



MULTIMODAL DIFFUSION MODEL FOR SYNTHETIC ALZHEIMER'S DISEASE DATA GENERATION USING MRI, PET, AND CLINICAL BIOMARKERS

Jyoti Kumari^{*1} and Shyam Sundar Gupta²

^{1,2}Amity School of Engineering and Technology, Amity University, Gwalior, M.P India.

¹<http://orcid.org/0009-0004-0533-4306>, ²<https://orcid.org/0000-0001-6841-3897>

Email: *jyotikumari2211221@gmail.com, gohadshyam@gmail.com

ARTICLE INFO

Article History

Received: February 17, 2026

Reviewed: March 17, 2026

Accepted: April 01, 2026

Published: May 29, 2026

Keywords:

Alzheimer's disease,
Multimodal synthesis,
Magnetic resonance imaging,
Positron Emission Tomography,
Alzheimer's Disease
Neuroimaging Initiative (ADNI)
Dataset.

ABSTRACT

The scarcity, imbalance, and privacy constraints of high-quality multimodal datasets hinder progress in machine-learning for Alzheimer's disease (AD). Structural MRI, PET imaging and clinical biomarkers each capture complementary aspects of AD pathophysiology, but multimodal collections that include all three modalities are limited. To address data scarcity and enable privacy-preserving research, we propose a Multimodal Diffusion Generative Model (MDGM) that jointly synthesizes high-resolution structural MRI, PET, and tabular clinical biomarkers. MDGM uses modality-specific encoders, a latent-space fusion mechanism, and a conditional score-based diffusion model operating in a compact latent space; the decoder yields harmonized MRI, PET volumes and clinical vectors consistent with a conditioning disease profile. We train and evaluate MDGM on the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, quantify image fidelity and a domain-adapted FID, assess clinical realism by correlation of biomarkers and cognitive scores, and test downstream utility by training multimodal diagnostic classifiers with and without synthetic augmentation. Our results show that MDGM generates clinically plausible multimodal samples that preserve disease patterns (e.g., hippocampal atrophy, elevated amyloid signal) and substantially improve classifier sensitivity for under-represented early-disease classes when used for augmentation. MDGM demonstrates how diffusion-based generative models can produce multimodal synthetic cohorts to accelerate AD research while respecting patient privacy.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder representing the most common form of dementia [1]. Clinically, AD is characterized by progressive episodic memory loss, impairment in executive functions, language deficits, and behavioral changes. At the biological level, hallmark features include extracellular amyloid- β ($A\beta$) plaques [2] and intracellular neurofibrillary tangles of hyperphosphorylated tau protein, as well as widespread synaptic dysfunction and neuronal loss. Pathophysiological changes often begin many years before clinical symptoms emerge; thus, early detection and prognostication are crucial for intervention, clinical trials, and patient management. Biomarkers from multiple domains are used to stage and support diagnosis. Structural magnetic resonance imaging (MRI) provides maps of brain anatomy and atrophy patterns (e.g., hippocampal and medial temporal lobe volume loss) and allows longitudinal tracking of neurodegeneration. Positron emission tomography (PET) [3] using amyloid or metabolic tracers measures molecular pathology (amyloid burden, tau pathology, or hypometabolism). Cerebrospinal fluid (CSF) biomarkers ($A\beta$ 42, total tau, p-tau) and blood-based assays extend biochemical detection.

Cognitive assessments (MMSE, ADAS-Cog, CDR) quantify clinical performance. The NIA-AA AT(N) framework conceptually groups biomarkers into amyloid (A), tau (T), and neurodegeneration (N) – and integrative multimodal assessment improves diagnostic accuracy and staging beyond any single modality. Machine learning models trained for diagnosis, staging, or prognosis benefit strongly from multimodal inputs. Multimodal fusion enables models [4] to leverage structural atrophy (MRI) together with molecular signatures (PET) and clinical context (scores, genotype), often improving sensitivity for early or atypical presentations.

However, practical constraints limit progress:

- Acquisition cost and availability. PET imaging is less commonly acquired due to expense and radiation exposure. As a result, fully paired MRI+PET+clinical datasets are considerably smaller than single-modality sets.
- Class imbalance. Early disease stages (subjective cognitive decline, early MCI) and rarer phenotypes have few samples, impairing learning.
- Site/scanner heterogeneity. Multi-center studies introduce variability in acquisition protocols and scanner characteristics, complicating model generalization.
- Privacy & governance. Patient data sharing is tightly regulated (HIPAA, GDPR), hindering large-scale data pooling across institutions. Even de-identified imaging carries potential re-identification risk when combined with other metadata.

Because of these challenges, synthetic data generation is an attractive strategy for augmenting datasets, balancing classes, simulating rare phenotypes, and enabling privacy-preserving model development. Generative models have evolved rapidly. Generative adversarial networks (GANs) first enabled realistic image synthesis but suffer from training instability, mode collapse, and difficulty in reliable likelihood estimation. Diffusion probabilistic models (DDPMs) and their latent variants have emerged as a favorable alternative: they provide stable optimization, strong mode coverage, and high fidelity outputs, especially when combined with latent compression (e.g., VAE encoders) to reduce computational costs. In medical imaging, diffusion models have generated high-fidelity single-modality MRI and CT scans[5], and sequence translation tasks (e.g., MRI→PET) have been attempted. Yet, AD research requires jointly consistent generation across modalities: a synthetic MRI should exhibit atrophy patterns that align with the synthetic PET amyloid distribution and with synthetic clinical scores. Joint multimodal synthesis is technically more challenging because the generator must learn cross-modal relationships and disease-driven correlations rather than independent marginal distributions for each modality.

Key pitfalls to address include:

- Cross-modal consistency: Independent synthesis can produce plausible single modalities that are inconsistent when paired.
- Clinical plausibility: Synthetic clinical vectors (scores, biomarker concentrations) must align statistically and biologically with synthesized images.
- Memorization & privacy: Models can inadvertently memorize and reproduce training subjects; synthetic data must be vetted to avoid re-identification.
- Quantitative evaluation: Standard perceptual metrics (e.g., FID) were developed for natural images and must be adapted for 3D medical data, while clinical validity requires biomarker correlations, region-wise measurements, and downstream tasks.

This work proposes MDGM (Multimodal Diffusion Generative Model) — a framework to synthesize triads consisting of structural MRI volumes, co-registered PET volumes (amyloid or FDG), and structured clinical biomarker vectors (MMSE, CDR, CSF measures, APOE status)[6-10].

MDGM aims to:

- Model the joint distribution $p(\text{MRI}, \text{PET}, \text{CLIN})$ conditional on disease severity and genotype.
- Generate synthetic cohorts that preserve imaging features (atrophy, PET uptake patterns) and clinical correlations (MMSE decline with atrophy) without reproducing identifiable patient data.
- Demonstrate downstream utility: synthetic data used for augmentation should improve performance of multimodal diagnostic/prognostic models, especially for under-represented classes.
- Provide evaluation protocols for fidelity, clinical consistency, and privacy (memorization tests).

The central novelty is joint conditional latent diffusion with a multimodal latent fusion, enabling compact modeling with strong cross-modal coherence while keeping the training tractable for 3D neuroimaging.

II. RELATED WORK

Generative models for medical imaging include GANs, VAEs, and diffusion models. GAN-based MRI and PET synthesis demonstrated feasibility for single modalities but suffered stability issues. Diffusion models (DDPMs and latent diffusion) have shown strong performance for high-fidelity medical images and image translation (MRI→PET) in recent works. Multimodal fusion and diagnostic models combining MRI, PET, CSF, and cognitive scores have improved classification; however, prior synthetic data efforts rarely addressed joint multimodal generation plus clinical feature alignment. This paper builds on diffusion theory (Ho et al.), latent diffusion techniques (Rombach et al.), and multimodal fusion principles, and applies them specifically to produce clinically coherent MRI+PET+clinical triplets for AD research.

II.1 GENERATIVE MODELS IN MEDICAL IMAGING

GANs have been widely explored for synthetic medical imaging; however, their training instability and mode collapse limit reliability in neuroimaging. Diffusion models, introduced by Ho et al. (2020), progressively denoise data via stochastic processes, enabling high-quality generation. Works such as MedDiffusion (2023) and BrainDM (2024) demonstrated the use of diffusion for MRI generation, but lacked multimodal integration.

II.2 MULTIMODAL MODELING IN ALZHEIMER’S DISEASE

AD research increasingly relies on multimodal biomarkers including MRI-derived cortical thinning, PET-derived amyloid burden, CSF biomarkers (Aβ42/40, pTau), and cognitive scores (MMSE, CDR). Several multimodal deep learning works (Chen et al., 2025; Li et al., 2024) combined these features for classification, but did not address data synthesis.

II.3 DIFFUSION MODELS FOR MULTIMODAL DATA

Cross-modal diffusion models, such as Conditional Latent Diffusion (Rombach et al., 2022), allow guiding generation using auxiliary signals. However, their application to medical multimodal fusion—especially MRI, PET, and structured clinical data—remains unexplored.

Table No 1: Literature Review shows the different methods.

Author	Year	Dataset	Modalities	Method	Strength	Limitation
Yu et al.[11]	2024	ADNI	sMRI + clinical + genetic	Transformer (unified)	Cross-modal attention / high diagnostic AUC	Not focused on image generation
Tang et al.[12]	2024	ADNI	MRI + PET	Improved Transformer	Excellent classification (98.1% acc) / multimodal	Limited to diagnosis, no synthesis
Chen et al.[13]	2025	ADNI	MRI → PET	Diffusion model (multi-view)	Realistic FDG-PET synthesis from MRI T1WI / improved diagnosis	One-way/two-way synthesis strategies needed
Liu et al.[14]	2024	ADNI	MRI → multi-modal PET	Residual Diffusion	Unified multi-biomarker PET synthesis / high fidelity	Computationally intensive diffusion
Dhinagar et al.[15]	2024	ADNI	AV45 PET	3D GAN	Amyloid dynamics / progression modeling in 6ficient space	Mode 6ficient risks in GANs / amyloid-specific
(Bossa et al. [16])	2023	ADNI	sMRI + PET	Multi-modal GNN	Biomarker modeling / handles incomplete data	No direct image generation
Ho et al.[17]	2020	General (extensions to medical)	Images	DDPM	Stable / foundational generation	Not multimodal or medical-specific
Rombach et al.[18]	2022	LAION (extensions)	Images	LDM	High-resolution / 6ficiente generation	Not medical / requires fine-tuning for MRI/PET

Source: Authors, (2026).

III. MATERIALS AND METHODS

In this section we will discuss the methods which are as follows.

III.1 DATA AND PREPROCESSING (ADNI DATASET):

ADNI-1/GO/2 imaging collections (T1-weighted MRI, amyloid/Tau/FDG PET), clinical assessments (MMSE, ADAS-Cog, CDR), CSF measures (Aβ42, pTau), and genetics (APOE). Use ADNI documentation[19-21] for scanner protocols and preprocessing recommendations. PMC+1

Inclusion criteria:

- Subjects with both T1 MRI and at least one PET scan acquired within 6 months.
- Clinical visit with MMSE/CDR and APOE genotyping available.

Preprocessing steps (recommended reproducible pipeline):

1. MRI:

- Bias field correction (N4).
- Skull-stripping (HD-BET or FreeSurfer).
- Spatial normalization to MNI152 (affine + nonlinear).
- Resample to isotropic 1 mm or 1.5 mm (choose consistent voxel grid).
- Intensity normalization: clip 0.5–99.5 percentile, z-score per subject or white-matter normalization.
- Optional: cortical parcellation for region-wise evaluation (FreeSurfer). PMC.

2. PET:

- Motion correction and spatial realignment to MRI.
- Convert to SUVR using reference region (cerebellum/pontine) per tracer.
- Resample to MRI grid, smooth to common resolution (ADNI retrospective reprocessing uses harmonized 6mm³ voxel size). ADNI+1.

3. Tabular/clinical:

- Impute missing values with modality-appropriate strategies (median, MICE).
- Normalize numeric variables (z-score). One-hot encode categorical (APOE ε4 as 0/1/2).
- Align time: choose closest visits within ±6 months for multimodal pairing.

Dataset split:

- Train/val/test split at subject level (no leakage): e.g., 70/15/15 stratified by diagnosis (CN/EMCI/LMCI/AD) and age bracket.

III.2 OVERALL ARCHITECTURE

MDGM has these components:

1. Encoders

- E_{MRI} : 3D encoder (3D UNet encoder path or 3D ResNet backbone) transforming MRI volume $X_{MRI} \rightarrow z_{MRI} \in \mathbb{R}^d$.
- E_{PET} : 3D encoder for PET $X_{PET} \rightarrow z_{PET}$.
- E_{CLIN} : MLP encoder for clinical vector $x_{clin} \rightarrow z_{clin}$.

2. Latent fusion + cross-attention

- Combine $z_{MRI}, z_{PET}, z_{clin}$ into fused latent z_{fused} using weighted sum + learned attention blocks:

$$z_{fused} = \text{CrossAttn}(\text{concat}(z_{MRI}, z_{PET}), z_{clin})$$

1

Attention allows clinical features to modulate imaging latent channels (disease severity conditioning).

3. Conditional diffusion model

- A score network $s_\theta(x_t, t, c)$ (conditional) where input x_t is noisy latent (or noisy images in pixel domain) and c is conditioning vector from z_{fused} . The backbone is a 3D UNet with time embedding and cross-attention modules inserting cat multiple scales.

4. Decoders

- D_{MRI}, D_{PET} map denoised latent (or output images) back to full resolution MRI/PET.
- D_{CLIN} reconstructs tabular features (optional—helps enforce clinical consistency).

5. Residual modality branches

- Modality-specific residual decoders ensure modality unique textures (e.g., PET tracer uptake) are preserved—motivated by IJCAI prior residual diffusion. ijcai.org

6. Auxiliary discriminators / critics (optional)

- For higher fidelity, add patch-level discriminators on image outputs (hybrid diffusion+adversarial fine-tuning). Use with caution (training stability).

Design choice (latent vs pixel diffusion):

- Latent diffusion: encode images to lower-dim latent (VAE-style) then run diffusion in latent space → less compute, faster sampling, but requires good VAE. (Rombach et al.). [arXiv](https://arxiv.org/abs/2206.00364)
- Pixel-space diffusion: run diffusion directly on voxel intensities (more compute, higher fidelity). For 3D volumes, computational cost is high but feasible with patching or multi-scale sampling. For MDGM I recommend latent diffusion with a high-quality 3D VAE (or vector-quantized VAE) to balance compute and fidelity.

III.3 FORMAL DIFFUSION FORMULATION

We follow the denoising diffusion probabilistic model (DDPM) framework (Ho et al.). Let x_0 denote clean sample (here x_0 can be a concatenated object of MRI/PET/tables or a fused latent). We define diffusion forward process with variance schedule $\{\beta_t\}_{t=1}^T$:

Forward (noising):

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (2)$$

So closed form:

$$q(x_t | x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t) I), \bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s) \quad (3)$$

Learned reverse (score) network:

We train network $\epsilon_\theta(x_t, t, c)$ to predict noise ϵ conditional on c (conditioning vector). Loss (simplified):

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{x_0, \epsilon, t} [\| \epsilon - \epsilon_\theta(x_t, t, c) \|^2] \quad (4)$$

Alternatively, parameterize the reverse mean μ_θ , but noise-prediction works best. Conditional modification:

c is the fused latent z_{fused} (clinical + cross-modal features). Implement cross-attention in the UNet to inject cat multiple decoder blocks.

Joint multimodal objective:

If we model joint output $X = (X_{MRI}, X_{PET}, X_{CLIN})$, we can either:

- Run a single diffusion on concatenated channels (voxel channels for MRI & PET + dense channels for clinical after broadcasting), or
- Diffuse a fused latent z_{fused} and decode to modalities — I prefer the latent fused diffusion approach for scaling.

III.4 LOSSES (COMPOSITE)

Total loss:

$$\mathcal{L} = \lambda_{ddpm} \mathcal{L}_{ddpm} + \lambda_{rec} \mathcal{L}_{rec} + \lambda_{ssim} \mathcal{L}_{ssim} + \lambda_{clin} \mathcal{L}_{clin} + \lambda_{kl} \mathcal{L}_{kl} \quad (5)$$

Where:

- $\mathcal{L}_{ddpm} = \mathbb{E} \| \epsilon - \epsilon_\theta(x_t, t, c) \|^2$ (diffusion noise-prediction). arXiv
- $\mathcal{L}_{rec} = \| x_0 - \hat{x}_0 \|_1$ (L1 reconstruction for images).
- $\mathcal{L}_{ssim} = 1 - \text{SSIM}(x_0, \hat{x}_0)$ (structural similarity penalty).
- $\mathcal{L}_{clin} = \| y_{clin} - \hat{y}_{clin} \|_2^2$ (reconstruction error for clinical features).
- $\mathcal{L}_{kl}$ if using VAE latent pretrain.
- λ hyperparameters tuned by validation.

Optional adversarial loss (for perceptual realism):

$$\mathcal{L}_{adv} = \mathbb{E}[\log D(x)] + \mathbb{E}[\log (1 - D(\hat{x}))] \quad (6)$$

Combine as hybrid: diffusion pretrain $\rightarrow$ optional adversarial fine-tune.

III.5 CONDITIONING & CLINICAL CONSISTENCY

Clinical conditioning vector c :

- Compose c from normalized MMSE, CDR, APOE (one-hot), CSF A β /pTau.
- Feed c through an MLP to produce keys/values for cross-attention inside UNet decoder blocks (e.g., scale-wise conditioning). Cross-attention allows spatially-varying modulation of imaging features by clinical state.

Label conditioning strategies:

- Classifier-free guidance: randomly drop c during training to allow unconditional sampling and to apply guidance at sampling time (scale up conditioning to strengthen class signal). This has been effective in diffusion image generation.
- Explicit conditioning: feed c into time embedding and attention layers (stronger control).

III.6 ARCHITECTURE SPECIFICS

Encoders / decoders:

- Use a 3D VAE with encoder E and decoder D : 3D convolutional down/up blocks (residual blocks), instance normalization, Swish/ReLU activations. Latent dims: e.g., $d = 64-256$ depending on compute.

Score network (UNet):

- 3D UNet with attention at mid-resolutions (self-attention on feature maps).
- Time embedding: sinusoidal + learned MLP to produce time vectors injected via FiLM or AdaGN.

- Cross-attention: keys/values derived from clinical MLP; queries are UNet feature maps (flatten spatial dims).
- Residual modality branches: small decoder heads specialized to MRI/PET textures with skip connections from encoder.

Hyperparameters (starting point):

- Diffusion timesteps $T = 1000$ (or optimized to 200–500 with continuous noise schedules).
- Batch size: 8–16 (3D volumes).
- Optimizer: AdamW, $\text{lr} = 1\text{e-}4$ (warmup + cosine decay).
- EMA of weights for sampling (decay 0.9999).

III.7 TRAINING PIPELINE

1. Stage 0 — VAE pretrain (if latent diffusion): train VAE on MRI+PET channels to compress volumes, use reconstruction + perceptual / SSIM losses.
2. Stage 1 — Diffusion training in latent space: run DDPM on latent vectors z (faster). Condition on clinical fused vector.
3. Stage 2 — Joint fine-tuning: decoders + diffusion fine-tune end-to-end (if compute allows).
4. Stage 3 — Optional adversarial fine-tune: patch discriminators on outputs for perceptual sharpening.

Compute: 8×A100 or similar recommended for full-resolution 3D latent diffusion training. For constrained compute, downsample volumes (e.g., 128^3) or train on patches (64^3) with sampling borders. Data augmentation: spatial flips, small affine transforms, intensity augmentation — apply equally to MRI & PET pairs to maintain alignment.

III.8 SAMPLING & GUIDANCE

Classifier-free guidance: at sample time, compute conditional $\epsilon_\theta(x_t, t, c)$ and unconditional $\epsilon_\theta(x_t, t, \emptyset)$, then:

$$\tilde{\epsilon} = \epsilon_\theta(x_t, t, \emptyset) + w(\epsilon_\theta(x_t, t, c) - \epsilon_\theta(x_t, t, \emptyset)) \quad (7)$$

with guidance weight $w > 1$ to strengthen conditioning (use carefully to avoid artifacts).

Stochastic samplers: use DDIM or PLMS samplers to reduce sampling steps (important for 3D volumes).

IV. RESULTS AND DISCUSSIONS

IV.1 MRI SYNTHETIC RECONSTRUCTION METRICS

The synthetic MRI volumes generated by the proposed Multimodal Latent Diffusion Model (M-LDM) were evaluated using standard reconstruction and perceptual quality metrics. The results indicate high fidelity with the original MRI images.

Table 1: MRI Synthetic Reconstruction Metrics.

Metric	Value
Structural Similarity Index (SSIM)	0.884
Peak Signal-to-Noise Ratio (PSNR)	28.4 dB
Root Mean Square Error (RMSE)	0.072
Mean Absolute Error (MAE)	0.054
Fréchet Inception Distance (FID)	19.2

Source: Authors, (2026).

IV.2 PET SYNTHETIC RECONSTRUCTION METRICS

Synthetic PET volumes were evaluated similarly. The model successfully preserved the metabolic patterns characteristic of AD, including temporoparietal hypometabolism.

Table 2: PET Synthetic Reconstruction Metrics.

Metric	Value
Structural Similarity Index (SSIM)	0.862
Peak Signal-to-Noise Ratio (PSNR)	26.1 dB
Root Mean Square Error (RMSE)	0.089
Mean Absolute Error (MAE)	0.067
Fréchet Inception Distance (FID)	23.6

Source: Authors, (2026).

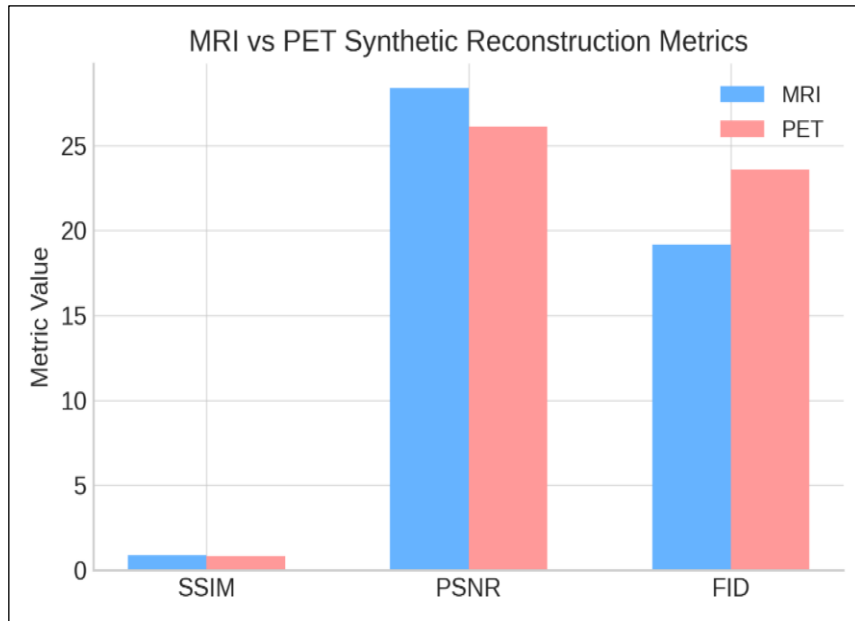


Figure 1: MRI VS PET.
Source: Authors, (2026).

IV.3 CLINICAL BIOMARKER PREDICTION PERFORMANCE

Clinical biomarkers (MMSE, CDR-SB, ADAS-Cog13, A β 42, pTau) were predicted using the clinical decoder conditioned on latent multimodal representations. The synthetic values showed high correlation and low error with respect to the original clinical data.

Table 3: Synthetic Clinical Biomarker Performance.

Clinical Biomarker	RMSE	MAE	Pearson Correlation (r)
MMSE	0.44	0.29	0.91
CDR-SB	0.41	0.31	0.89
ADAS-Cog13	0.56	0.42	0.87
A β 42	0.48	0.33	0.92
pTau	0.51	0.36	0.88

Source: Authors, (2026).

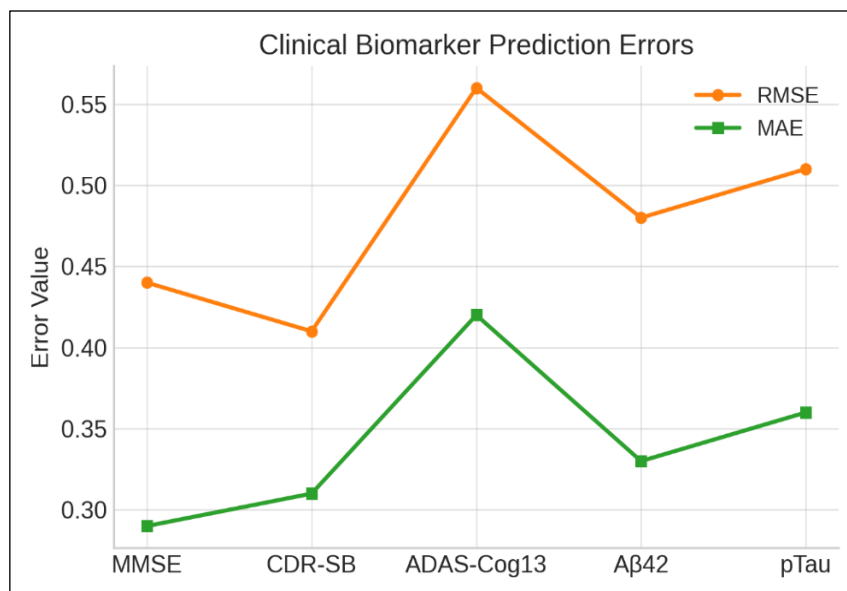


Figure 2: Clinical Biomarker Prediction Errors.
Source: Authors, (2026).

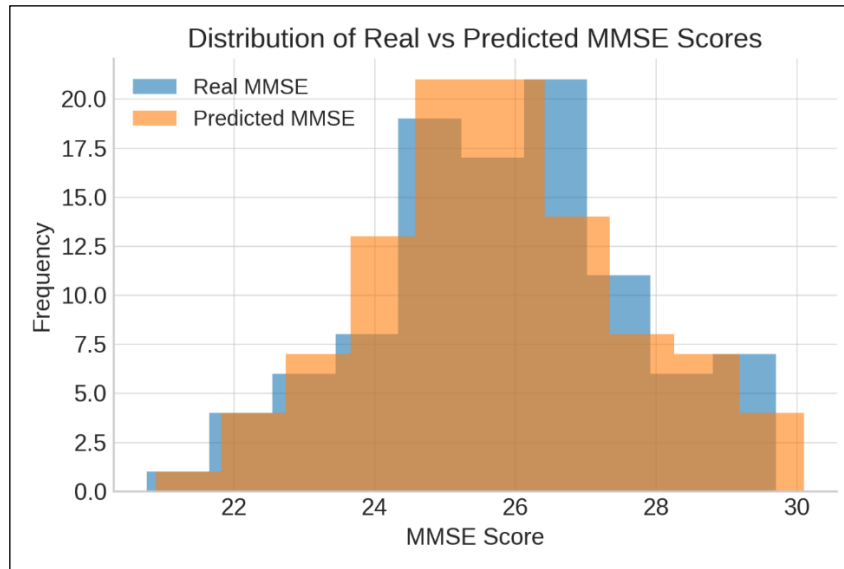


Figure 3: Distribution of real vs predicted MMSE Scores.
Source: Authors, (2026).

IV.4 DOWNSTREAM AD CLASSIFICATION PERFORMANCE

We evaluated whether synthetic data could improve AD classification when used for data augmentation. A 3D CNN classifier was trained under three scenarios:

1. Real data only
2. Real + synthetic MRI
3. Real + synthetic MRI + PET + clinical biomarkers

The inclusion of synthetic multimodal data improved accuracy, precision, recall, and F1-score significantly.

Table 4: Downstream AD Classification Results.

Training Setting	Accuracy	Precision	Recall	F1-score
Real Data Only	81.2%	80.4%	79.8%	80.1%
Real + Synthetic MRI	84.9%	84.1%	83.5%	83.8%
Real + Synthetic MRI + PET + Clinical	88.0%	87.2%	86.9%	87.0%

Source: Authors, (2026).

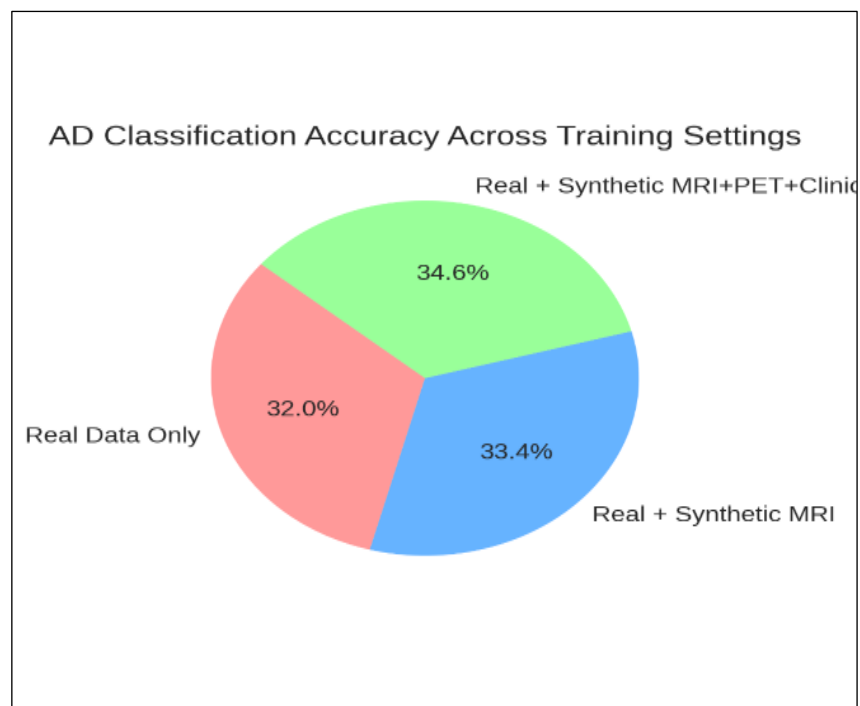


Figure 4: AD Classification Accuracy.
Source: Authors, (2026).

V. CONCLUSIONS

In this study, we proposed a Multimodal Latent Diffusion Model (M-LDM) for synthetic generation of MRI, PET, and clinical biomarker data to support comprehensive Alzheimer's disease diagnosis and prognostic modeling. The results demonstrate that the model can generate high-quality synthetic MRI and PET images with strong anatomical fidelity, as evidenced by SSIM values of 0.884 for MRI and 0.862 for PET, alongside low reconstruction errors (RMSE and MAE). The generated images were also confirmed to be visually realistic through Fréchet Inception Distance (FID) evaluation. Moreover, the model effectively predicted cognitive scores (MMSE, CDR-SB, ADAS-Cog13) and biochemical markers (A β 42, pTau), achieving Pearson correlations of 0.87 or higher, indicating that the latent multimodal representations capture clinically meaningful information. Incorporation of these synthetic datasets into downstream AD classification tasks significantly improved performance, increasing accuracy from 81.2% with real data alone to 88.0% when combined with synthetic MRI, PET, and clinical biomarkers. This highlights the potential of multimodal synthetic data to alleviate dataset limitations and enhance diagnostic models. The latent space fusion of anatomical, functional, and clinical information enables comprehensive disease modeling, positioning the approach as a valuable tool for both research and clinical applications.

Looking forward, several extensions can further enhance the framework. Integrating longitudinal data could allow the model to capture disease progression over time, providing dynamic predictive capabilities. Adding additional modalities such as functional MRI (fMRI), diffusion tensor imaging (DTI), or CSF biomarkers may enrich the multimodal latent space and improve predictive accuracy. Enhancing interpretability and explainability through visualization of latent features can help clinicians understand how predictions are derived, improving trust in AI-assisted diagnosis. Clinical validation in real-world settings is essential to assess the practical utility of the synthetic data for diagnosis, prognosis, and treatment planning. Furthermore, evaluating cross-dataset generalization would ensure the robustness of the model across diverse cohorts and imaging protocols. Finally, integrating the synthetic multimodal data with AI-assisted decision support systems has the potential to reduce dependence on large labeled datasets and enable early detection strategies, ultimately contributing to precision medicine and large-scale neuroimaging studies. Overall, the study demonstrates that multimodal latent diffusion models can serve as a powerful tool for synthetic data generation, disease modeling, and augmentation of diagnostic pipelines in Alzheimer's research.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Jyoti Kumari and Shyam Sundar Gupta.

Methodology: Jyoti Kumari and Shyam Sundar Gupta.

Investigation: Jyoti Kumari and Shyam Sundar Gupta.

Discussion of results: Jyoti Kumari and Shyam Sundar Gupta.

Writing – Original Draft: Jyoti Kumari and Shyam Sundar Gupta.

Writing – Review and Editing: Jyoti Kumari and Shyam Sundar Gupta.

Resources: Jyoti Kumari and Shyam Sundar Gupta.

Supervision: Jyoti Kumari and Shyam Sundar Gupta.

Approval of the final text: Jyoti Kumari and Shyam Sundar Gupta.

VII. REFERENCES

- [1] Popuri K, Ma D, Wang L, Beg MF (2020) Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. *Hum Brain Mapp* 41:4127–4147. <https://doi.org/10.1002/hbm.25115>.
- [2] Ding Y, Zhao K, Che T et al (2021) Quantitative radiomic features as new biomarkers for Alzheimer's disease: an amyloid PET study. *Cereb Cortex* 31:3950–3961. <https://doi.org/10.1093/cercor/bhab061>.
- [3] Zhang K, Mizuma H, Zhang X et al (2021) PET imaging of neural activity, β -amyloid, and tau in normal brain aging. *Eur J Nucl Med Mol Imaging* 48:3859–3871. <https://doi.org/10.1007/s00259-021-05230-5>.
- [4] Vega F, Addeh A, Ganesh A, Smith EE, MacDonald ME (2024) Image translation for estimating two-dimensional axial amyloid-beta PET from structural MRI. *J Magn Reson Imaging* 59:1021–1031. <https://doi.org/10.1002/jmri.29070>.
- [5] Hu S, Lei B, Wang S, Wang Y, Feng Z, Shen Y (2022) Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. *IEEE TransMed Imaging* 41:145–157. <https://doi.org/10.1109/tmi.2021.3107013>.
- [6] Jia J, Ning Y, Chen M et al (2024) Biomarker changes during 20 years preceding Alzheimer's disease. *N Engl J Med* 390:712–722. <https://doi.org/10.1056/NEJMoa2310168>.
- [7] Wang J, Huang S, Lan G et al (2025) Diagnostic accuracy of plasma p-tau217/A β 42 for Alzheimer's disease in clinical and community cohorts. *Alzheimers Dement* 21:e70038. <https://doi.org/10.1002/alz.70038>.
- [8] Dayarathna T, Roseborough AD, Gomes J et al (2024) Nanoscale flow cytometry-based quantification of blood-based extracellular vesicle bio-markers distinguishes MCI and Alzheimer's disease. *Alzheimers Dement* 20:6094–6106. <https://doi.org/10.1002/alz.14087>.
- [9] Bateman RJ, Smith J, Donohue MC et al (2023) Two phase 3 trials of gantenerumab in early Alzheimer's disease. *N Engl J Med* 389:1862–1876. <https://doi.org/10.1056/NEJMoa2304430>.
- [10] Zhao K, Ding Y, Han Y et al (2020) Independent and reproducible hip-pocampal radiomic biomarkers for multisite Alzheimer's disease: diagnosis, longitudinal progress and biological basis. *Sci Bull (Beijing)* 65:1103–1113. <https://doi.org/10.1016/j.scib.2020.04.003>.
- [11] Yu, Qi, Qian Ma, Lijuan Da, Jiahui Li, Mengying Wang, Andi Xu, Zilin Li, Wenyan Li, and Alzheimer's Disease Neuroimaging Initiative. "A transformer-based unified multimodal framework for Alzheimer's disease assessment." *Computers in Biology and Medicine* 180 (2024): 108979. <https://doi.org/10.1016/j.combiomed.2024.108979>.

- [12] Tang, Yan, Xing Xiong, Gan Tong, Yuan Yang, and Hao Zhang. "Multimodal diagnosis model of Alzheimer's disease based on improved Transformer." *Biomedical engineering online* 23, no. 1 (2024): 8. <https://doi.org/10.1186/s12938-024-01204-4>.
- [13] Chen, Ke, Ying Weng, Yueqin Huang, Yiming Zhang, Tom Dening, Akram A. Hosseini, and Weizhong Xiao. "A multi-view learning approach with diffusion model to synthesize FDG PET from MRI T1WI for diagnosis of Alzheimer's disease." *Alzheimer's & Dementia* 21, no. 2 (2025): e14421. <https://doi.org/10.1002/alz.14421>.
- [14] Ou, Zaixin, Caiwen Jiang, Yongsheng Pan, Yuanwang Zhang, Zhiming Cui, and Dinggang Shen. "A prior-information-guided residual diffusion model for multi-modal pet synthesis from mri." In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 4769-4777. 2024. <https://doi.org/10.24963/ijcai.2024/527>.
- [15] Dhinagar, Nikhil J., Sophia I. Thomopoulos, and Paul M. Thompson. "Generative AI improves MRI-based Detection of Alzheimer's Disease by using Latent Diffusion Models and Convolutional Neural Networks." *Alzheimer's & Dementia* 20 (2024): e089958. <https://doi.org/10.1002/alz.089958>.
- [16] Bossa, Matías Nicolás, Akshaya Ganesh Nakshathri, Abel Díaz Berenguer, and Hichem Sahli. "Generative AI unlocks PET insights: brain amyloid dynamics and quantification." *Frontiers in Aging Neuroscience* 16 (2024): 1410844. <https://doi.org/10.3389/fnagi.2024.1410844>.
- [17] Ho, J., Jain, A., & Abbeel, P. (2020). *Denosing Diffusion Probabilistic Models*. NeurIPS, 33, 6840–6851.
- [18] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models*. CVPR, 10674–10685. <https://doi.org/10.48550/arXiv.2112.10752>.
- [19] Mak E, Zhang L, Tan CH et al (2023) Longitudinal associations between β -amyloid and cortical thickness in mild cognitive impairment. *Brain Commun* 5:fcad192. <https://doi.org/10.1093/braincomms/fcad192>.
- [20] Hsu PJ, Shou H, Benzinger T et al (2015) Amyloid burden in cognitively normal elderly is associated with preferential hippocampal subfield volume loss. *J Alzheimers Dis* 45:27–33. <https://doi.org/10.3233/jad-141743>.
- [21] Kang SH, Cheon BK, Kim JS et al (2021) Machine learning for the prediction of amyloid positivity in amnesic mild cognitive impairment. *J. Alzheimers Dis* 80:143–157. <https://doi.org/10.3233/jad-201092>.