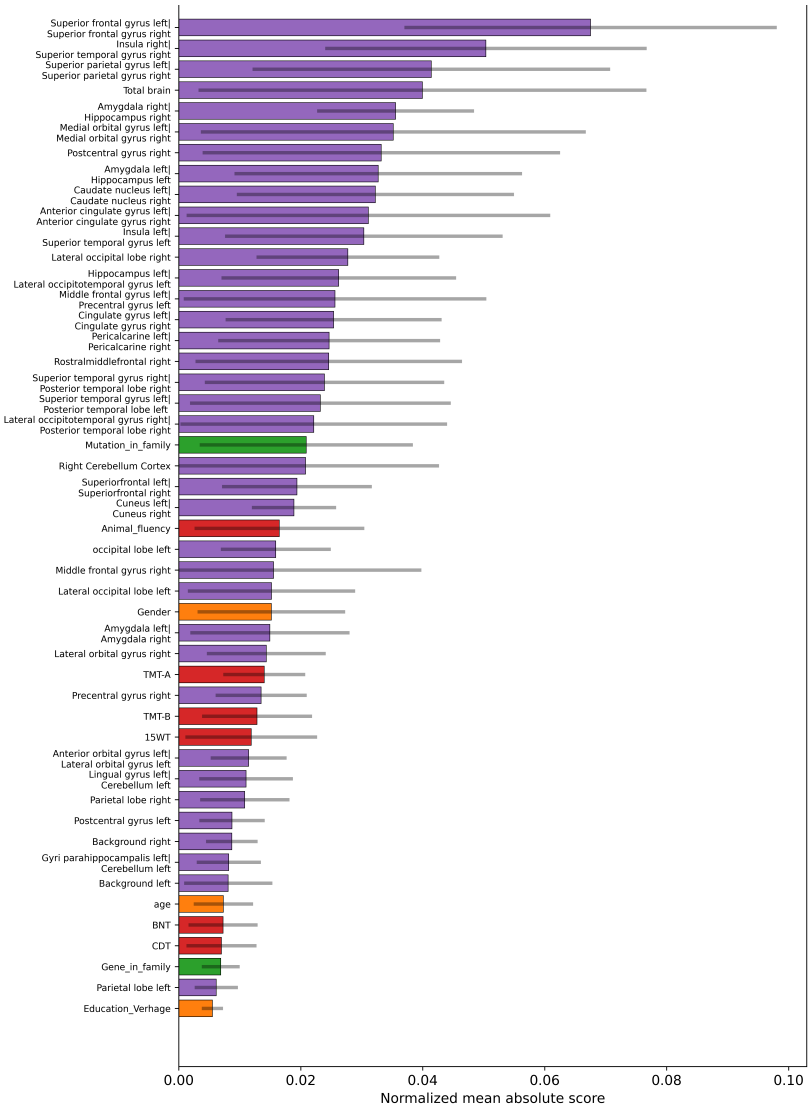


Figure S4: Complete training-set group-level feature importance ranking of *Multi-ICNN* for 2-year phenoconversion prediction. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations. The model includes 47 features.

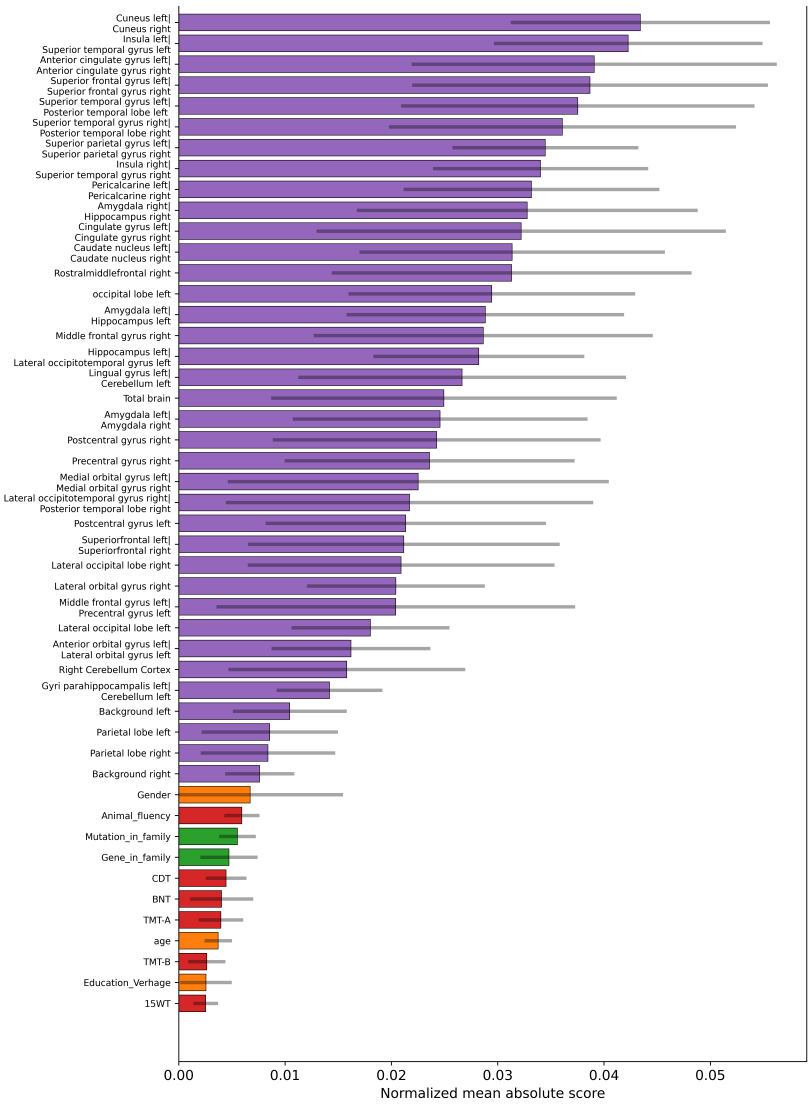
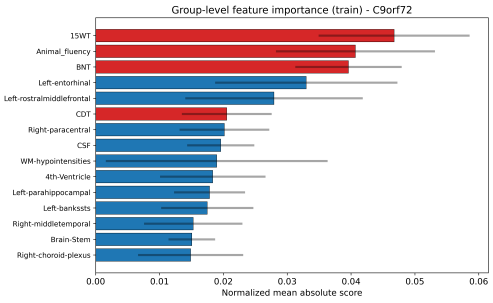


Figure S5: Complete training-set group-level feature importance ranking of *Multi-ICNN* for 5-year phenoconversion prediction. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations. The model includes 47 features.

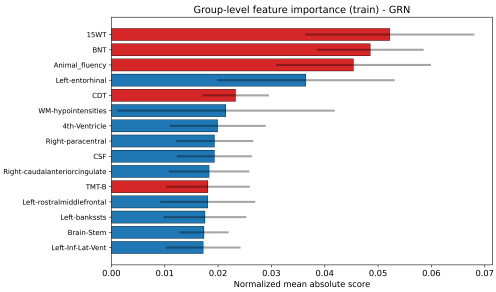
4. Group-level feature importance in genetic subgroups

To investigate potential differences across genetic subtypes, group-level feature importance was evaluated within genetic subgroups using models trained on the full cohort. For each subgroup, feature importance was computed using only samples belonging to that genetic group. For FTD vs. NC classification, each subgroup included mutation carriers and non-carriers from families associated with the same gene. For the phenoconversion-prediction tasks, feature importance was computed separately for mutation carriers within each subgroup. The *TARDBP* subgroup was not analysed separately because of its very small sample size.

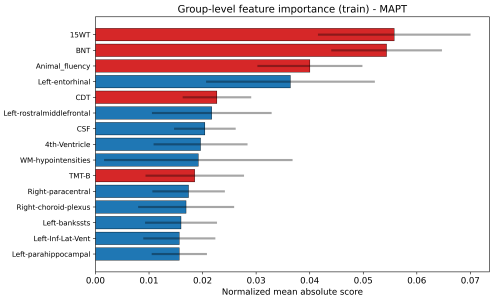
The top-15 subgroup results for *Multi-EBM* are shown in Fig. S6; the corresponding *Multi-ICNN* results are shown in Figs. S7–S9. Although the relative ranking of individual biomarkers varied across genetic subgroups, no clear subgroup-specific pattern was consistently observed across tasks. These results should therefore be considered exploratory.



(a) *C9orf72* subgroup

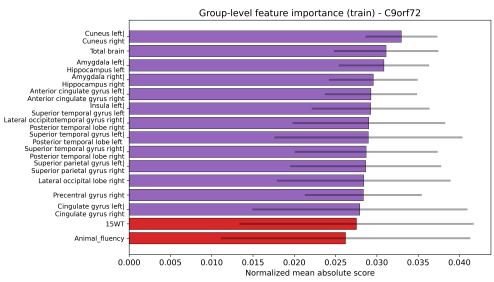


(b) *GRN* subgroup

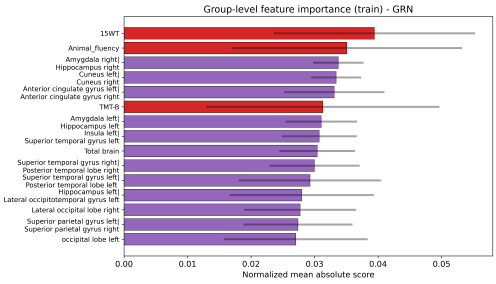


(c) *MAPT* subgroup

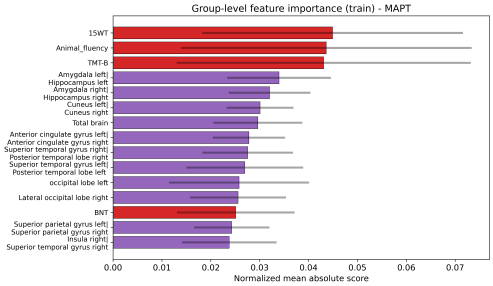
Figure S6: Group-level feature importance of *Multi-EBM* within genetic subgroups for FTD vs. NC classification. Within each subgroup, the 15 highest-ranked features are shown. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations.



(a) *C9orf72* subgroup

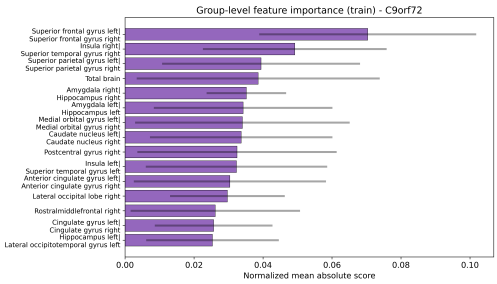


(b) *GRN* subgroup

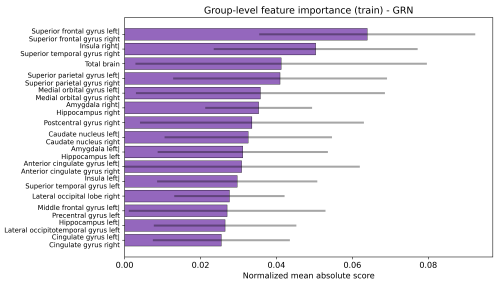


(c) *MAPT* subgroup

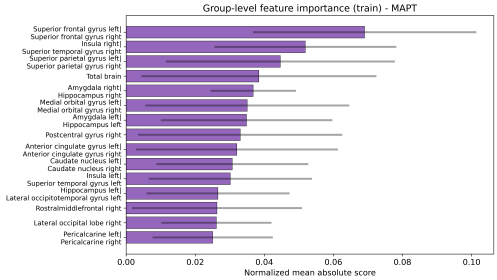
Figure S7: Group-level feature importance of *Multi-ICNN* within genetic subgroups for FTD vs. NC classification. Within each subgroup, the 15 highest-ranked features are shown. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations.



(a) *C9orf72* subgroup

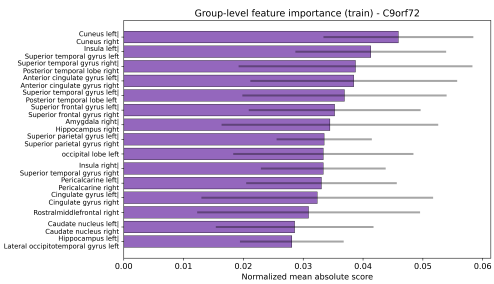


(b) *GRN* subgroup

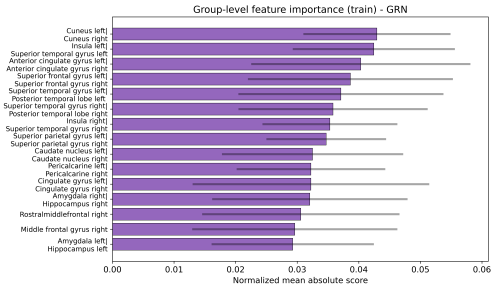


(c) *MAPT* subgroup

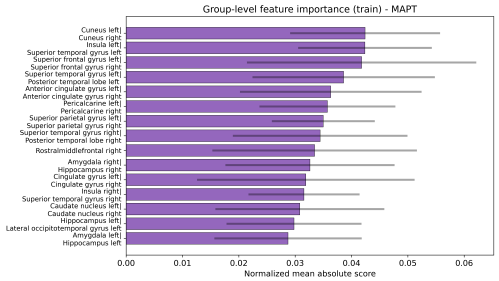
Figure S8: Group-level feature importance of *Multi-ICNN* within genetic subgroups for 2-year phenoconversion prediction. Within each subgroup, the 15 highest-ranked features are shown. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations.



(a) *C9orf72* subgroup



(b) *GRN* subgroup



(c) *MAPT* subgroup

Figure S9: Group-level feature importance of *Multi-ICNN* within genetic subgroups for 5-year phenoconversion prediction. Within each subgroup, the 15 highest-ranked features are shown. feature importance scores are averaged over 10 iterations, with error bars indicating the standard deviation across iterations.