INDIANA UNIVERSITY
School of Medicine

FENS Forum 2026 · Barcelona · 6 to 10 July

# SNP prioritization using deep learning for Alzheimer's disease risk prediction: Validation in 5,570 independent individuals

**Taeho Jo**
Indiana Alzheimer Disease Research Center and Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA tjo@iu.edu

## Background

Alzheimer's disease (AD) is the most common dementia. With anti-amyloid therapies, early risk prediction before symptom onset is increasingly important.

Germline genomic data is fixed from birth, enabling lifelong risk prediction. AD heritability is 50 to 70 percent, and the APOE region accounts for the largest portion. APOE ε4 raises AD risk about 3-fold in heterozygotes and up to 15-fold in homozygotes. Whole-genome sequencing (WGS) is high-dimensional. The genome has both local interactions between nearby loci and long-range functional relationships, which raw-input methods struggle to capture.

Most genome-based AD models reach only modest performance, and typically analyze local patterns or functional annotations in isolation.

## Objective and contributions

- Develop DuAL-Net (Dual Approach Local-global Network), a hybrid framework that captures local SNP interactions and global functional relationships at the same time to prioritize AD-associated SNPs.
- Show that annotation-guided scoring improves identification of disease-relevant variants.
- Demonstrate that the rankings generalize to a fully independent population.
- Provide a generalizable framework extensible to other genomic regions and diseases.

## Methods

### Cohorts

- Training. ADNI and ADNI-WGS-2 with ADSP Follow-Up Study. n=1,050 (607 AD, 443 CN) after excluding MCI.
- Validation. ADSP Alzheimer's Disease Centers (ADC), Release 5, NHW. n=5,570 (3,837 AD, 1,733 CN), fully independent of training.

### Data and preprocessing

- Region. APOE ±50 kb on chromosome 19, 14,094 SNPs, GRCh38. Standard ADSP QC and KNN imputation (k=5).
- Annotation. Ensembl release 108 covering genomic context, gene biotype, variant consequence, and clinical significance.

### DuAL-Net framework

- Local. Nonoverlapping 100-SNP windows. Random Forest and TabNet combined by out-of-fold stacking with a logistic regression metamodel. Holdout accuracy gives each window a local score.
- Global. SNP groups defined by shared annotations are scored the same way. Each SNP gets a global score averaged over its groups.
- Combined score = α × local + (1 − α) × global. α is optimized by nested CV in the inner loop, with outer 5-fold stratified CV for evaluation.
- Prioritization yields top and bottom panels of 100, 500, and 1,000 SNPs. ADNI rankings are transferred to ADC by genomic position for external validation. Baselines are logistic regression on all SNPs and univariate F-test selection with Random Forest.

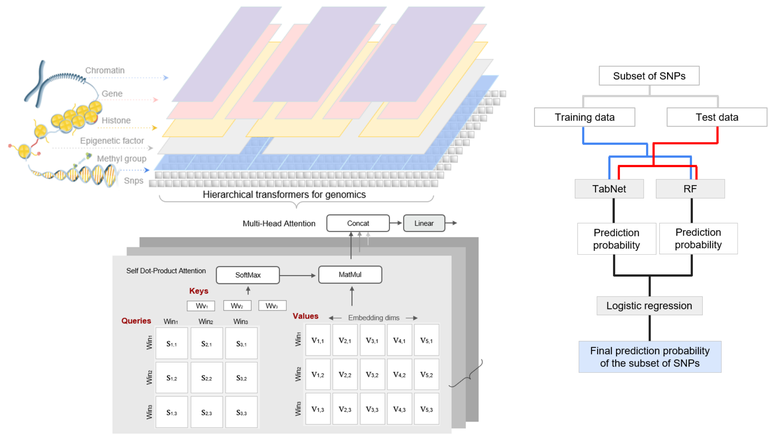


*Figure 1. DuAL-Net architecture. Hierarchical transformers with multi-head self-attention encode each SNP subset with genomic and epigenomic context. TabNet and Random Forest predictions are combined by logistic regression into a final probability.*

### Table 1. Cohort demographics and APOE distribution

| Characteristic | Train CN (443) | Train AD (607) | Valid CN (1,733) | Valid AD (3,837) |
|---|---|---|---|---|
| Age, mean (SD) | 72.6 (6.3) | 74.0 (7.2) | 82.3 (5.7) | 68.1 (10.6) |
| Female, n (%) | 208 (47.0) | 352 (58.0) | 642 (37.0) | 1,780 (46.4) |
| ε2/ε2 | 2 (0.5) | 1 (0.2) | 8 (0.5) | 3 (0.1) |
| ε2/ε3 | 57 (12.9) | 19 (3.1) | 162 (9.3) | 74 (1.9) |
| ε2/ε4 | 7 (1.6) | 14 (2.3) | 18 (1.0) | 88 (2.3) |
| ε3/ε3 | 257 (58.0) | 194 (32.0) | 1,048 (60.5) | 1,239 (32.3) |
| ε3/ε4 | 110 (24.8) | 278 (45.8) | 454 (26.2) | 1,850 (48.2) |
| ε4/ε4 | 10 (2.3) | 101 (16.6) | 35 (2.0) | 526 (13.7) |
| ε4 carrier, n (%) | 127 (28.7) | 393 (64.7) | 507 (29.3) | 2,464 (64.2) |

### Table 2. Model component comparison

| Model | AUC | 95% CI | Accuracy |
|---|---|---|---|
| Global-only | 0.652 | 0.624 to 0.681 | 0.610 |
| Local-only | 0.682 | 0.623 to 0.740 | 0.653 |
| **Integrated (α=0.66)** | **0.698** | **0.659 to 0.737** | **0.660** |

## Conclusions

- Integrating local genomic context with global functional annotations improves SNP prioritization for AD.
- ADNI-trained rankings generalize to an independent cohort of 5,570 individuals.
- A foundation for genome-wide extension and multimodal risk prediction with clinical and imaging biomarkers.
- APOE region only, no PRS comparison, AUC 0.698 is not yet sufficient for standalone clinical use, and validation in diverse ancestries is needed.

## Key results

### Local and global analysis

- 140 windows, mean accuracy 0.612 (range 0.530 to 0.762), 56 windows at 0.60 or higher.
- Top annotation groups: pathogenic 0.695, uncertain significance 0.695, drug response 0.669, protein-coding 0.630.

### Integrated model

- Optimal α = 0.66. Integrated AUC = 0.698 (95% CI 0.659 to 0.737), accuracy 0.660.
- Significantly outperformed global-only (p = 0.010) and both baselines (logistic regression p = 0.019, univariate p = 0.026).

### Established variants and independence

- rs429358 (APOE ε4) ranked in the top 100 in all 5 folds. rs7412 (APOE ε2) was also identified.
- rs7412 showed very low LD with rs429358 ($r^2$ = 0.035), confirming independent signals.
- After removing rs429358 and 34 LD proxies, top-ranked SNPs still beat bottom-ranked SNPs, signal beyond APOE ε4.

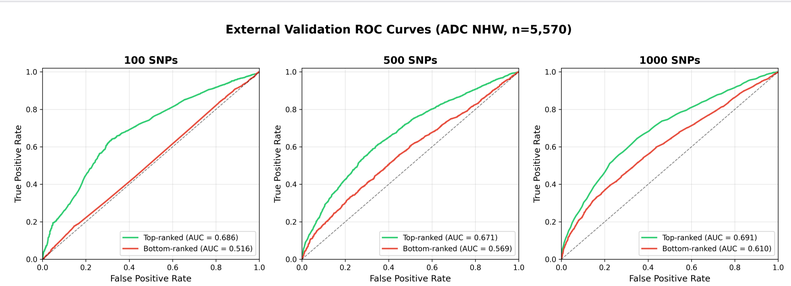


*Figure 2. External validation in the independent ADC cohort (NHW, n=5,570). Top-ranked SNPs (green) consistently exceed bottom-ranked SNPs (red) across 100, 500, and 1,000 SNP panels.*

### Table 3. Top-ranked versus bottom-ranked SNP panels

| Cohort | SNP subset | AUC 100 | AUC 500 | AUC 1,000 |
|---|---|---|---|---|
| Training (1,050) | Top-ranked | 0.695 | 0.649 | 0.666 |
| Training (1,050) | Bottom-ranked | 0.479 | 0.481 | 0.527 |
| **Validation (5,570)** | **Top-ranked** | **0.686** | **0.671** | **0.691** |
| Validation (5,570) | Bottom-ranked | 0.516 | 0.569 | 0.610 |

## Resources and funding

- Code: github.com/taehojo/DuAL-Net
- Web server: jolab.ai/dualnet
- Reference: https://doi.org/10.34133/csbj.0010.
- Funding: Alzheimer's Association AARG 22-974053 and NIH P30 AG072976, U19 AG024904, U01 AG068057, U19 AG074879, U01 AG072177, U24 AG074855.