Decoding Phenotypes: A Framework for Fusing Genomic Language Models and Neuroimaging

Tianli Tao[1,2], Ziyang Wang[3], Emma Robinson[1], Rachel Sparks[1], and Le Zhang[3]

[1] School of Biomedical Engineering and Imaging Science, Faculty of Life Science and Medicine, King's College London, UK
[2] School of Computer Science and Digital Technologies, Aston University, UK
[3] School of Engineering, College of Engineering and Physical Sciences, University of Birmingham, UK

Abstract. Neuroimaging and genetic testing are two clinical references for nervous system diseases, offering complementary diagnostic information. However, integrating genomic and neuroimaging data for disease diagnosis is challenging due to cross-modality heterogeneity. Existing imaging–genetics approaches mainly encode genetic information as hard-coded labels, which lose the local sequence context around disease-associated variants. To address this limitation, we propose GeneFuse, a multimodal learning framework that aligns genetic representations from pretrained Genomic Language Models (GLMs) with neuroimaging features. GeneFuse integrates two components: (1) Genotype-Conditioned Feature Modulation (GCFM) to use genomic embeddings to modulate image feature maps; and (2) Uncertainty-aware Genomic Residual Fusion (U-GRF) for combining imaging and genotypic features with soft gating. We evaluate GeneFuse on early cognitive decline identification (NC vs. MCI) and dementia screening (NC vs. AD). In the APOE-centered setting, GeneFuse achieves AUROCs of 0.77 and 0.83, outperforming existing imaging–genetics fusion methods. These results indicate that GLM-derived genomic embeddings provide additional information to imaging.

Keywords: Alzheimer's Disease · Genomic Language Models · Multimodal Fusion · Imaging-Genetics

1 Introduction

Integration of neuroimaging and genetic data is critical for precise diagnosis and understanding of nervous system diseases [15]. Neuroimaging captures macroscopic structural phenotypes, such as brain structure and cortical atrophy during aging. Meanwhile, genomics provides complementary information on the underlying disease processes [12]. Despite this complementarity, effective data fusion remains challenging due to the substantial heterogeneity in the data representation. Many imaging–genetics fusion approaches represent genetic information as tabular features [9]. However, this approach may lose the information on the complex interactions between genomic factors and imaging phenotypes.

Recently, genomic language models (GLMs) have emerged as a promising approach for DNA sequence representation. Foundation models such as the Nucleotide Transformer (NT-v2) [3], DNABERT [8] and HyenaDNA [14] treat genomic sequences as linguistic structures and leverage self-supervised learning on large-scale genomic datasets to learn context-sensitive representations around disease-associated variants.

Despite the promise of GLM-derived representations, their effective integration with high-dimensional volumetric medical imaging features remains unexplored. Existing approaches often depend on handcrafted radiomics or naive late-stage fusion strategies, which may not fully exploit cross-modal interactions between genomics and imaging biomarkers [11]. Bridging this gap involves two fundamental challenges. *(1) Semantic misalignment*: Unlike image-text pairs (e.g., CLIP [17]), gene-image pairs lack explicit semantic alignment. *(2) Coupling uncertainty*: genotype–phenotype relationships are often weak or indirect. Genetic features may not always manifest as observable imaging information due to genetic resilience or environmental factors. Naive fusion strategies carry an inherent risk: when imaging features alone are highly informative, incorporating potentially noisy genomic signals can degrade predictive performance.

In this work, we propose a framework that aligns GLM-derived genomic representations with volumetric neuroimaging features. **Our contributions**: **(1)** We adapt feature-wise linear modulation (FiLM) [16] to align imaging-genomic features through Genotype-Conditioned Feature Modulation (GCFM), which uses genomic embeddings to generate channel-wise affine parameters to modulate volumetric imaging features in an appropriate manner. **(2)** We present Uncertainty-aware Genomic Residual Fusion (U-GRF), a lightweight module that supplies the imaging branch's single-pass predictive entropy to a learned soft gate controlling the contribution of genotypic information. The gate is data-driven and is not structurally constrained to vary monotonically with entropy. **(3)** We integrate these components into a network and validate it on Alzheimer's disease (AD) detection tasks using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset in single- and multi-locus settings [7].

2 Methods

2.1 Unimodal Representation and Encoders

Our framework is designed to process two distinct modalities: volumetric imaging and genomic sequences. In this section, we detail the data construction and encoding networks for each modality.

Volumetric Image Encoding. To extract high-level anatomical representations from MRI of the brain, we use a 3D TransUNet backbone [2]. In our implementation, the three encoder stages use channel widths $C_{1/2} = 32$, $C_{1/4} = 64$, and $C_{1/8} = 128$. Each stage contains two $3 \times 3 \times 3$ convolutional layers with instance normalization and GELU activation, followed by stride-2 downsampling. A two-layer transformer bottleneck with hidden dimension 256,

four attention heads, and an MLP expansion ratio of 4 is applied at the deepest scale. The decoder mirrors the three scales using trilinear upsampling, skip-feature concatenation, and a $3 \times 3 \times 3$ convolutional refinement block followed by GCFM (Section 2.2) at each decoder resolution. After each GCFM block, globally pooled features from the three decoder resolutions are concatenated and projected to an imaging latent vector $z_{img} \in \mathbb{R}^{d_{img}}$, where $d_{img} = 256$. Given an MRI scan $X_{mri} \in \mathbb{R}^{H \times W \times D}$, the backbone produces decoder feature maps $F_s \in \mathbb{R}^{C_s \times H_s \times W_s \times D_s}$ at resolutions $s \in \{1/2, 1/4, 1/8\}$, which are genotype-conditioned before final aggregation and diagnosis.

Semantic Genomic Encoding. Unlike traditional association studies that encode Single Nucleotide Polymorphisms (SNPs) as discrete tabular values (i.e., $\{0, 1, 2\}$), we encode local genomic sequence using a pretrained GLM [3]. This approach preserves nucleotide context around disease-associated variants.

Genomic Tokenization & Context Modeling: We focus on the APOE locus region as it is a well-established genetic risk factor for late-onset Alzheimer's disease (AD) [10]. Variant coordinates follow GRCh37/hg19 and all extracted sequences are represented in the forward-reference (+) orientation. The APOE setting includes the two defining variants rs429358 (chr19:45,411,941) and rs7412 (chr19:45,412,079). For each target variant and subject, we extract a patient-specific DNA sequence window of length $L = 1024$ bp. Following Beagle phasing, the two haplotypes are encoded separately by the GLM and then likewise averaged to form the APOE representation. The multi-locus setting uses $K = 5$ loci (APOE, PVRL2, ABCA7, CD33, and APOC1), processes their representations independently, and aggregates them by permutation-invariant mean pooling.

Genomic Backbone: To encode each patient-specific sequence window into a dense representation, we employ a pretrained GLM. In this work, we instantiate the backbone using the Nucleotide Transformer v2 (NT-v2-100M) [3]. We denote the resulting subject-level genomic latent vector as $z_{gene} \in \mathbb{R}^{d_g}$. During training, the parameters of NT-v2 are frozen to preserve the pretrained genomic representations and reduce computational overhead.

2.2 Cross-Modal Feature Fusion

To effectively bridge the gap between imaging and genomic biomarkers, we introduce two fusion modules: Genotype-Conditioned Feature Modulation (GCFM) and Uncertainty-aware Genomic Residual Fusion (U-GRF). These modules work together to ensure genetic information guides visual feature extraction without introducing errors when imaging biomarkers are sufficient for the task.

Genotype-Conditioned Feature Modulation (GCFM). GCFM is an imaging-genetics adaptation of FiLM [16]. Different from a late FiLM fusion baseline that conditions only the final imaging representation, GCFM injects GLM-derived conditioning into multi-scale spatial feature maps of the 3D volumetric image encoding. This allows genomic information to recalibrate imaging features before the final representation.

In our implementation, GCFM is applied at the 1/2, 1/4, and 1/8 decoder resolutions, after skip-feature concatenation and convolutional refinement at each

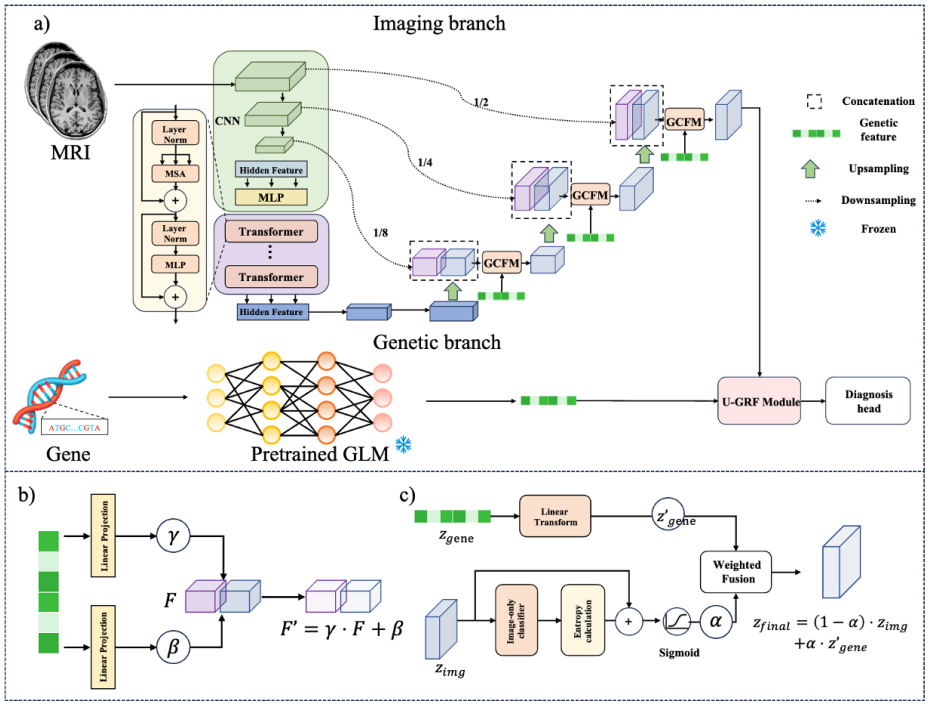

Fig. 1. Overview of the proposed GeneFuse framework. (a) The imaging branch extracts multi-scale MRI features with an encoder–decoder backbone, while a frozen pretrained GLM encodes patient-specific genomic sequences. GCFM is placed on the decoder path at the 1/8, 1/4, and 1/2 resolutions, after upsampling, skip-feature concatenation, and convolutional refinement. (b) GCFM predicts channel-wise affine parameters (γ, β) from genomic embeddings and modulates imaging features by $F' = \gamma \cdot F + \beta$. (c) U-GRF supplies single-pass imaging predictive entropy to a learned soft gate α that combines z_{img} with the transformed genomic vector z'_{gene}.

resolution. Formally, at a decoder stage with channel dimension C_s, given the genomic embedding $z_{gene} \in \mathbb{R}^{d_g}$ (where $d_g = 512$ for NT-v2-100M) and the spatial feature map $F^s_{\text{img}} \in \mathbb{R}^{C_s \times H_s \times W_s \times D_s}$, GCFM predicts channel-wise scale (γ_s) and shift (β_s) parameters by two separate Multi-Layer Perceptrons (MLPs). Each MLP consists of two linear layers with ReLU activation: $\text{Linear}(d_g, 256) \rightarrow \text{ReLU} \rightarrow \text{Linear}(256, C_s)$. Thus: $\gamma_s = \mathcal{M}^s_\gamma(z_{gene})$, $\beta_s = \mathcal{M}^s_\beta(z_{gene})$, where $\mathcal{M}^s_\gamma$ and $\mathcal{M}^s_\beta$ project the genomic latent space to $\mathbb{R}^{C_s}$. The parameters are applied to the imaging feature map via $F'^s_{\text{img}} = \gamma_s \odot F^s_{\text{img}} + \beta_s$, where $\odot$ denotes channel-wise multiplication with broadcasting over H_s, W_s, D_s. Because the affine parameters are channel-wise and spatially shared, GCFM does not directly predict a voxel-level attention map. Spatially heterogeneous effect arises indirectly from modulating feature channels with different spatially structured

activations. The modulated multi-scale imaging features are then aggregated to produce the imaging latent vector z_{img} for downstream fusion and classification.

Uncertainty-aware Genomic Residual Fusion (U-GRF). U-GRF is an entropy-conditioned fusion strategy. It treats single-pass predictive entropy as an additional conditioning signal for a learned soft gate, rather than as a calibrated estimate of model uncertainty.

The fusion process operates in three steps. First, initial logits l_{img} from the imaging branch define $p = \text{softmax}(l_{img})$ and the predictive entropy $E_{img} = -\sum p \log p$. This quantity summarizes the confidence of a single forward pass but does not capture calibrated epistemic uncertainty. Next, the gate is computed as $\alpha = \sigma(\text{MLP}_{gate}([z_{img} \oplus E_{img}]))$, where σ ensures $\alpha \in [0, 1]$. Because entropy is only one learned input, the architecture neither guarantees that the gate uses it nor enforces a monotonic relationship between E_{img} and genomic weight. Finally, $z'_{gene} = \text{Transform}(z_{gene})$ and $z_{final} = (1-\alpha) \cdot z_{img} + \alpha \cdot z'_{gene}$. The Transform$(\cdot)$ function is a single linear projection layer, Linear(d_g, d_{img}), which maps z_{gene} from 512 to 256 dimensions.

3 Experiments

Dataset and Preprocessing. Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [7]. We selected subjects with paired T1-weighted (T1w) magnetic resonance imaging (MRI) scans and whole-genome sequencing (WGS). The final dataset contained $N = 182$ subjects spanning three diagnostic groups based on baseline clinical assessments: 52 Normal Controls (NC), 53 subjects with Mild Cognitive Impairment (MCI), and 77 subjects with AD.

In this work, two clinically motivated diagnosis tasks are considered: NC vs. MCI, corresponding to early identification of cognitive decline, and NC vs. AD , corresponding to dementia screening. All MRI scans underwent a standard pre-processing pipeline, including AC-PC correction, skull stripping, bias field correction and affine registration to the MNI152 atlas space. Finally, the images were cropped and resized to a resolution of $128 \times 128 \times 128$ and normalized. The genomic pre-processing pipeline followed the pipeline described in [11], including variant filtering using `VCFtools` [4], genotype phasing using `Beagle` (v5.5) [1] and personal genome construction using `vcf2diploid` [18].

Implementation Details. Experiments were conducted on the ADNI dataset with stratified 5-fold cross-validation. To prevent data leakage, data splitting was performed at the subject level, so scans and genomic windows from the same subject never appear in both training and test folds. Model training was performed using PyTorch for 50 epochs with AdamW ($lr = 10^{-4}, wd = 10^{-2}$). The total training objective combines the classification loss with an L1 regularization term on the fusion gate α: $\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda_{reg} \cdot \frac{1}{N} \sum |\alpha|$. Accuracy and F1-Score are computed from the positive-class probability using a fixed threshold of 0.5. The architecture and optimization hyperparameters reported above are fixed across folds and methods. Statistical significance for AUROC comparisons is

Table 1. Performance comparison on NC vs. MCI and NC vs. AD tasks. We report mean ± std. Best results are in bold text. $^{\dagger}p_{\text{Holm}} < 0.05$ vs. Image-only and $^{\ddagger}p_{\text{Holm}} < 0.05$ vs. FiLM Fusion using paired DeLong tests on subject-level out-of-fold AUROC predictions.

	NC vs. MCI			**NC vs. AD**		
Method	AUROC	Acc.	F1	AUROC	Acc.	F1
Baselines						
GeneFuse (Image-only)	0.69 ± 0.09	0.56 ± 0.07	0.48 ± 0.10	0.71 ± 0.05	0.61 ± 0.07	0.56 ± 0.13
GeneFuse (GLM-only)	0.63 ± 0.08	0.50 ± 0.06	0.45 ± 0.09	0.71 ± 0.03	0.58 ± 0.08	0.54 ± 0.03
Image + SNP scalar	0.70 ± 0.09	0.57 ± 0.07	0.49 ± 0.11	0.73 ± 0.09	0.58 ± 0.05	0.53 ± 0.06
Existing Imaging–Genetics Methods						
Stage-wise DNN [22]	0.70 ± 0.02	0.58 ± 0.05	0.52 ± 0.08	0.74 ± 0.07	0.64 ± 0.05	0.49 ± 0.06
MMDL [20]	0.71 ± 0.08	0.60 ± 0.07	0.53 ± 0.07	0.76 ± 0.10	0.62 ± 0.08	0.58 ± 0.12
MADDi [6]	0.72 ± 0.07	0.61 ± 0.03	0.55 ± 0.04	0.77 ± 0.12	0.67 ± 0.07	0.60 ± 0.09
Fusion Strategies						
Late Concat	0.67 ± 0.08	0.52 ± 0.03	0.46 ± 0.09	0.76 ± 0.08	0.60 ± 0.09	0.51 ± 0.14
FiLM Fusion [16]	0.72 ± 0.03	0.59 ± 0.07	0.52 ± 0.05	0.73 ± 0.03	0.61 ± 0.05	0.54 ± 0.16
Cross-Attention	0.74 ± 0.07	0.61 ± 0.04	0.54 ± 0.06	0.77 ± 0.07	0.63 ± 0.08	0.48 ± 0.08
Residual (α=0.05)	0.73 ± 0.08	0.60 ± 0.07	0.53 ± 0.11	0.76 ± 0.02	0.62 ± 0.04	0.56 ± 0.03
Ablation Study (Ours w/ NT-V2)						
GCFM Only	0.75 ± 0.06	0.63 ± 0.06	0.54 ± 0.09	$0.81 \pm 0.06^{\ddagger}$	0.65 ± 0.07	0.52 ± 0.05
U-GRF Only	0.74 ± 0.07	0.62 ± 0.07	0.58 ± 0.10	0.78 ± 0.04	0.68 ± 0.07	0.52 ± 0.07
GeneFuse (Full)	$\mathbf{0.77 \pm 0.05}$	$\mathbf{0.66 \pm 0.04}$	$\mathbf{0.60 \pm 0.06}$	$\mathbf{0.83 \pm 0.05}^{\dagger\ddagger}$	$\mathbf{0.71 \pm 0.03}$	$\mathbf{0.66 \pm 0.04}$
GLM Comparison (GeneFuse)						
w/ DNAGPT [21]	0.75 ± 0.06	0.63 ± 0.06	0.57 ± 0.09	0.79 ± 0.03	0.64 ± 0.07	0.62 ± 0.10
w/ GROVER [19]	0.74 ± 0.05	0.62 ± 0.05	0.56 ± 0.10	0.80 ± 0.07	0.65 ± 0.04	0.55 ± 0.11
w/ GENA-LM [5]	0.73 ± 0.08	0.60 ± 0.02	0.54 ± 0.11	0.77 ± 0.06	0.63 ± 0.08	0.44 ± 0.16

assessed on subject-level out-of-fold predictions using paired DeLong tests for correlated ROC curves, avoiding tests based only on the five fold-level summary values. The family of four primary comparisons (GeneFuse vs. Image-only and vs. FiLM Fusion for each of the two tasks) is corrected using the Holm procedure. We compare GeneFuse with unimodal baselines (Image-only, GLM-only), image fusion with scalar SNP encodings, representative imaging–genetics methods (MADDi [6], MMDL [20], stage-wise DNN [22]), common fusion strategies (Late Concat, FiLM Fusion [16], Cross Attention, Residual), and alternative GLM encoders (DNAGPT [21], GROVER [19], GENA-LM [5]). For the settings of fusion baselines, Late Concat concatenates the final imaging and genomic latent vectors, and Residual adds a fixed-weight transformed genomic residual to the imaging representation. All fusion baselines use the same imaging backbone under the same data split.

Experimental Results. Table 1 reports the APOE-centered comparison. GeneFuse achieves AUROCs of 0.77 (NC vs. MCI) and 0.83 (NC vs. AD), improving over the image-only baseline by Δ0.08 and Δ0.12 AUROC, respectively. Under paired DeLong tests on subject-level out-of-fold predictions, only the NC vs. AD improvement over Image-only remains significant after Holm correction ($p_{\text{Holm}} < 0.05$). Image + SNP provides only a small gain, suggesting that scalar SNP encoding is less effective than GLM-derived sequence repre-

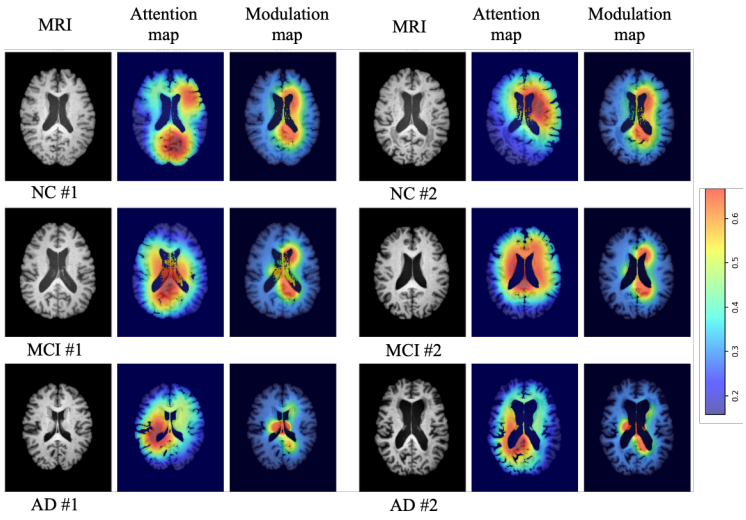

Fig. 2. GCFM modulation-energy visualization, computed as $|F' - F|$ and averaged across channels.

Table 2. Ablation of GCFM insertion resolutions on the decoder path using NT-v2.

Decoder Resolution(s)	NC vs. MCI			NC vs. AD		
	AUROC	Acc.	F1	AUROC	Acc.	F1
1/2 only	0.72 ± 0.07	0.56 ± 0.09	0.49 ± 0.10	0.78 ± 0.08	0.64 ± 0.05	0.59 ± 0.02
1/4 only	0.71 ± 0.07	0.59 ± 0.07	0.53 ± 0.04	0.73 ± 0.07	0.60 ± 0.02	0.39 ± 0.15
1/8 only	0.74 ± 0.03	0.57 ± 0.08	0.55 ± 0.08	0.76 ± 0.08	0.65 ± 0.08	0.50 ± 0.05
1/4 + 1/8	0.75 ± 0.06	0.64 ± 0.06	0.57 ± 0.03	0.79 ± 0.06	0.68 ± 0.07	0.58 ± 0.06
1/2 + 1/4 + 1/8	$\mathbf{0.77 \pm 0.05}$	$\mathbf{0.66 \pm 0.04}$	$\mathbf{0.60 \pm 0.06}$	$\mathbf{0.83 \pm 0.05}$	$\mathbf{0.71 \pm 0.03}$	$\mathbf{0.66 \pm 0.04}$

sentations. Compared with existing imaging–genetics baselines, late concatenation, FiLM-style conditioning, cross-attention, and a fixed residual fusion rule, GeneFuse obtains the most favorable overall balance in all three metrics. The ablation study suggests complementary roles for the two modules: GCFM conditions multi-scale imaging features on genomic context, while U-GRF provides an entropy-conditioned, learned residual weighting mechanism.

Qualitative Visualization. Fig. 2 visualizes the spatial energy of the feature change induced by GCFM. We extract intermediate imaging features before and after modulation and compute $|F' - F|$, averaged across channels. Because GCFM applies channel-wise affine parameters shared over spatial locations, the resulting pattern reflects the spatial structure already present in the amplified or suppressed feature channels. As illustrated in Fig. 2 genomic conditioning qualitatively produces spatially structured modulation responses around periventricular white matter and medial temporal regions.

Table 3. Performance comparison across different AD-related gene loci.

Gene Locus	NC vs. MCI			NC vs. AD		
	AUROC	Acc.	F1	AUROC	Acc.	F1
Multi-locus	$\mathbf{0.79 \pm 0.04}$	$\mathbf{0.68 \pm 0.06}$	$\mathbf{0.62 \pm 0.08}$	$\mathbf{0.84 \pm 0.02}$	$\mathbf{0.72 \pm 0.04}$	0.63 ± 0.05
APOE	0.77 ± 0.05	0.66 ± 0.04	0.60 ± 0.06	0.83 ± 0.05	0.71 ± 0.03	$\mathbf{0.66 \pm 0.04}$
PVRL2	0.72 ± 0.07	0.60 ± 0.08	0.53 ± 0.11	0.77 ± 0.08	0.64 ± 0.09	0.48 ± 0.13
ABCA7	0.73 ± 0.05	0.63 ± 0.05	0.44 ± 0.13	0.79 ± 0.08	0.65 ± 0.08	0.60 ± 0.12
CD33	0.72 ± 0.04	0.61 ± 0.02	0.53 ± 0.12	0.78 ± 0.03	0.64 ± 0.09	0.49 ± 0.13
APOC1	0.68 ± 0.09	0.59 ± 0.09	0.42 ± 0.10	0.71 ± 0.09	0.62 ± 0.10	0.56 ± 0.14

GCFM Insertion Scale. An ablation study was performed on the insertion resolution of GCFM along the CNN-Transformer U-Net decoder (Table 2). Each block follows upsampling, skip-feature concatenation, and convolutional refinement. Single-resolution modulation is competitive in most settings, while using all three decoder resolutions gives the highest observed performance. This suggests that genomic conditioning can act on features spanning fine anatomical detail and deeper semantic information.

Impact of Genomic Encoders. To assess the importance of GLM selection, we compared NT-v2 to alternative genomic encoders (Table 1, bottom section). NT-v2 achieves the highest mean performance across both tasks, with modest AUROC gains over other GLMs.

Analysis of Different Gene Loci Performance. The diagnostic contributions of different AD risk loci are shown in Table 3. The single-locus APOE result corresponds to GeneFuse in Table 1. The multi-locus setting uses $K = 5$ loci and provides the highest observed AUROC and accuracy. This suggests that the additional AD-related loci may contribute complementary information for MRI-genomic fusion.

4 Conclusion

GeneFuse shows that GLM-derived risk-locus embeddings can condition MRI features more effectively than scalar SNP encoding or late fusion in this cohort. The multi-locus results further suggest that AD-related loci beyond APOE may contribute additional diagnostic information. This study is limited by a single cohort and small task-specific samples, which restrict statistical power and generalizability. Future work will evaluate matched-information genomic baselines, external cohorts and calibrated uncertainty estimates, as well as GLM fine-tuning on larger imaging–genomics datasets.

Acknowledgments. This work was supported by the King's College London–China Scholarship (K-CSC) PhD Scholarship Programme. The authors acknowledge the use of resources provided by the Isambard-AI National AI Research Resource (AIRR) [13]. Isambard-AI is operated by the University of Bristol and is funded by the UK Government's Department for Science, Innovation and Technology (DSIT) via UK Research

and Innovation; and the Science and Technology Facilities Council [ST/AIRR/I-A-I/1023].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Browning, B.L., Tian, X., Zhou, Y., Browning, S.R.: Fast two-stage phasing of large-scale sequence data. The American Journal of Human Genetics **108**(10), 1880–1890 (2021)
2. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
3. Dalla-Torre, H., Gonzalez, L., Mendoza-Revilla, J., Lopez Carranza, N., Grzywaczewski, A.H., Oteri, F., Dallago, C., Trop, E., de Almeida, B.P., Sirelkhatim, H., et al.: Nucleotide transformer: building and evaluating robust foundation models for human genomics. Nature Methods **22**(2), 287–297 (2025)
4. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.: The variant call format and vcftools. Bioinformatics **27**(15), 2156–2158 (2011)
5. Fishman, V., Kuratov, Y., Shmelev, A., Petrov, M., Penzar, D., Shepelin, D., Chekanov, N., Kardymon, O., Burtsev, M.: Gena-lm: a family of open-source foundational dna language models for long sequences. Nucleic Acids Research **53**(2), gkae1310 (2025)
6. Golovanevsky, M., Eickhoff, C., Singh, R.: Multimodal attention-based deep learning for alzheimer's disease diagnosis. Journal of the American Medical Informatics Association **29**(12), 2014–2022 (2022)
7. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)
8. Ji, Y., Zhou, Z., Liu, H., Davuluri, R.V.: Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome. Bioinformatics **37**(15), 2112–2120 (2021)
9. Li, Y., Niu, D., Qi, K., Liang, D., Long, X.: An imaging and genetic-based deep learning network for alzheimer's disease diagnosis. Frontiers in Aging Neuroscience **17**, 1532470 (2025)
10. Liu, C.C., Kanekiyo, T., Xu, H., Bu, G.: Apolipoprotein e and alzheimer disease: risk, mechanisms and therapy. Nature Reviews Neurology **9**(2), 106–118 (2013)
11. Liu, Q., Zeng, W., Zhu, H., Li, L., Wong, W.H., Initiative, A.D.N.: Leveraging genomic large language models to enhance causal genotype-brain-clinical pathways in alzheimer's disease. medRxiv pp. 2024–10 (2024)
12. Mashhour, M.A., Youssef, I., Wahed, M.A., Mabrouk, M.S.: The intersection of genetics and neuroimaging: A systematic review of imaging genetics in neurological disease for personalized treatment. Journal of Molecular Neuroscience **75**(2), 66 (2025)

13. McIntosh-Smith, S., Alam, S.R., Woods, C.: Isambard-AI: A leadership class supercomputer optimised specifically for artificial intelligence. arXiv preprint arXiv:2410.11199 (2024). https://doi.org/10.48550/arXiv.2410.11199
14. Nguyen, E., Poli, M., Faizi, M., Thomas, A., Birch-Sykes, C., Wornow, M., Patel, A., Rabideau, C., Massaroli, S., Bengio, Y., et al.: Hyenadna: long-range genomic sequence modeling at single nucleotide resolution. arXiv preprint arXiv:2306.15794 (2023)
15. Parvin, N., Joo, S.W., Jung, J.H., Mandal, T.K.: Multimodal ai in biomedicine: Pioneering the future of biomaterials, diagnostics, and personalized healthcare. Nanomaterials **15**(12), 895 (2025)
16. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
17. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
18. Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., Leng, J., Bjornson, R., Kong, Y., Kitabayashi, N., et al.: Alleleseq: analysis of allele-specific expression and binding in a network framework. Molecular systems biology **7**(1), 522 (2011)
19. Sanabria, M., Hirsch, J., Joubert, P.M., Poetsch, A.R.: Dna language model grover learns sequence context in the human genome. Nature Machine Intelligence **6**(8), 911–923 (2024)
20. Venugopalan, J., Tong, L., Hassanzadeh, H.R., Wang, M.D.: Multimodal deep learning models for early detection of alzheimer's disease stage. Scientific Reports **11**(1), 3254 (2021)
21. Zhang, D., Zhang, W., Zhao, Y., Zhang, J., He, B., Qin, C., Yao, J.: Dnagpt: A generalized pre-trained tool for versatile dna sequence analysis tasks. arXiv preprint arXiv:2307.05628 (2023)
22. Zhou, T., Thung, K.H., Zhu, X., Shen, D.: Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Human Brain Mapping **40**(3), 1001–1016 (2019)