G2DBridge: A Multimodal Framework Linking Genetics to Disease through Imaging Intermediates

Okan Bilge Ozdemir[1], Raelynn Chen[1], Olivia Wu[1], Ruowang Li[1]

1. Cedars-Sinai Health Sciences University, Department of Computational Biomedicine, Los Angeles, CA, USA

Abstract

Genetic-based risk prediction is becoming increasingly available for a wide range of common diseases thanks to the growth of large-scale biobanks and population-scale genetic studies. However, despite widespread availability, genetic risk prediction typically offers only modest predictive power, as complex biological processes that mediate disease manifestation are often not fully captured by genetic variation alone. In contrast, imaging-based prediction can achieve substantially higher predictive accuracy for many diseases but remains costly and unavailable for the majority of populations. Here, we present G2DBridge, a predictive machine learning framework that links genetic variation to disease risk by modeling imaging-derived phenotypes as intermediate traits. Using Alzheimer's Disease Neuroimaging Initiative (ADNI) for training and Alzheimer's Disease Sequencing Project (ADSP) for external validation, G2DBridge outperforms state-of-the-art polygenic risk scores in Alzheimer's Disease (AD) prediction while relying on the same genetic data in the target data, providing additional predictive power without requiring new data modalities. The framework generalizes across cohorts, enables risk stratification with clear separation between cases and controls, and highlights biologically meaningful regions consistent with AD pathology. This approach offers a scalable strategy for applying genetically inferred imaging features in precision medicine.

Keywords: Genomics; PRS; DeepLearning; Image Derived Phenotypes, Alzheimer's disease, Neuroimaging

Introduction

Accurate prediction of disease risk is essential for early diagnosis, patient stratification in clinical trials, and targeted prevention strategies[1–5]. Genetic risk prediction models, such as polygenic risk scores (PRS), are widely used due to the increasing availability of large-scale genetic data from research biobanks and clinical repositories[5–9]. However, traditional PRS methods typically assume linear genetic effects[10–13], which limit their predictive power and fail to capture individual variability, especially when key intermediate biological processes are missing from the models[14–16].

Imaging phenotypes, derived from modalities such as neuroimaging, cardiac MRI, PET, or CT scans, have emerged as powerful biomarkers for disease prediction[17–21]. These features capture detailed structural and functional characteristics of biological systems, providing richer phenotypic information than genetic data alone[18,22]. Across diverse diseases, including neurodegenerative, cardiovascular, and metabolic disorders, imaging-based models often outperform purely genetic models in predictive accuracy[23–25]. However, the limited availability of imaging data in large-scale or routine clinical settings, due to cost, accessibility, and technical challenges[25–27], constrains the widespread adoption of imaging-based approaches.

Existing studies have explored integrating genetic, imaging, and disease modalities to improve predictive performance[2,17,20,22,23]. Large-scale genome-wide association studies (GWAS) and derived PRS remain state-of-the-art for genetics-driven risk prediction[9,15,28]. Meanwhile, machine learning models using imaging data have demonstrated markedly superior prediction compared to genetics-only models[19,21,23,24], but are typically trained on small imaging datasets, which restricts their broader applicability. Multi-modal

[1] *Correspondence: ruowang.li@csms.edu

integration methods that jointly model genetics and imaging have been proposed, particularly in neurodegenerative diseases[20,22,29,30], but they generally require full access to all data modalities in the target cohort, limiting real-world usability[21]. Other approaches leverage imag-derived phenotypes (IDPs) as intermediate traits[31] but often rely on linear PRS frameworks[32], which lack the flexibility to capture non-linear and heterogeneous effects in imaging phenotypes. In contrast, deep neural networks (DNNs) provide a promising alternative[11–13], as they can naturally capture complex, non-linear relationships among genetic and imaging variables without the need to specify interaction terms explicitly. DNNs have achieved state-of-the-art performance across a range of scientific domains, including image recognition, segmentation, and object detection, and are increasingly being adopted in biomedical prediction tasks[33].

To overcome these challenges, we propose G2DBridge, a generalizable predictive modeling framework that leverages imaging-derived phenotypes as an intermediate bridge between genetic variation and disease outcomes. G2DBridge introduces several innovations compared to existing multi-modal data integration methods. First, G2DBridge learns latent imaging-derived representations using limited imaging data and transfers this knowledge to datasets where only genetic data are available, thereby improving disease prediction even when imaging is absent. Second, unlike traditional genetic risk prediction models that assume additive linear effects, G2DBridge leverages deep learning to capture non-linear and high-order genetic effects as mediated through imaging modalities on disease outcomes. By introducing imaging-derived intermediate representations, it reduces the risk of overfitting when modeling complex genetic interactions in datasets with limited imaging availability. Third, G2DBridge includes a feature extraction module that learns low-dimensional embeddings from different imaging modalities and integrates them into a unified framework. This enables multi-imaging integration while maintaining scalability to settings with incomplete imaging data.

In this study, we applied G2DBridge to Alzheimer's disease (AD), a neurodegenerative disorder with well-established genetic and neuroimaging biomarkers[34–36]. We trained and validated our models using matched genetic data and IDPs, including FreeSurfer (FS) for brain volumes and cortical thickness[26], Automatic Segmentation of Hippocampal Subfields (ASHS) for hippocampal regions[27], and diffusion tensor imaging (DTI) for white matter pathways[29,30] from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We further performed external validation using the Alzheimer's Disease Sequencing Project (ADSP), where only genetic data were available. Our results demonstrate that G2DBridge substantially improves AD risk prediction compared to conventional PRS, when applied to the same genetic-only inputs. This integrative framework leverages limited imaging data to improve genetic risk prediction, enabling reliable and cost-effective disease risk assessment in large-scale biobanks and clinical settings.

Results

Overview of the proposed approach

We introduce G2DBridge, a generalizable predictive framework that links genetic variation to disease risk by modeling intermediate traits, specifically IDPs from genotype data (Figure 1). The framework leverages matched genetic, imaging, and clinical outcome data to learn cross-modal predictive relationships, effectively bridging genetics to disease through imaging-derived features.

In the training phase, where all modalities are available, G2DBridge learns an intermediate representation of IDPs that connects genetic variants to disease risk. This process consists of three stages: (i) training neural networks(NN) to predict IDPs directly from genetic variants, enabling the generation of imaging-like features even in cohorts without neuroimaging data. (ii) extracting low-dimensional latent representations of each IDP from hidden layers to provide compact yet expressive embeddings. and (iii) integrating these embeddings through a downstream classification network to predict disease outcomes. In the application phase, only genetic data from the target dataset are required. The pretrained networks are transferred to infer imaging features and used for disease prediction, thereby improving risk prediction without requiring imaging data.

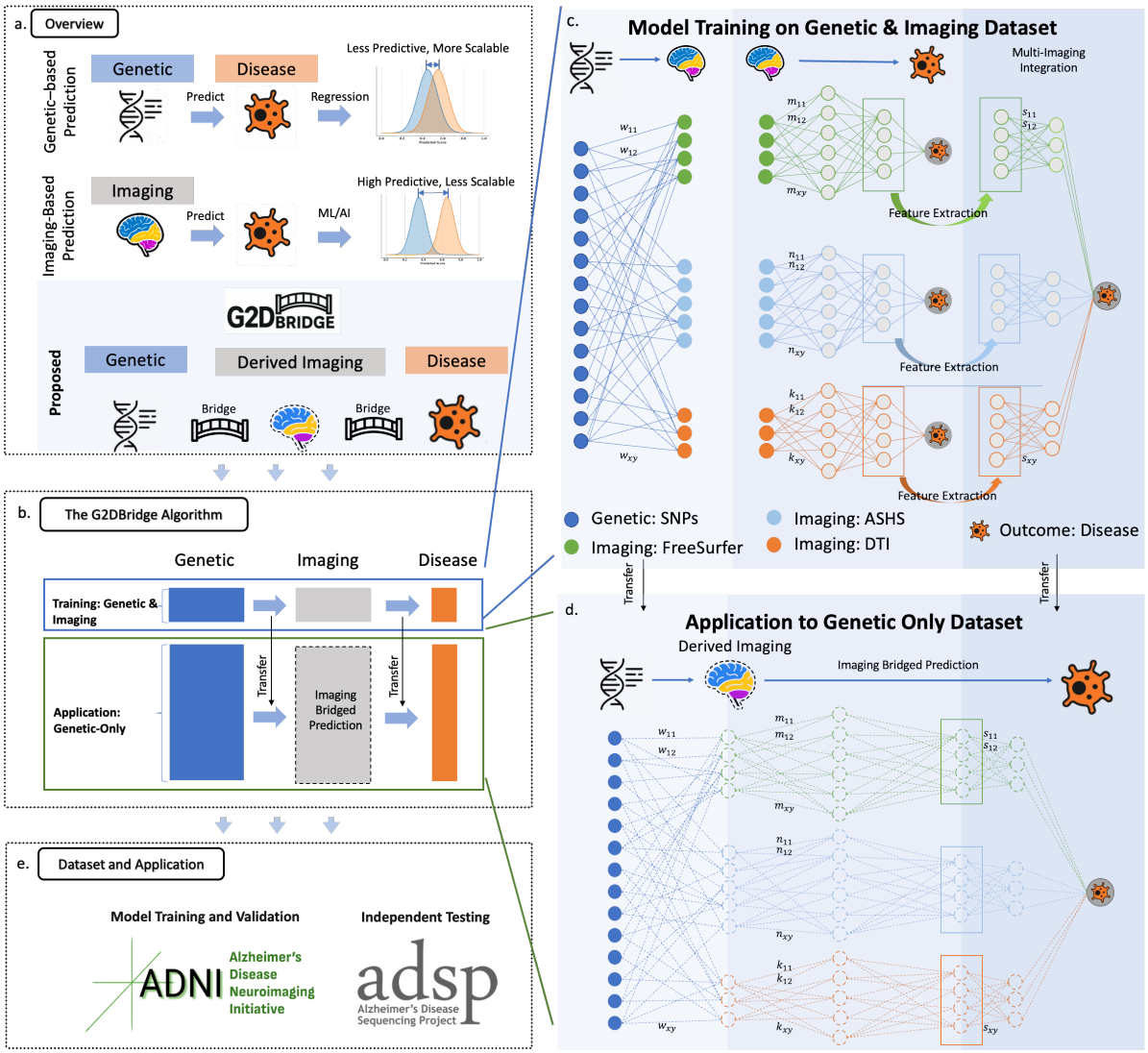


Figure 1 Overview of the G2DBridge framework. *(a) Schematic illustration of genetics-based prediction, imaging-based prediction, and the integrative G2DBridge approach. (b) Architecture of the G2DBridge algorithm, showing training with genetic and imaging data and application with genetic data only. (c) Model training procedure using genetic and imaging data. (d) Application of the trained model to genetic-only datasets for disease prediction. (e) Datasets used for training, validation, and testing: ADNI and ADSP.*

AD Classification with IDPs

To evaluate the utility of IDPs for AD prediction, we compared five machine learning models (Logistic Regression, Random Forest, Lasso, XGBoost, and NN) across three IDP sources (ASHS, DTI, and FS). Figure 2-a summarizes the resulting AUC for each model-IDP combination, as well as for an integrated IDP model based on a NN.

In the ASHS and DTI analyses, XGBoost achieved the highest performance with AUCs of 0.838 ± 0.100 and 0.952 ± 0.036, respectively. In FreeSurfer, a neural network yielded the best performance with an AUC of 0.922 ± 0.024, followed by Lasso with 0.914 ± 0.022. Linear models, including logistic regression and Lasso, underperformed compared to the more flexible machine learning models, with 0.8%–10.5% lower AUC compared to the best model for each IDP. The multi-IDP integrated model achieved the highest

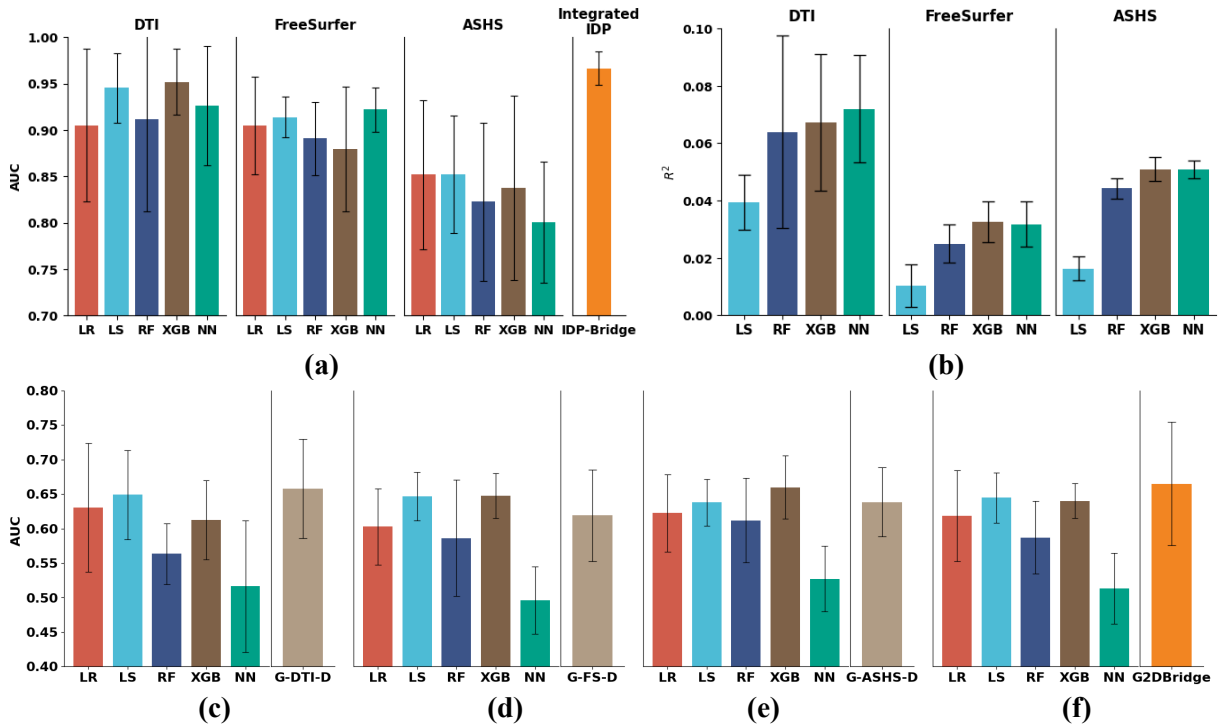


Figure 2. Performance of IDP based AD classification, genetic IDP prediction, and IDP-mediated G2DBridge models.

(a) *AD classification using individual and integrated IDPs. AUC performance of five machine learning models, LR, RF, Lasso (LS), XGBoost and NN, across three IDP sources (DTI, FreeSurfer, ASHS). Error bars represent standard deviations from five-fold cross-validation. The integrated IDP model (IDP-Bridge) is shown on the right.*
(b) *Genetic prediction of IDPs. Explained variance (R^2) of four machine learning models, LS, RF, XGBoost, and NN for three IDP sources (DTI, FreeSurfer, ASHS). Error bars represent standard deviations from five-fold cross-validation.*
(c–f) *Comparison of individual IDP-mediated models and the G2DBridge framework. Mean AUC values with standard deviations across five-fold cross-validation. Panels show* ***(c)*** *G-DTI-D,* ***(d)*** *G-FS-D,* ***(e)*** *G-ASHS-D and* ***(f)*** *proposed G2DBridge, each benchmarked against genetic-only baselines (LR, RF, LS, XGB and NN).*
Error bars indicate 95% confidence intervals (CI) across folds.

accuracy with an AUC of 0.967 ± 0.018, showing that combining multiple IDPs substantially improves predictive power and that structural IDPs are robust predictors of AD diagnosis, consistent with prior studies[37–40].

Prediction of IDPs from Genetic Data

We next evaluated the prediction of IDPs from genetic variants by comparing four machine learning methods using explained variance (R^2) and mean squared error (MSE). Figure 2-b summarizes the R^2 results across models and IDPs, with corresponding MSE results in Supplementary Figure 1.

Across the three IDP sources, NN consistently provided the best balance between variance explained and low error. For DTI, NN attained the highest R^2, 0.072 ± 0.019, while XGBoost (0.067 ± 0.024) and random forest (0.064 ± 0.034) followed with greater variability (Supplementary Figure 2). Similar patterns were observed in FreeSurfer and ASHS, where XGBoost achieved the highest R^2 (0.033 ± 0.007 and 0.051 ± 0.004, respectively), and NN were close behind (0.032 ± 0.008 and 0.051 ± 0.003) while minimizing error. In contrast, Logistic regression and the Lasso model trailed with lower R^2 and higher MSE. Overall, these findings establish NN as the most stable method across all three IDPs, and these results demonstrate that

IDPs can both be predicted from genetic data and used to classify AD, motivating the evaluation of an integrated framework that leverages both components.

Figure 2c–f shows the performance of each IDP-mediated model compared with the genetic-only baseline. The baseline models achieved moderate performance, with mean AUCs generally in the 0.56–0.66 range, and the strongest baseline differing by IDP (DTI: LS = 0.649, FreeSurfer: XGB = 0.648, ASHS: XGB = 0.659, all-IDP baseline: LS = 0.645). In contrast, the IDP-mediated models (rightmost columns) showed mixed results across IDPs. G-DTI-D improved slightly over the best genetic-only model (AUC = 0.658 ± 0.082, +1.4% vs the best baseline), whereas G-FS-D (AUC = 0.619 ± 0.076) and G-ASHS-D (AUC = 0.638 ± 0.057) were lower than their strongest genetic-only counterparts. The corresponding area under the precision–recall curve (AUPRC) results are provided in Supplementary Figure 3.

In the multi-IDP setting, G2DBridge combined latent representations from all three IDPs using an integrated neural network architecture. This integrated model achieved an AUC of 0.665 ± 0.102, exceeding the genetic-only baselines by +7.47% (LR), +13.22% (RF), +3.12% (Lasso), +3.87% (XGB), and +29.53% (NN). Its performance was higher than G-DTI-D (AUC = 0.658), G-ASHS-D (AUC = 0.638), and G-FS-D (AUC = 0.619), indicating that integrating multiple IDPs can provide additional predictive power beyond individual IDP-mediated models in this setting.

External Validation on ADSP Dataset

To evaluate generalizability, models trained on the ADNI dataset were externally validated on the independent ADSP cohort. As a benchmark, we first examined nine PRS models from the PGS Catalog (Figure 3, left). These models, all trained on datasets independent of ADSP, achieved AUCs ranging from 0.51 to 0.63, with the highest accuracy observed for PGS004146 (AUC = 0.629). Despite representing state-of-the-art PRS benchmarks, overall performance remained modest.

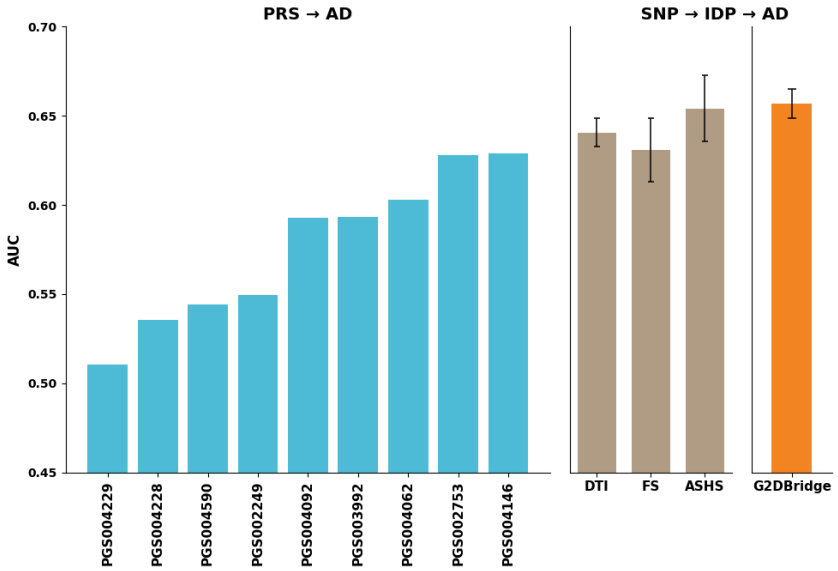


Figure 3 External validation on ADSP. *(Left) Performance of nine polygenic risk score (PRS) models from the PGS Catalog for AD classification in ADSP; model identifiers (e.g., PGS004146) correspond to catalog entries. (Right) Performance of IDP-mediated models (DTI, FreeSurfer, ASHS) and the integrated G2DBridge model trained on ADNI and tested on ADSP. Error bars indicate 95% confidence intervals (CI) across folds.*

In contrast, our IDP-mediated models, which estimate imaging-derived phenotypes from genetic data before classifying AD, demonstrated improved performance (Figure 3, right). Models based on individual

IDPs achieved AUCs of 0.641 (DTI), 0.631 (FreeSurfer), and 0.654 (ASHS). The corresponding AUPRC results are provided in Supplementary Figure 4.

Finally, the integrated G2DBridge framework, which combines genetically predicted IDPs through a neural network, achieved the highest accuracy with an AUC of 0.657. This outperformed both the PRS baselines and individual IDP-mediated models.

Together, these results demonstrate that G2DBridge generalizes effectively to independent cohorts and highlights the value of leveraging genetically derived imaging representations for disease risk prediction.

Risk Stratification Analysis

AUC summarizes ranking ability but does not show how prediction scores distribute between cases and controls. Risk stratification analysis addresses this by revealing group separation and areas of overlap or misclassification. We therefore examined the distributions of normalized prediction scores for each model (Figure 4). All PRS models showed statistically significant differences between cases and controls, with the strongest among them being PGS002753, PGS004062, and PGS003992. However, none matched the discriminative strength of G2DBridge, which achieved the most significant separation (Welch's t-test, $p < 1.23 \times 10^{-300}$). These results indicate that G2DBridge captures the relationship between genetic risk and AD status more strongly than standard PRS approaches.

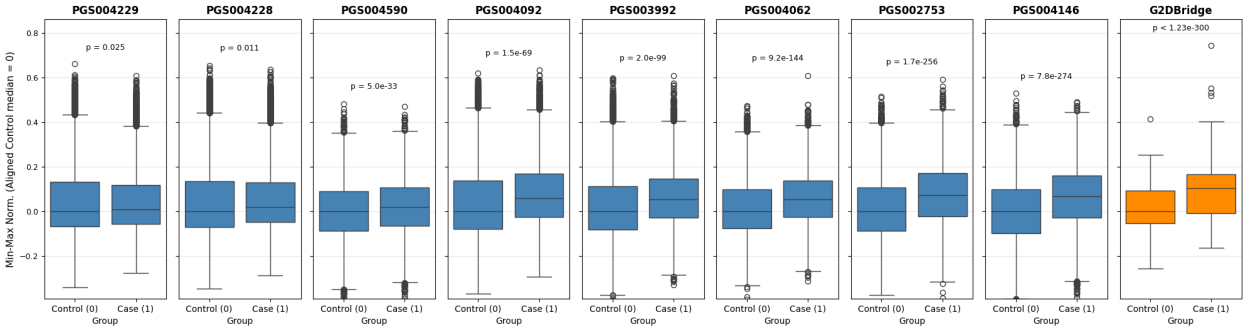


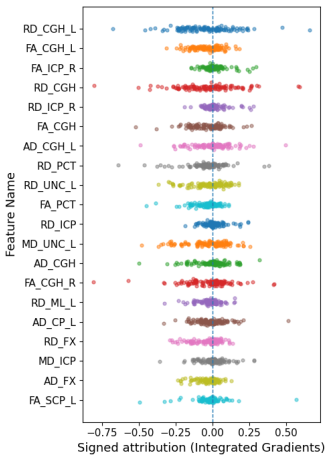
Figure 4 Risk stratification using PRS models and G2DBridge. *Box plots showing the distribution of min–max normalized prediction scores across case and control groups. Scores are aligned such that the median of the control group is set to zero. PGS identifiers correspond to models from the PGS Catalog. Welch's t-test was used to assess group differences; p-values are shown above each panel.*

Feature Attribution and Consistency Analysis

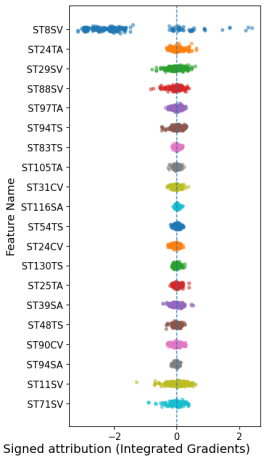
To better understand which imaging features contributed most to AD classification, we applied gradient-based attribution analysis and visualized feature importance using two complementary methods.

First, we generated beeswarm plots of the top 20 features per fold, with results from fold 1 shown in Figure 5 (a–c) (DTI, FreeSurfer, ASHS; additional folds in Supplementary Figure 5-7). These plots display subject-level signed Integrated Gradient (IG) values, where positive attributions push the model toward predicting AD and negative values toward control.

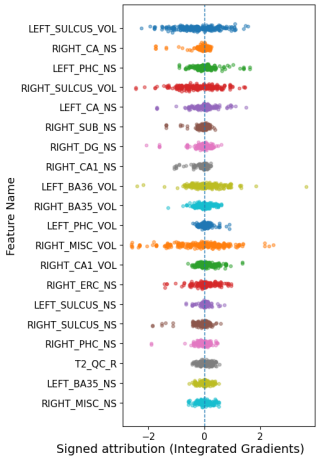
Across modalities, distinct patterns emerged. For DTI, diffusivity metrics from the cingulum–hippocampal tract (RD_*CGH, MD_CGH*) and uncinate fasciculus (*MD_UNC_L*) were consistently among the strongest predictors, aligning with prior studies of white matter disruption in AD. In FreeSurfer, cortical and subcortical features from temporal regions (e.g., *ST119TS, ST11SV, ST8SV, ST29SV*) repeatedly ranked


RD_CGH_L
FA_CGH_L
FA_ICP_R
RD_CGH
RD_ICP_R
FA_CGH
AD_CGH_L
RD_PCT
RD_UNC_L
FA_PCT
RD_ICP
MD_UNC_L
AD_CGH
FA_CGH_R
RD_ML_L
AD_CP_L
RD_FX
MD_ICP
AD_FX
FA_SCP_L
Feature Name
−0.75 −0.50 −0.25 0.00 0.25 0.50
Signed attribution (Integrated Gradients)

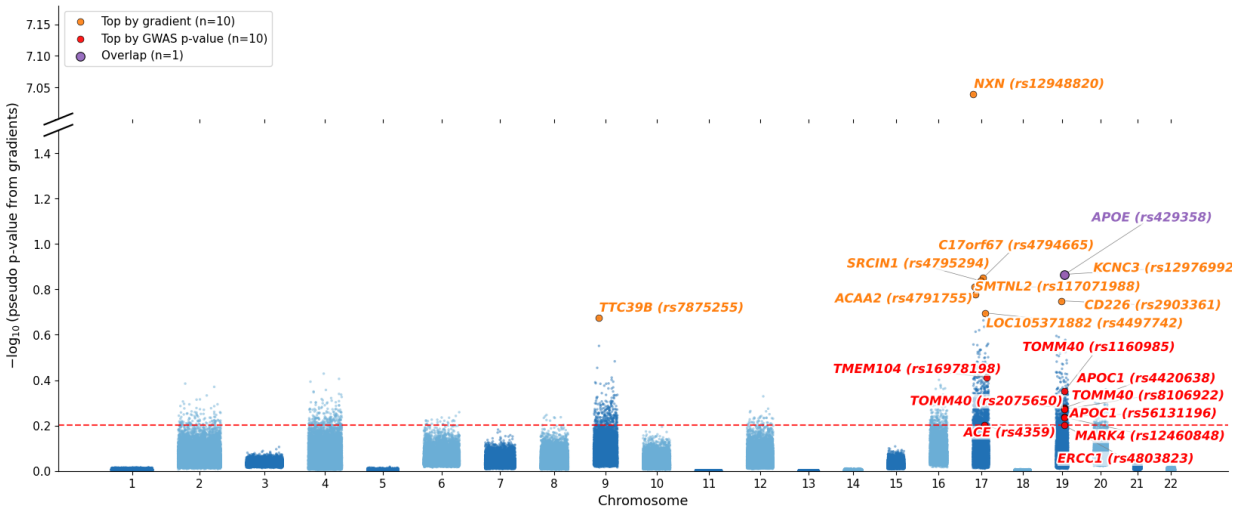
(a)


ST8SV
ST24TA
ST29SV
ST88SV
ST97TA
ST94TS
ST83TS
ST105TA
ST31CV
ST116SA
ST54TS
ST24CV
ST130TS
ST25TA
ST39SA
ST48TS
ST90CV
ST94SA
ST11SV
ST71SV
Feature Name
−2 0 2
Signed attribution (Integrated Gradients)

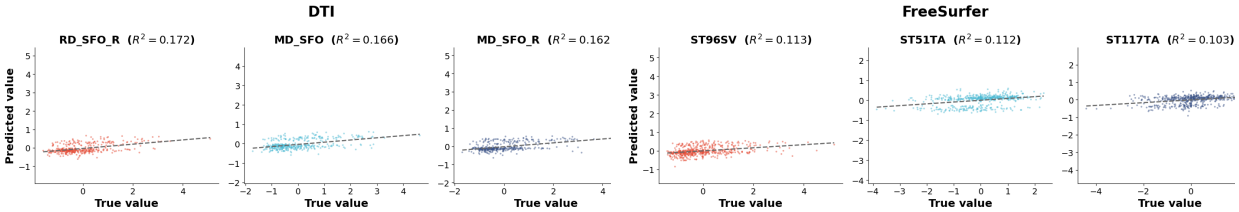
(b)


LEFT_SULCUS_VOL
RIGHT_CA_NS
LEFT_PHC_NS
RIGHT_SULCUS_VOL
LEFT_CA_NS
RIGHT_SUB_NS
RIGHT_DG_NS
RIGHT_CA1_NS
LEFT_BA36_VOL
RIGHT_BA35_VOL
LEFT_PHC_VOL
RIGHT_MISC_VOL
RIGHT_CA1_VOL
RIGHT_ERC_NS
LEFT_SULCUS_NS
RIGHT_SULCUS_NS
RIGHT_PHC_NS
T2_QC_R
LEFT_BA35_NS
RIGHT_MISC_NS
Feature Name
−2 0 2
Signed attribution (Integrated Gradients)

(c)


Top by gradient (n=10)
Top by GWAS p-value (n=10)
Overlap (n=1)
NXN (rs12948820)
APOE (rs429358)
C17orf67 (rs4794665)
SRCIN1 (rs4795294)
KCNC3 (rs12976992)
SMTNL2 (rs117071988)
ACAA2 (rs4791755)
CD226 (rs2903361)
TTC39B (rs7875255)
LOC105371882 (rs4497742)
TOMM40 (rs1160985)
TMEM104 (rs16978198)
APOC1 (rs4420638)
TOMM40 (rs2075650)
TOMM40 (rs8106922)
APOC1 (rs56131196)
ACE (rs4359)
MARK4 (rs12460848)
ERCC1 (rs4803823)
−log10 (pseudo p-value from gradients)
Chromosome

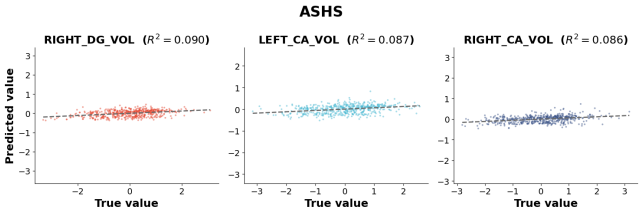
(d)


DTI
RD_SFO_R ($R^2 = 0.172$)
MD_SFO ($R^2 = 0.166$)
MD_SFO_R ($R^2 = 0.162$)
Predicted value
True value

(e)

FreeSurfer
ST96SV ($R^2 = 0.113$)
ST51TA ($R^2 = 0.112$)
ST117TA ($R^2 = 0.103$)
Predicted value
True value

(f)


ASHS
RIGHT_DG_VOL ($R^2 = 0.090$)
LEFT_CA_VOL ($R^2 = 0.087$)
RIGHT_CA_VOL ($R^2 = 0.086$)
Predicted value
True value

(g)

Figure 5. Feature attribution, SNP importance, and modality-specific predictive fidelity.
(a–c) *Beeswarm plots of subject-level signed Integrated Gradients (IG) for the top 20 features from fold 1, shown separately by IDP modality (DTI, FreeSurfer, ASHS). Points are jittered for visibility. Positive IG values push the model toward Alzheimer's disease (AD), while negative values push away.*
(d) *Manhattan plot of gradient-based SNP importance with GWAS and gene annotations. Each point is a SNP positioned by chromosome and genomic coordinate. The y-axis shows gradient-derived importance transformed as −log10 of a pseudo p-value computed from min-max normalized gradients, so larger values indicate higher model importance. Labeled SNPs are the union of two sets: orange labels mark the top gradient-ranked variants above the percentile threshold, and red labels mark variants prioritized by the lowest GWAS p-values among those same high-gradient candidates (GWAS filtering applied using the external p-value threshold). Variants meeting both criteria are highlighted as the overlap category. The red dashed horizontal line indicates the 99th-percentile cutoff (top 1%) of transformed gradient-importance scores. Gene symbols for labeled SNPs were retrieved from dbSNP and displayed next to rsIDs. The plot highlights the canonical AD locus on chr19 (APOE/TOMM40/APOC1 region) alongside additional labeled genes for follow-up.*
(e–g) *True versus predicted values for the best-performing feature in each modality using the best-performing model:* ***(e)*** *DTI,* ***(f)*** *FreeSurfer,* ***(g)*** *ASHS. Each point represents one subject, and the dashed line shows the linear fit between true and predicted values.*

highly, reflecting known cortical thinning and atrophy in AD. ASHS features (e.g., *LEFT_BA36_VOL, RIGHT_BA36_VOL*) highlighted medial temporal lobe structures, including parahippocampal gyrus, entorhinal cortex, Cornu Ammonis, dentate gyrus, and subiculum, regions long established as early markers of AD-related neurodegeneration.

Together, these analyses confirm that the model relies on biologically meaningful imaging markers, with DTI capturing microstructural tract damage, FreeSurfer reflecting cortical atrophy, and ASHS identifying medial temporal lobe degeneration.

Figure 5-d summarizes SNP-level signals using a Manhattan-style layout. Each point represents a SNP positioned by chromosome and genomic coordinate, and the y-axis reports a transformed model-derived importance score computed as $-\log_{10}(p_{pseudo})$, where p_{pseudo} is derived from min-max normalized gradient magnitudes. This transformation makes high-gradient SNPs visually comparable to conventional GWAS Manhattan plots. GWAS summary statistics were downloaded from the NHGRI-EBI GWAS Catalog (GCST013197), corresponding to Wightman, D. P. et al.[41], and used only as an external reference for interpreting SNP-level signals; GWAS p-values are not plotted in this panel.

To improve readability in the presence of one extreme value, the y-axis is shown with a broken scale: most variants cluster in the lower range, whereas a single high-scoring locus is displayed in the upper panel. Labeled variants correspond to (i) the top gradient-ranked SNPs above the chosen percentile threshold and (ii) the top GWAS-supported SNPs among those same high-gradient candidates, using an external GWAS p-value filter. The union of these labeled SNPs was annotated using dbSNP to obtain mapped gene symbols, and these genes were printed to facilitate downstream manuscript edits and biological interpretation.

The labeled set includes expected Alzheimer's disease loci in the chr19 *APOE* region (*APOE, TOMM40, APOC1, MARK4*), demonstrating that the gradient-based ranking recovers known AD-associated genomic signals. In addition, the plot highlights other genes in the labeled set (e.g., *ACE, TTC39B, KCNC3, SRCIN1, SMTNL2, TMEM104, ERCC1, NXN*), which may reflect cohort-specific effects, linkage with nearby causal variation, or model sensitivity to correlated genomic features. Overall, Fig. 5d provides a compact view of how model-derived SNP importance aligns with established AD genetics while also surfacing additional candidates for follow-up.

Figures 5e–g visualize NN predictions for each modality (DTI, FreeSurfer, ASHS). Predictions were concatenated across cross-validation folds, feature-wise R^2 was computed, and the top three target features were selected for display. The true–predicted scatter plots show a clear positive, approximately linear association between ground truth and NN estimates, indicating good agreement; the dispersion around the fitted regression line reflects residual error and potential differences in calibration across targets.

Discussion

In this study, we introduced G2DBridge, a predictive framework that links genetic variation to disease risk through IDPs as intermediate traits. By combining the broad availability of genetic data with the biological specificity of neuroimaging, G2DBridge enables scalable prediction while retaining interpretability. Although AD served as the case study because of its strong genetic and imaging biomarkers, the approach is generalizable to other neurodegenerative, cardiovascular, and psychiatric diseases where paired genetic and imaging data exist.

Our results highlight three key findings. First, IDPs are informative predictors of AD status, and integrating multiple IDPs improves classification compared with single-modality models (Figure 2-a). The multi-IDP model achieved the highest overall accuracy, showing that combining modalities provides a more complete representation of disease pathology[37–40]. Second, IDPs can be predicted from genetic data (Figure 2-b and Supplementary Figure 2), with NN outperforming other machine learning methods across modalities in both explained variance and error. Third, and most importantly, genetically predicted IDPs improve disease risk prediction compared with PRS. In ADNI internal validation, genetically inferred IDPs yielded consistently higher AUCs than genetic-only models (Figure 2-c-d-e-f). In ADSP external validation, G2DBridge outperformed multiple state-of-the-art PRS models (Figure 3), achieving superior accuracy while using the same target genetic data. This represents a free gain in predictive power. By learning intermediate imaging representations, G2DBridge extracts additional signal without requiring new data types or increased genotyping cost. From a clinical perspective, the framework also provides added value in risk stratification. Distributional analyses demonstrated significantly clearer separation between AD cases and controls for G2DBridge than for PRS models (Figure 4).

Interpretability analyses further showed that G2DBridge relies on biologically meaningful markers (Figure 5). Across modalities, distinct patterns emerged. For DTI, diffusivity metrics from the cingulum–hippocampal tract (*RD_CGH, MD_CGH*) and uncinate fasciculus (*MD_UNC_L*) were consistently among the strongest predictors, aligning with prior studies of white matter disruption in AD[42–44]. In FreeSurfer, cortical and subcortical features from temporal regions (e.g., *ST119TS, ST11SV, ST8SV, ST29SV*) repeatedly ranked highly, reflecting known cortical thinning and atrophy in AD[45–47]. ASHS features (e.g. *LEFT_BA36_VOL, RIGHT_BA36_VOL*) highlighted medial temporal lobe structures, including parahippocampal gyrus, entorhinal cortex, Cornu Ammonis, dentate gyrus, and subiculum regions, long established as early markers of AD-related neurodegeneration[48,49].

This work has several limitations. First, the analyses pooled participants across ancestries for model training and evaluation due to limited sample sizes in several individual ancestry groups. As larger datasets become available, ancestry-specific models will likely deliver improved and more generalizable performance[49,50]. Second, we used IDPs from structural and diffusion MRI rather than raw images to reduce dimensionality and improve robustness; however, models operating directly on raw MRI may capture finer signal when adequate data, harmonization, and computation are available[50]. Future work could also include PET or functional imaging to provide more biological insight[51]. Finally, while explainable AI helped improve interpretability, it does not establish causality. Future studies could integrate causal inference methods to better understand biological mechanisms[49].

In summary, G2DBridge generalizes across cohorts, improves prediction relative to PRS without additional data requirements, and identifies biologically meaningful imaging features. By bridging genetics, imaging, and disease outcomes, the framework offers a scalable and clinically relevant strategy for advancing genetic risk prediction and moving toward personalized medicine.

Methods

Data

We used two large and well-characterized cohorts: the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Alzheimer's Disease Sequencing Project (ADSP). All models were trained and tuned on ADNI with five-fold cross-validation, and ADSP was reserved exclusively for external validation.

ADNI. ADNI is a multicenter longitudinal study of older adults spanning the spectrum of AD. It provides genetic, neuroimaging, and clinical data. For this study, we constructed binary outcome models using genetic variants and IDPs. Three IDP sets were analyzed: diffusion tensor imaging (DTI), FreeSurfer morphometric measures, and Automatic Segmentation of Hippocampal Subfields (ASHS). These modalities capture complementary aspects of brain structure and were used to evaluate the contribution of single vs. multi-IDP integration. After quality control and intersection, 203,909 SNPs remained(Supplementary Table 2). IDPs were standardized using a single z-score scaler fitted on the with-genome subset for each IDP set, and the same transformation was applied consistently when generating model inputs. The analytic samples were DTI with D=265 features and N=385 participants, FreeSurfer with D=323 features and N=534 participants, and ASHS with D=35 features and N=534 participants. The intersection common to all three IDP sets contained N=355 participants. The genotype-only cohort contained N=651 individuals for the ADNI dataset.

ADSP. ADSP is a large consortium integrating genetic and clinical data across diverse AD studies. We used ADSP version 5 solely as an independent test set. To ensure comparability, identical quality-control procedures were applied to ADNI and ADSP. Single-nucleotide variants with minor allele frequency below 0.05 or with high missingness were removed, and analyses were restricted to variants present in both datasets. The genotype-only cohort contained N=25113 for the ADSP dataset. Individuals present in both the ADSP and ADNI datasets were removed from the ADSP dataset before performance evaluation.

Machine learning models. We implemented logistic regression, Lasso regression, random forest, XGBoost and feedforward neural networks to predict both continuous (IDPs) and binary (AD status) outcomes.

<u>Logistic regression</u>. Standard logistic regression (generalized linear model, no penalty) was used as a linear baseline for binary classification.

<u>Lasso regression</u>. Lasso regression (L1 penalty) was applied to enable variable selection in both regression and classification settings, particularly suitable for high-dimensional SNP data. The penalty parameter (α) was optimized using glmnet.

<u>Random forest.</u> Random forest classifiers and regressors were used to capture nonlinear relationships and assess feature importance. Hyperparameters including n_estimators, max_depth, max_features, and min_samples_split were tuned using Optuna.

<u>XGBoost</u>. Extreme Gradient Boosting (XGBoost) was employed as a powerful ensemble method that combines multiple decision trees using gradient boosting to capture complex nonlinear relationships in the data. Its regularization capabilities help prevent overfitting, making it particularly effective for high-

dimensional genomic data. Hyperparameters such as n_estimators, learning_rate, max_depth, and subsample were optimized using Optuna.

Neural networks. Feedforward neural networks were implemented in PyTorch. Architectures varied in depth, width, activation functions, and dropout rates, all optimized with Optuna. Training used the Adam optimizer with early stopping based on validation loss.

Hyperparameter optimization was performed with five-fold cross-validation, using AUC for classification and R^2/MSE for regression.

Interpretability. To improve interpretability, we used gradient-based attribution methods from Captum (PyTorch library), focusing on Integrated Gradients. These analyses estimate how much each input feature contributes to the model's output, allowing us to quantify feature importance at both the subject and cohort levels.

G2DBridge framework

G2DBridge is a three-stage neural network framework that links genotype data to AD risk via genetically predicted IDPs.

Stage 1: Genetics to IDPs. Three parallel multilayer perceptron map the standardized SNP vector $x \in R^p$ to a predicted IDP $\widehat{y^{(r)}}$, where $r \in \{\text{DTI,FS,ASHS}\}$. Each network has its own parameters $\theta^{(r)} = \{W^{(l,r)} b^{(l,r)}\}$, L_r hidden layers and nonlinear activations $\theta^{(l,r)}$.

Stage 2: IDPs to AD. For each modality, a classifier predicts AD status from the standardized IDP vectors. The final hidden layer of each classifier provides an embedding $f^{(r)}$, representing a compact modality-specific feature vector.

Stage 3: The three IDP embeddings are concatenated into a fused representation

$$z = \left[f^{(\text{DTI})}, f^{(\text{FS})}, f^{(\text{ASHS})}\right] \in R^d, \quad d = d_{\text{DTI}} + d_{\text{FS}} + d_{\text{ASHS}}$$

and passed through a multilayer classifier with K hidden layers. The final output is the predicted probability of AD,

$$\hat{p} = \sigma\left(w^{(o)T} h^{(K)} + b^{(o)}\right)$$

The model was implemented in PyTorch with the Adam optimizer. Hyperparameters (layer widths, depth, learning rate, dropout, activation functions) were tuned with Optuna. We used five-fold cross-validation with early stopping on validation loss. Performance was reported as mean AUC (classification) or mean R^2 (regression) across folds.

Application to genetic-only data. Once trained on cohorts with paired genetic and imaging data, G2DBridge can be applied to genetic-only datasets. In this setting, the first stage of the framework predicts IDPs directly from the SNP vector, IDP $\widehat{y^{(r)}}$. The resulting genetically inferred IDPs are passed through

the second and third stages of the network (IDP→AD classifiers and multi-IDP integration) to yield AD risk probabilities.

Polygenic risk scores. All polygenic scores used in this work were obtained from the PGS Catalog and computed using previously published SNP effect sizes from large-scale GWAS studies. Before scoring, SNP alleles were strand-matched and harmonized with reference panels to ensure consistent allele orientation across datasets. Genome-wide PRS were then calculated across all variants passing quality control using the *plink2 --score* function.

We considered multiple Alzheimer's disease PRS derived from large European cohorts, as well as one score trained on multi-ancestry samples. These polygenic scores were generated using various methodologies. A detailed summary of all polygenic scores, including their PGS identifiers, training sample sizes, number of variants, and methodologies, is provided in Table 1.

Table 1 Comparison with PGS Catalog PRS

PGS Catalog ID	Training Sample Size	Number of Variant	PRS methodology
PGS002249[52]	55,134 European	249,273	LD clumping
PGS002753[53]	455,258 European	1,092,011	PRS-CS
PGS003992[54]	472,868 European	1,136,212	DBSLMM-auto
PGS004062[54]	472,868 European	691,136	Megaprs.auto
PGS004092[54]	472,868 European	1,109,233	PRS-CS-auto
PGS004146[54]	472,868 European	915,771	SBayesR-auto
PGS004228[55]	63,926 European	8863	PRSice-2
PGS004229[55]	63,926 European	8858	PRSice-2
PGS004590[56]	2,240 Hispanic or Latin American 8,006 African American 8,036 East Asian 487,511 European	363	Pruning and Thresholding (P+T)

Risk stratification analysis. For risk stratification, we compared the distributions of model prediction scores between AD cases and controls. Prediction scores were min–max normalized, aligning the median of controls to zero for comparability. Group differences were assessed using Welch's t-test, and significance values were reported for each model.

Data and Code availability

We provide the scripts used to perform the model training and inference proposed in the article in the GitHub repository https://github.com/Cedars-CIG/G2DBridge.

Acknowledgments

O.B.O, R.C, and R.L are supported by the Department of Computational Biomedicine, Cedars Sinai.

Author Contributions

O.B.O. and R.L. contributed to method development, model training, and writing the paper, while R.C. and O.W. contributed to data processing.

Declaration of Interests

The authors declare no competing interests.

References

1. Zarnani, K. *et al.* Discovering markers of healthy aging: a prospective study in a Danish male birth cohort. *Aging (Albany NY)* **11**, 5943 (2019).
2. Patel, K. *et al.* Unsupervised deep representation learning enables phenotype discovery for genetic association studies of brain imaging. *Commun. Biol.* **7**, 414 (2024).
3. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
4. Lewis, C. M. & Vassos, E. Prospects for using risk scores in polygenic medicine. *Genome Med.* **9**, 96 (2017).
5. Schwarzerova, J. *et al.* A perspective on genetic and polygenic risk scores—advances and limitations and overview of associated tools. *Brief. Bioinform.* **25**, bbae240 (2024).
6. Patel, K. *et al.* New phenotype discovery method by unsupervised deep representation learning empowers genetic association studies of brain imaging. *medRxiv* 2012–2022 (2022).
7. Schulz, M.-A. *et al.* Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nat. Commun.* **11**, 4238 (2020).
8. Inouye, M. *et al.* Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).
9. Chatterjee, N., Shi, J. & Garc\'\ia-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
10. Elgart, M. *et al.* Non-linear machine learning models incorporating SNPs and PRS improve polygenic prediction in diverse human populations. *Commun. Biol.* **5**, 856 (2022).
11. Sigurdsson, A. I. *et al.* Deep integrative models for large-scale human genomics. *Nucleic Acids Res.* **51**, e67–e67 (2023).
12. Kelemen, M. *et al.* Performance of deep-learning based approaches to improve polygenic scores. *medRxiv* 2010–2024 (2024).
13. Clark, K., Leung, Y. Y., Lee, W.-P., Voight, B. & Wang, L.-S. Polygenic risk scores in Alzheimer's disease genetics: methodology, applications, inclusion, and diversity. *Journal of Alzheimer's Disease* **89**, 1–12 (2022).
14. Khera, A. V *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
15. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
16. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
17. Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
18. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
19. Cooper, L. A. D. *et al.* Novel genotype-phenotype associations in human cancers enabled by advanced molecular platforms and computational analysis of whole slide images. *Laboratory investigation* **95**, 366–376 (2015).
20. Suk, H.-I. & Shen, D. Deep learning-based feature representation for AD/MCI classification. in *International conference on medical image computing and computer-assisted intervention* 583–590 (Springer, 2013).

21. Basaia, S. *et al.* Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin.* **21**, 101645 (2019).
22. Zhang, D. *et al.* Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856–867 (2011).
23. Wen, J. *et al.* Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med. Image Anal.* **63**, 101694 (2020).
24. Lee, E. E. *et al.* Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **6**, 856–864 (2021).
25. Smith, S. M. *et al.* An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* **24**, 737–745 (2021).
26. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424 (2018).
27. Yushkevich, P. A. *et al.* Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum. Brain Mapp.* **36**, 258–287 (2015).
28. Li, R., Chen, Y., Ritchie, M. D. & Moore, J. H. Electronic health records and polygenic risk scores for predicting disease risk. *Nat. Rev. Genet.* **21**, 493–502 (2020).
29. Basser, P. J., Mattiello, J. & LeBihan, D. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* **66**, 259–267 (1994).
30. Mori, S. & Zhang, J. Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron* **51**, 527–539 (2006).
31. Shen, L. *et al.* Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* **53**, 1051–1063 (2010).
32. Yang, X. *et al.* Multi-organ imaging-derived polygenic indexes for brain and body health. *medRxiv* 2023–2024 (2024).
33. Ching, T. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **15**, 20170387 (2018).
34. Jack Jr, C. R. *et al.* NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & dementia* **14**, 535–562 (2018).
35. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
36. Frisoni, G. B., Fox, N. C., Jack Jr, C. R., Scheltens, P. & Thompson, P. M. The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* **6**, 67–77 (2010).
37. Lei, B. *et al.* Alzheimer's disease diagnosis from multi-modal data via feature inductive learning and dual multilevel graph neural network. *Med. Image Anal.* **97**, 103213 (2024).
38. Adarsh, V., Gangadharan, G. R., Fiore, U. & Zanetti, P. Multimodal classification of Alzheimer's disease and mild cognitive impairment using custom MKSCDDL kernel over CNN with transparent decision-making for explainable diagnosis. *Sci. Rep.* **14**, 1774 (2024).
39. Archetti, D. *et al.* A machine learning model to harmonize volumetric brain MRI data for quantitative neuroradiologic assessment of Alzheimer disease. *Radiol. Artif. Intell.* **7**, e240030 (2024).
40. Lee, G., Nho, K., Kang, B., Sohn, K.-A. & Kim, D. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci. Rep.* **9**, 1952 (2019).
41. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).
42. Mayo, C. D. *et al.* Relationship between DTI metrics and cognitive function in Alzheimer's disease. *Front. Aging Neurosci.* **10**, 436 (2019).
43. Gozdas, E. *et al.* Focal white matter disruptions along the cingulum tract explain cognitive decline in amnestic mild cognitive impairment (aMCI). *Sci. Rep.* **10**, 10213 (2020).

44. Kantarci, K. *et al.* White matter integrity determined with diffusion tensor imaging in older adults without dementia: influence of amyloid load and neurodegeneration. *JAMA Neurol.* **71**, 1547–1554 (2014).
45. Yang, H. *et al.* Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. *Gen. Psychiatr.* **32**, e100005 (2019).
46. Li, K., Xie, D., Zhang, Z., Fu, C. & Li, C. Voxel-based and surface-based morphometry in the cortical thickness, and cortical and subcortical grey matter volume in patients with mild to moderate Alzheimer's disease. *Front. Aging Neurosci.* **17**, 1546977 (2025).
47. Sattari, N., Faeghi, F., Shekarchi, B. & Heidari, M. H. Assessing the changes of cortical thickness in alzheimer disease with mri using freesurfer software. *Basic Clin. Neurosci.* **13**, 185 (2022).
48. da Silva Filho, S. R. B. *et al.* Neuro-degeneration profile of Alzheimer's patients: a brain morphometry study. *Neuroimage Clin.* **15**, 15–24 (2017).
49. Hasan Saif, F., Al-Andoli, M. N. & Bejuri, W. M. Y. W. Explainable AI for Alzheimer detection: A review of current methods and applications. *Applied Sciences* **14**, 10121 (2024).
50. Taiyeb Khosroshahi, M. *et al.* Explainable artificial intelligence in neuroimaging of Alzheimer's disease. *Diagnostics* **15**, 612 (2025).
51. Christodoulou, R. C., Woodward, A., Pitsillos, R., Ibrahim, R. & Georgiou, M. F. Artificial Intelligence in Alzheimer's Disease Diagnosis and Prognosis Using PET-MRI: A Narrative Review of High-Impact Literature Post-Tauvid Approval. *J. Clin. Med.* **14**, 5913 (2025).
52. Lourida, I. *et al.* Association of lifestyle and genetic risk with incidence of dementia. *JAMA* **322**, 430–437 (2019).
53. Mars, N. *et al.* Systematic comparison of family history and polygenic risk across 24 common diseases. *The American Journal of Human Genetics* **109**, 2152–2162 (2022).
54. Monti, R. *et al.* Evaluation of polygenic scoring methods in five biobanks shows larger variation between biobanks than methods and finds benefits of ensemble learning. *The American Journal of Human Genetics* **111**, 1431–1447 (2024).
55. Green, R. E. *et al.* Investigating associations between blood metabolites, later life brain imaging measures, and genetic risk for Alzheimer's disease. *Alzheimers Res. Ther.* **15**, 38 (2023).
56. Lake, J. *et al.* Multi-ancestry meta-analysis and fine-mapping in Alzheimer's disease. *Mol. Psychiatry* **28**, 3121–3132 (2023).