Review
ISSN 2413-0516

Artificial Intelligence–Integrated fMRI Methodologies for Early Diagnosis, Prognostic Modeling, and Precision Therapeutic Strategy Development in Alzheimer's Disease: A Comprehensive Narrative Review

Simin Mahakizadeh[1], Gholamreza Hassanzadeh[2,3,4], Bahador Makkiabadi[5], Fatemeh Haji Ali Asgari[6], Lida Shafaghi[3*]

[1]Department of Anatomy, School of Medicine, Alborz University of Medical Sciences, Karaj, Iran.
[2]Department of Digital Health, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.
[3]Department of Neuroscience and Addiction Studies, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran.
[4]Department of Anatomy, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.
[5]Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.
[6]Information Technology Department, Electronic Health Group, Tehran University of Medical Sciences, Tehran, Iran.
*Correspondence to: Dr. Lida Shafaghi (E-mail: lshafaghi@razi.tums.ac.ir)
(Submitted: 11 April 2025 – Revised version received: 30 May 2025 – Accepted: 25 June 2025 – Published online: 26 October 2025)

Abstract

Objective: This narrative review examines applications of artificial intelligence to fMRI in Alzheimer's disease, with emphasis on deep-learning methods. It summarizes how studies approach identification of preclinical functional patterns and assessment of treatment responsiveness, and documents common modeling choices, evaluation practices, and limitations.

Methods: This narrative review collates machine- and deep-learning uses of fMRI in Alzheimer's disease under three themes: disease-state delineation, therapy-response prediction, and progression modeling. Reported evaluation practices include patient-level site-held-out or external validation with leakage control, calibration (Brier, ECE), decision-curve analysis, scanner harmonization audits, and uncertainty/interpretability assessments. Approaches seen include CNNs, GNNs, and multimodal fusion; robustness strategies include self-supervision, domain-invariant training, and federated learning.

Results: Studies have used AI models to characterize patterns in functional connectivity, explore links to clinical trajectories, and examine responder/non-responder distinctions. Multimodal combinations of fMRI with clinical, genetic, or molecular measures often report higher cross-validated performance than fMRI alone, though findings are heterogeneous and sensitive to analytic choices. Persistent constraints include limited interpretability, small and fragmented datasets, and cross-platform variability, which together limit generalizability and clinical applicability.

Conclusion: AI-driven analysis of fMRI data shows promise in supporting the study and potential clinical management of Alzheimer's disease, including early detection and personalized treatment strategies. Future work should focus on improving model interpretability, standardization, and ethical oversight to ensure reliable and responsible application.

Keywords: Alzheimer's disease, fMRI, artificial intelligence, machine learning, deep learning

Introduction

Functional MRI (fMRI) occupies a central position in contemporary Alzheimer's disease (AD) research not because it rivals structural MRI in depicting atrophy, but because it captures systems-level dysfunction that precedes clinical decline.[1,2] Well before hippocampal volume breaches radiological thresholds, fMRI reveals reproducible alterations across large-scale networks—most prominently the default mode, salience, and frontoparietal control systems—manifesting as attenuated interregional synchrony; state-dependent non-stationarity in *dynamic functional connectivity* (time-varying coupling with altered dwell times and transition architecture);[3–5] and asymmetries in *effective connectivity* (directed influence inferred from statistical or biophysical models) within hippocampal–posterior cingulate–medial prefrontal circuitry.[6–8] Considered jointly, these phenomena constitute a coherent functional signature with potential to (i) distinguish preclinical risk from normative aging, (ii) physiologically stratify mild cognitive impairment (MCI) subtypes, and (iii) forecast progression from prodrome to dementia.[6,9–11]

The translational appeal of such biomarkers is evident: they are non-invasive, repeatable, and mechanistically anchored, linking cognitive phenotypes to circuit physiology that can be monitored longitudinally.[12–14] Inter-individual variability arising from demographic, vascular, arousal, motion, and pharmacologic factors can equal or exceed disease effects.[15,16] Moreover, pipelines predicated on stationarity and linearity underspecify intrinsically non-stationary, nonlinear network dynamics, depressing effect sizes and eroding out-of-site reproducibility.[16,17] Fragility is compounded by information leakage, over-parameterization at small N, and performance reporting that privileges discrimination while neglecting calibration, clinical utility, and decision-curve net benefit.

Artificial intelligence offers a principled route to fMRI biomarker development in AD by aligning model inductive biases with the data's fundamental structures—spatial patterning, temporal dynamics, and network topology—and by emphasizing representations that generalize across scanners and institutions.[18–22] Convolutional encoders impose locality and translation invariances to detect distributed deviations

along medial temporal–posterior midline circuits without information loss from coarse averaging;[23–25] spatiotemporal and latent-state formulations (e.g., sequence models, state-space abstractions) recast BOLD as trajectories through network states, summarizing coordination changes via occupancy, persistence, and switching;[26–28] These architectural choices are coupled with learning principles for domain robustness—self-supervised/contrastive objectives to reduce label dependence, harmonization-aware training to attenuate site/style variance without suppressing biological signal, and federated optimization to exploit multi-institutional heterogeneity while preserving privacy and local governance. Under this framework, the target biomarker is not a region-wise amplitude but a calibrated, uncertainty-aware representation of *where* coordination reorganizes, *when* it does so, and *along which edges*—with attribution stable to nuisance perturbations and transportable to held-out sites—thereby supporting discrimination across the normal-at-risk to MCI to AD continuum, forecasting conversion risk, and enabling physiologically grounded treatment stratification.

Claims of clinical readiness warrant a stringent evidentiary standard. Models should be evaluated at the patient level with site-held-out splits and auditable leakage control; produce calibrated risk estimates accompanied by decision-curve analyses benchmarked against standard care; demonstrate robustness across vendors, sequences, and preprocessing choices with explicit residual-bias reporting; and exhibit interpretability that maps stably onto known AD physiology rather than nuisance variance (e.g., motion, vascular pulsatility).[29–31] Meeting these standard shifts emphasis from within-cohort discrimination to transportable, trustworthy tools capable of informing prognosis, enriching trials, and individualizing monitoring.

Adopting this translational lens, the present review pursues four aims: (i) delineate the functional phenotype of early AD in fMRI—its network-level alterations, temporal instabilities, and directional asymmetries; (ii) analyze methodological sources of brittleness and realistic performance ceilings; (iii) evaluate AI approaches—convolutional, spatiotemporal/state-space, and graph-based models; representation learning under label scarcity; harmonization-aware training; and federated paradigms—against explicit criteria of generalization, calibration, robustness, and interpretability; and (iv) articulate methodological and governance requirements, including prospective designs with pre-specified analysis plans, standardized acquisition and metadata, transparent reporting of failure modes, and open, auditable code. Our thesis is deliberately pragmatic: AI-enhanced fMRI can advance earlier case finding, trajectory forecasting, and treatment selection/monitoring in AD, provided methodological sophistication is coupled to rigorous evaluation and operational discipline.

Methods

This narrative, methods-aware review offers a conceptual and methodological synthesis of artificial-intelligence (AI) applications to functional MRI (fMRI) in Alzheimer's disease (AD), oriented to three translation-relevant aims: preclinical/prodromal case-finding, progression forecasting—especially MCI→AD conversion—and treatment selection/monitoring. We did not pursue exhaustive retrieval or meta-analysis. Literature was identified in PubMed, Scopus, and IEEE Xplore (2004–2025), supplemented by backward/forward citation chaining and targeted screening of preprints reporting complete methods (last search: October 2025). Queries combined AD and fMRI/connectivity terminology with a broad AI lexicon (convolutional and graph neural networks, transformers, state-space/sequence models, self-supervised learning) and were augmented by terms indexing translational stringency (calibration, decision-curve analysis, site-held-out/external validation, uncertainty quantification, harmonization); a representative verbatim PubMed strategy is provided in the Supplement to support reproducibility.

Eligible contributions were peer-reviewed studies and substantial preprints in which fMRI was a primary analytic signal for AI/ML tasks in diagnosis/risk stratification, disease-course modelling, or treatment response. Acceptable representations included voxel/region activity maps, static and dynamic functional connectivity, effective connectivity, and graph/mesoscale indices; multimodal reports were retained when fMRI contributed core information to inference. Records in which AI or fMRI was peripheral, items without full text, and non-English publications were set aside; preprints were treated qualitatively and flagged as non–peer-reviewed.

From each study we abstracted cohort context (single vs multi-site, sample size, vendors/sequences), fMRI paradigms and preprocessing/harmonization, representational choices, algorithmic family (CNNs; spatiotemporal/sequence and state-space models; GNNs/transformers; multimodal hybrids), and strategies for representation learning and robustness (self-supervised/contrastive or masked pretraining; ComBat/ComBat-GAM and domain-invariant training; test-time adaptation; federated optimization). Evidence was synthesized qualitatively through a methods-aware lens emphasizing leakage control and patient-level validation granularity; external or site-held-out testing; domain-shift handling with residual site-effect quantification where available; and calibration/uncertainty (Brier score, expected calibration error) alongside interpretability and attribution stability. Given heterogeneity in datasets, representations, outcomes, and validation designs, we undertook thematic synthesis across spatial, temporal, and topological dimensions and contrasted within-site with out-of-site performance. The review was unregistered, omitted PRISMA flow and formal risk-of-bias scoring by design, and may be subject to publication and language biases; its aim is a theory- and methods-driven integration delineating architectures, representational choices, and evaluative practices most likely to yield clinically credible, transportable AI-enabled fMRI biomarkers for AD.

Computational Frameworks and AI Algorithms Applied to fMRI Biomarkers in AD

We summarize results by model family, moving from CNNs to temporal models, graph methods, transformers with self-supervision, and multimodal fusion. For each, we note the fMRI inputs used, why the fit is appropriate, how studies evaluate performance, and the main findings. Overall, external validation lags internal results, dynamic and network-aware

features often help, and robustness and calibration matter more than the specific architecture.[32]

Classical Machine Learning Baselines (SVM, Logistic/Elastic-Net, ELM, MKL)

Classical ML for AD fMRI treats diagnosis as supervised learning on hand-crafted features: static functional connectivity (sFC) derived from Fisher-z Pearson/partial correlations or precision/tangent embeddings; regional activity (ALFF/fALFF, ReHo); ICA loadings; and graph metrics (degree, clustering, participation, efficiency, modularity).[33–35] Graphs are standardized (fixed-density, MST, or tangent-space) and signals rigorously preprocessed.[36–38] Within each training fold, PCA/ICA/stability selection compresses feature to avoid leakage, feeding linear/kernel SVMs, elastic-net logistic/ridge, MKL integrators, Gaussian processes, or ELMs.[39,40] The result is transparent, sample-efficient baselines whose validity hinges on careful feature engineering and strict control of site/motion confounds.[38,41]

These models consistently achieve reliable discrimination among cognitively normal, mild cognitive impairment (MCI), and AD groups when features emphasize the default-mode and medial-temporal networks, and they show incremental gains when global integration or segregation metrics accompany local nodal measures. Prediction of MCI-to-AD conversion improves further when static connectivity features are augmented with regional activity measures or low-order dynamic connectivity statistics, although performance typically declines under patient-level, site-held-out evaluation unless harmonization techniques such as ComBat or domain-adaptive scaling are incorporated. The main advantages of these approaches lie in their transparency, interpretability, and resilience in modest sample settings: sparse weight maps readily map to neurobiologically meaningful circuits, and cross-validation with nested feature construction provides leakage resistance. Nevertheless, their dependence on handcrafted representations and sensitivity to scanner, site, and motion confounds limit external generalizability, and most studies still prioritize discrimination over calibration or clinical utility. When carefully regularized and evaluated, however, these classical models establish reproducible benchmarks for assessing newer deep-learning systems and remain indispensable for validating whether algorithmic biomarkers truly align with established AD neurophysiology.

Convolutional Neural Networks as the Mostly Applied Framework

Convolutional neural networks. Conceptually, CNNs are a good fit to fMRI because their core inductive biases—local receptive fields, weight sharing, and hierarchical feature composition—match how disease-related signal in AD tends to appear: as spatially patterned deviations in regional activity and inter-regional coupling that recur across brains. In practice, fMRI inputs to CNNs fall into two main families.

(1) Volumetric maps: 3D stacks such as ALFF/fALFF, ReHo, VMHC or voxel-wise seed/ICA maps are fed to 3D CNNs or residual CNNs; the network learns spatial motifs spanning hippocampus–posterior cingulate/precuneus–medial prefrontal territories while suppressing noise via pooling and normalization. Methodologically, pipelines extract these maps after standard preprocessing (slice timing/motion correction, band-pass ~0.01–0.1 Hz, nuisance regression), sometimes replacing fully connected layers with global average pooling to tame overfitting and preserve spatial locality for explainability.

(2) Connectome "images": ROI×ROI correlation/precision matrices are treated as images and passed to specialized CNNs that convolve along rows and columns to capture each region's connectivity "fingerprint," thereby respecting graph symmetry and reducing parameters relative to dense MLPs. The literature shows these designs are the workhorses for AD fMRI.

For connectome CNNs, Meszlényi et al. introduced a connectome-convolutional network that shared $1\times N$ and $N\times 1$ filters across the matrix; on an aMCI dataset it outperformed matched deep and shallow baselines and could integrate multiple connectivity metrics, illustrating why weight sharing over ROI fingerprints improves data efficiency in small-N fMRI settings.[42] For volumetric CNNs, several rs-fMRI AD/ADNI studies report strong discrimination or severity readouts: Qureshi et al. used 3D-CNNs on group-ICA spatial maps to classify dementia severity (CDR 0.5–1 vs 2–3) with ~0.92 balanced accuracy and highlighted networks (medial frontal, sensorimotor, executive control, dorsal attention, visual) whose involvement is neurobiologically plausible for AD.[43] Duc et al. extended this idea by jointly predicting diagnosis and MMSE from rs-fMRI ICA maps with a 3D deep model, showing that CNN features can carry both categorical and continuous clinical signal.[44] Ramzan et al. trained residual CNNs on rs-fMRI to perform six-stage classification (CN, SMC, EMCI, MCI, LMCI, AD) in ADNI, demonstrating multi-class staging capacity of volumetric CNNs, though later work cautions that rigorous, subject-level splits are essential to avoid leakage.[45] More recent ADNI work by Zhou et al. implemented a compact 3D-CNN with attention to interpretability (CAMs), reaching ~93% accuracy/AUC~0.95 and emphasizing the importance of careful class balance and multimodal co-variates.[21] Explicit explainability studies now show that Grad-CAM on 3D CNNs trained on ALFF/fALFF/ReHo/VMHC maps tends to highlight hippocampus and precuneus in AD—anatomy consistent with default-mode disruptions—while global attention patterns dominate in controls, reinforcing that CNN attributions can align with known AD circuitry when preprocessing is consistent.[46] Finally, CNNs have moved beyond cross-sectional staging toward prognosis: a 3D-CNN that integrates rs-fMRI with clinical/demographic data improved prediction of MCI→AD conversion over imaging alone, supporting the idea that learned spatial features complement clinical risk factors.[47] Taken together, the methodological picture is coherent: CNNs succeed in fMRI-AD when the representation encodes neurophysiologically meaningful structure (volumetric activity maps or ROI×ROI fingerprints), preprocessing and band-pass filtering stabilize inputs, subject-level validation prevents leakage, and attribution methods verify that discriminative features concentrate in canonical AD networks rather than nuisance signals.[46]

Autoencoders

Conceptually, AEs learn a compact latent code z that preserves the information in high-dimensional fMRI inputs by training

an encoder–decoder to reconstruct the data; the bottleneck enforces parsimonious, structured representations, while convolutional AEs exploit spatial weight sharing for voxelwise maps and matrix/patch AEs can respect ROI×ROI connectome structure. Variational AEs (VAEs) add a generative prior $p(z)$ with a KL term so that latent variables capture population distributions rather than just point embeddings—useful for normative modeling (learning "healthy" manifolds and scoring patient deviations) and for stabilizing representations across scanners/sites. This inductive setup matches rs-fMRI's regime (high dimensionality, low SNR, site heterogeneity): the encoder can denoise nuisance variance while retaining network-level signal (e.g., DMN/limbic coupling) that shallow hand-crafted pipelines must engineer explicitly. In the AD fMRI literature, AEs are routinely used either (i) unsupervised pretraining before a classifier head (or fine-tuning), or (ii) direct representation learning for subtyping, prognosis, or anomaly/normative scoring. A representative rs-fMRI study used a stacked sparse AE + graph-CNN to classify AD stages from ADNI, illustrating why an AE bottleneck plus topology-aware layers can improve small-N data efficiency compared with purely supervised CNNs.[48] For subtyping, a 2024 Scientific Reports paper fed ADNI functional-connectivity matrices to an autoencoder + hierarchical clustering pipeline and uncovered two reproducible AD subtypes—one with global FC loss ("malignant") and one with limbic-system–focused disruption ("benign")—with graph-theoretic and cognitive differences, showing how AE latent spaces can expose disease heterogeneity that aligns with network pathology.[49]

On the generative/normative side, VAEs trained on rs-fMRI learn low-dimensional spatiotemporal trajectories that summarize network dynamics, a mechanism increasingly repurposed for AD by using likelihood/reconstruction-error as deviation scores or by transferring VAE latents into AD classifiers.[50] Very recent preprints push this further: BrainVAE targets preclinical AD classification from rs-fMRI (including white-matter signals), and a VAE-based intervention framework models how lifestyle/physiological changes might alter FC-mediated AD risk—evidence that probabilistic AE latents can support hypothesis-driven biomarkers as well as prediction.[51] Finally, field-level reviews that focus on fMRI + deep learning in AD consistently list AEs (with CNNs) among the most frequently used architectures, reflecting their role as data-efficient, unsupervised feature learners that bridge noisy fMRI measurements and biologically interpretable network-level disease signatures.[52]

Temporal Hybrids (CNN-LSTM/CRNN) on Dynamic Functional Connectivity

Conceptually, these models marry convolutions (to learn spatial/connectomics motifs from each fMRI time-slice or windowed connectome) with recurrent units—typically LSTMs—to capture temporal dependencies in how networks fluctuate, making them well-suited to AD where pathology unfolds as non-stationary disconnection rather than a single static pattern. Pipelines usually (i) build sliding-window dynamic FC (dFC) or short sequences of ROI×ROI matrices/feature maps; (ii) pass each frame through a CNN (or 1D/2D conv over rows/columns to respect connectome topology); (iii) feed the per-frame embeddings to an LSTM/GRU or a lightweight temporal attention layer; and (iv) train for staging (CN↔MCI↔AD) or conversion risk with subject-level, site-aware splits. This architecture exploits the idea that AD selectively perturbs transitions among intrinsic connectivity states—especially those anchored in the DMN–limbic axis—so temporal trajectories (dwell times, switching, drift) add signal beyond static FC. Methodologically strong reports show consistent gains over static-feature baselines: a CRNN trained on ADNI rs-fMRI dFC achieved superior discrimination by extracting convolutional features per window and modeling their sequence with recurrence; ablations confirmed the temporal component was necessary for the improvement.[53] A recent CNN-LSTM study extended this to multi-stage fMRI classification, demonstrating that coupling spatial filters with sequence modeling can stabilize performance across class granularity.[54]

Independent biomarker work further validates the target signal: dynamic/multilayer FC distinguishes preclinical and symptomatic AD and aligns with amyloid/tau and cognition, justifying temporal modeling in the first place.[55] Practical lessons recur across papers: use leakage-proof, site-aware validation; nest any unsupervised transforms inside CV; and mitigate cross-scanner drift via harmonization (e.g., ComBat variants) and, where possible, light test-time adaptation—all of which narrow the common drop seen when moving from subject-random to patient-level, site-held-out tests.[56] In short, CNN-LSTM/CRNN architectures operationalize AD's time-varying network pathology by pairing spatially meaningful filters with sequence learning, yielding more generalizable fMRI biomarkers when embedded in robust, multi-site-aware pipelines.[53]

Graph Neural Networks (GNNs)

Methodologically, GNNs cast an rs-fMRI dataset into graphs whose nodes are atlas ROIs and edges are functional couplings (e.g., Pearson/precision/tangent), then learn by message passing: each layer aggregates neighbors' features (spatial GCN), optionally with attention (GAT) to weight informative connections; spatiotemporal variants add temporal convolutions/recurrent blocks over dynamic graphs to model window-wise connectivity changes. This inductive bias fits AD's network-level "disconnection"—especially DMN/limbic hubs—better than Euclidean CNNs because topology (who-connects-to-whom) is explicit, pooling can summarize communities, and attention can expose disease-relevant subgraphs. 2025 field review outlines two dominant representations—subject-level graphs (ROIs as nodes) and population graphs (subjects as nodes with phenotyped edges)—and argues they are well matched to AD heterogeneity and multimodal integration.[57] On rs-fMRI specifically, spatiotemporal GCNs have scaled to multi-site cohorts: the STGC-GCAM framework (2,272 participants across ADNI/China sites) learned dynamic FC graphs and used a Grad-CAM–style readout on nodes; it reported AUCs of 0.98 (CN vs MCI), 0.95 (CN vs AD), 0.96 (MCI vs AD) and stratified sMCI→pMCI risk (HR≈3.9) while localizing biomarkers in DMN, sensorimotor, and visual networks and linking them to cognition, Aβ and metabolism via mediation analyses.[20] Attention-based designs are also emerging: the Brain Graph Attention Network (BGAN) improves AD classification by letting the model learn edge/node saliencies over FC graphs, representative of a broader

GAT trend in connectome learning.[58] Complementing these, graph transformers extend attention across space–time; a spatiotemporal graph transformer on rs-fMRI explicitly learns temporal dependencies on top of graph structure for AD diagnosis.[59] Beyond staging, dynamic-weighted GNNs built directly on time-varying FC demonstrate early eMCI detection on ADNI by feeding windowed adjacency as edge weights and local clustering as node features, underscoring that AD signal lies in state transitions as much as in static connectivity.[60] Finally, GNNs have been used in AD-oriented brain-age modeling from rs-fMRI, validating sensitivity to medial-temporal circuitry (hippocampus/parahippocampus/amygdala) that is central to AD.[61] Taken together, the GNN literature shows that when graphs are carefully constructed (atlas choice, density control, dynamic windows), validated site-aware, and paired with interpretable readouts (e.g., graph Grad-CAM), they capture AD-relevant topology and time-varying dysconnectivity more directly than voxel-space models—yielding accurate staging, conversion risk stratification, and biologically coherent biomarkers.[20]

Transformer-Based Models (Vision/Sequence; CNN–ViT Hybrids)

Transformer architectures in AD fMRI model relations rather than only local patterns by using self-attention over tokens that represent spatial patches, connectivity tiles, or temporal segments, enabling the capture of diffuse DMN/medial-temporal perturbations and non-stationary dynamics that precede atrophy. Vision transformers (ViT) tokenize 2D "images" of ROI×ROI functional connectomes or voxel-wise correlation/ALFF/ReHo maps into patch sequences with sinusoidal or learned positional encodings; AD-focused variants such as OViTAD report multi-stage classification (CN/MCI/AD) from rs-fMRI (often with sMRI) with attention concentrating on posterior midline/medial-temporal tokens, and improved performance when transformer depth is paired with preprocessing that aggressively standardizes rs-fMRI inputs and harmonizes across sites. Sequence transformers operate directly on parcel time-series or on sliding-window dynamic FC (dFC) tensors, attending across windows to summarize state occupancy, dwell, and transition structure; recent spatiotemporal transformer formulations on ADNI dFC show gains for MCI classification and conversion risk, especially when a contrastive pretraining stage stabilizes temporal embeddings prior to supervised fine-tuning. Beyond grid tokens, spatiotemporal graph transformers treat the brain as a time-varying graph $G_t = (V, E_t)$ and combine message passing with temporal self-attention; these models leverage graph positional encodings (e.g., Laplacian eigenvectors) and report superior rs-fMRI diagnosis across AD stages under site-held-out validation, attributing importance to posterior midline and medial-temporal subgraph. In parallel, hybrid CNN→ViT stacks use convolutions to extract local motifs (e.g., hippocampus–PCC–mPFC) and transformer layers for global integration; multiclass staging (four to six categories) and multimodal fusion with clinical variables are common, with improved robustness to atlas heterogeneity via learned positional codes. For multimodal forecasting, transformer-based unified frameworks fuse sMRI/clinical/genetic streams and, when available, fMRI tokens via cross-attention, producing joint heads for diagnosis and MCI→AD conversion; these systems emphasize external testing and deliver more stable calibration when paired with temperature scaling or deep-ensemble uncertainty. Training hygiene that consistently helps includes (i) self-supervised pretraining (masked autoencoding or contrastive) on large rs-fMRI corpora to learn scanner-agnostic embeddings before AD fine-tuning, (ii) site-aware splits with patient-level, site-held-out evaluation, and (iii) harmonization or domain-invariant objectives to reduce scanner/style signatures—the latter substantially shrinking the performance gap between random and site-held-out splits. Attention maps from well-calibrated transformers repeatedly localize to DMN hubs and medial temporal lobe and remain stable under perturbation tests; in contrast, models trained without harmonization show site-dependent drift and degraded calibration on external scanners, underscoring the need for pretraining, harmonization, and explicit calibration in AD fMRI transformer pipelines.

Self-Supervised Learning (SSL: Contrastive and Masked Pretraining)

Conceptually, SSL learns fMRI representations without labels by posing pretext tasks that enforce meaningful invariances: (i) contrastive objectives pull together two augmented "views" of the same subject/session and push apart others, and (ii) masked objectives (e.g., MAE) reconstruct missing patches/entries, forcing the encoder to model global structure. For rs-fMRI this is attractive because labels are scarce, SNR is low, and disease signal is distributed across networks (DMN/limbic) rather than confined to single regions; carefully designed augmentations (time-window jitter, small noise, density-preserving edge augmentations) can suppress nuisance variance while preserving functional topology. Recent rs-fMRI methods show how to do this on graphs (ROIs as nodes, FC as edges): unsupervised contrastive graph learning preserves connectivity while augmenting raw BOLD, yielding robust subject embeddings, and diffusion-augmented graph contrastive learning further stabilizes features for downstream diagnosis—both directly applicable to AD connectomes[62] Masked modeling has also been adapted to dynamic FC matrices (RS-MAE), reducing redundancy and learning temporally aware representations that transfer to brain-disorder classification; this scaffolds AD tasks where non-stationary coupling changes matter.[63] At the field level, medical-imaging reviews and neuroinformatics surveys now highlight SSL as a data-efficient route to generalizable features and interpretable decision boundaries—often competitive with fully supervised baselines after light fine-tuning.[64] In AD-focused literature specifically, rs-fMRI reviews note the emergence of SSL pipelines alongside CNN/GNN baselines, and cross-modality evidence (e.g., SSL on ADNI PET with downstream AD progression prediction) demonstrates clear gains in low-label regimes—supporting the rationale for rs-fMRI AD adoption.[6] Practical insight: SSL works best when augmentations respect brain geometry (atlas topology, symmetry, edge density), pretraining is done site-mixed to learn domain-invariant codes, and evaluation includes site-held-out testing; under these conditions, SSL encoders provide strong starting points for AD staging and MCI→AD risk models while keeping neurobiological alignment through topology-aware objectives.[62]

Effective-Connectivity–Centric Methods (DCM Surrogates; Granger/VAR; Neural-ODE Surrogates)

Directed influence complements sFC/dFC by estimating causal flow.[65] Spectral/stochastic DCM on reduced subnetworks, multivariate autoregression/Granger families on parcel series, and neural-ODE surrogates infer edge directionality and its reweighting with disease.[66] Reported alterations include diminished hippocampal to PCC drive, altered mPFC inbound/outbound influence, and redistributed control within the medial-temporal–posterior midline axis.[67] When fed to classical or deep classifiers, these features often yield additive gains for staging and conversion risk beyond undirected connectivity.[68] Trade-offs include computational burden (full DCM), linearity/weak stationarity assumptions (VAR), and the need for rigorous priors and model checks; studies that align directed findings with dynamic-state metrics report the most stable effects.[68]

Hybrid and Ensemble Pipelines (CNN/LSTM; GNN + Transformer; Multimodal Fusion)

Hybrid systems for Alzheimer's disease (AD) integrate spatial representation with temporal inference—and, when available, molecular and cognitive priors—to model *where* and *when* coordination fails within large-scale networks.[68] Architecturally, they pair convolutional or vision-transformer (ViT) encoders that capture distributed disruptions along the medial temporal–posterior midline axis with recurrent or transformer heads that parameterize state occupancy, dwell-time shortening, transition structure, and evolving directional asymmetries (e.g., hippocampal to PCC/MPFC influence);[69] graph-centric variants replace the image backbone with GNN message passing so temporal attention operates on edge-time trajectories rather than adjacency surrogates.[70] Multimodal fusion links rs-fMRI's dynamic/topological sensitivity to structural and molecular specificity via (i) *early fusion* (feature-level concatenation under modality-aware normalization and missingness masks), (ii) *late fusion* (calibrated decision aggregation that is strata-aware for site/cohort), or (iii) *cross-modal attention* in which PET amyloid/tau, sMRI atrophy maps, CSF indices, and neuropsychological scores modulate the salience of fMRI-derived states, thereby biasing the model toward AD-consistent etiologic explanations rather than nuisance variance.[71] To reduce estimator variance and improve transport under patient-level, site-held-out evaluation, ensembles (bagging/stacking across CNN/ViT/GNN backbones, parcellations, and preprocessing variants) are coupled to probability calibration (temperature scaling or isotonic regression), with residual-site audits to quantify domain imprint in latent spaces.[32,72] The most rigorous implementations expose multi-head outputs—stage classification, individualized conversion risk, and both epistemic and aleatoric uncertainty—together with internal consistency checks (modality ablation and counterfactual dropout) to verify that attributions remain concentrated on AD-relevant circuitry and do not migrate to site or motion proxies when inputs are perturbed.[73] When embedded within harmonization-aware training and evaluated with leakage-proof splits, robustness stress tests, and decision-curve analyses against volumetry and cognitive baselines, such hybrid systems deliver what translation demands: calibrated, etiologically grounded, patient-level inferences that enrich trials, trigger earlier referral, and provide mechanistically plausible explanations in the networks AD actually disrupts.[74]

Representation Learning and Pretraining (Contrastive/Self-Supervised; Masked Modeling; VAEs)

Self-supervised pretraining on large unlabeled fMRI corpora offers a principled response to label scarcity and site heterogeneity in Alzheimer's disease imaging.[64] Contrastive objectives—at instance and graph levels—learn invariances to biologically plausible augmentations (temporal jitter, segment masking, ROI/edge dropout, spectral/bandpass or nuisance-regression variants), yielding embeddings that transfer with few labels to AD tasks (preclinical screening, MCI→AD conversion) and outperform purely supervised baselines under patient-level, site-held-out evaluation.[75] Masked reconstruction (e.g., masked autoencoders over time frames or connectome sub-blocks) complements contrastive learning by enforcing local–global consistency, reducing site imprint, and stabilizing saliency on posterior midline and medial-temporal circuitry.[76] Generative approaches (VAEs) provide low-dimensional priors for downstream classifiers, enable anomaly-style scoring for prodromal detection, and expose epistemic/aleatoric uncertainty via latent sampling.[77] Reported gains are largest when pretraining spans multiple vendors and cognitive states, augmentations are audited to preserve AD-relevant signal, and fine-tuning incorporates domain-aware alignment, leakage-proof splits, post-hoc calibration, and decision-curve analysis against volumetric and cognitive baselines.[78]

Harmonization and Domain Adaptation (Statistical and Deep)

Cross-site heterogeneity is a dominant determinant of apparent success.[32,72] Statistical harmonization (ComBat/ComBat-GAM) removes additive/multiplicative site effects at feature or embedding level;[79] deep domain-invariant training (adversarial heads, MMD/CORAL objectives, style-transfer augmentation) suppresses scanner "style" while retaining biological signal.[80] Test-time adaptation adjusts normalization, batch statistics, or small adapter layers to new sites without labels. Results consistently indicate narrower generalization gaps under patient-level site-held-out evaluation when harmonization is applied; over-aggressive correction can attenuate disease effects, so several studies audit residual site variance and assess alignment of attributions before and after harmonization.[72,74]

Federated and Privacy-Preserving Learning

Federated optimization trains models across institutions without centralizing raw scans.[81] Architectures mirror centralized counterparts (CNNs, transformers, GNNs) but incorporate secure aggregation; some introduce per-site batch-norm or lightweight adapters for personalization.[82]

Reported findings show parity with pooled training on balanced multi-site cohorts and improved robustness under heterogeneous conditions when harmonization and site-aware normalization are integrated.[83] Differential-privacy constraints on gradients bound leakage at a modest performance cost;[84] communication-efficient strategies (partial model sharing, layer freezing) are used when bandwidth limits apply.[85]

Uncertainty Estimation, Calibration, and Interpretability

High-quality reports increasingly pair discrimination with calibration (Brier score, expected calibration error) and uncertainty estimation (Monte-Carlo dropout, deep ensembles, Laplace/posterior approximations).[86] Well-calibrated outputs enable actionable thresholds for enrichment or monitoring and reduce over-confident errors on out-of-distribution sites.[87] Interpretability spans gradient-based saliency (integrated gradients, SmoothGrad), perturbation/occlusion tests on nodes/edges/time segments, and counterfactual editing (deleting or amplifying specific edges/subgraphs).[88] Stability analyses probe whether attributions persist under motion/physiology perturbations, alternative parcellations, resampled training folds, and scanner swaps; more robust pipelines show attribution concentration on DMN/medial-temporal circuits across such perturbations.[14]

Synthesis of the Prior Works

Alzheimer's disease appears on fMRI less as focal loss than as reconfigured coordination within the medial temporal–posterior midline–medial prefrontal loop, with unstable state occupancy/switching and weakened hippocampus to posterior cingulate/mPFC influence.[89] Methods converge when they are shaped to this biology. Convolutional encoders capture distributed spatial motifs; temporal modules summarize trajectories through intrinsic connectivity states; and graph-based message passing preserves who-connects-to-whom.[90] Transformer variants help chiefly by integrating long-range dependencies across space–time.[69] Across all of them, performance depends less on architectural novelty than on representation learning and domain robustness: self-supervised pretraining on mixed-site data, variational/contrastive bottlenecks that retain network geometry while down-weighting scanner "style," and harmonization embedded in training rather than bolted on afterward.[64] Effective-connectivity summaries add directional nuance when restricted to biologically motivated subnetworks, especially when their shifts cohere with dynamic-state irregularities.[65]

Credible evaluation anchors the synthesis: patient-level, site-held-out splits (ideally with an external test site), leakage-proof pipelines, and reporting that treats calibration, uncertainty, and decision utility as first-class outcomes.[32,78] Under that discipline, AI-derived representations yield transportable, mechanism-aligned readouts—calibrated risk for preclinical triage, individualized MCI to AD forecasts with honest uncertainty, and longitudinal markers (state occupancy/switching, community-level saliency) that can change before volumetry.[87] The remaining leverage is infrastructural rather than algorithmic: standardized, leakage-guarded benchmarks for dynamic/graph analyses; scanner-agnostic foundation encoders released with audits; scalable, uncertainty-aware surrogates for directionality; and prospective studies that quantify impact on clinical decisions rather than headline accuracy.[34,72,75]

Challenges, Clinical Implications, and Opportunities for Personalized Therapy in AI-Based Neuroimaging System

Despite rapid progress, clinical deployment of AI-enabled fMRI for Alzheimer's disease is limited by small, inconsistently standardized datasets; pronounced cross-site domain shift that can exceed disease signal; heterogeneous ground truth and variably defined endpoints; small-*N*/large-p regimes that invite leakage and optimism; and reporting that favors discrimination over calibration and decision utility.[91] Model-level gaps persist in interpretability (rarely stress-tested for test–retest, architectural, or perturbation stability with convergence on hippocampal–DMN circuitry), fairness and privacy (imbalanced cohorts, inadequate federation and provenance), and reproducibility (opaque pipelines, incomplete code/weights/environments).[41] The net effect is unstable or poorly calibrated risk estimates, weak portability across scanners and sites, and fragile biological plausibility for identifying responders and non-responders—undermining trial enrichment and treatment allocation. A credible path to personalization requires curated longitudinal multimodal cohorts with harmonized acquisition and preprocessing; patient-level, site-held-out or external validation with explicit leakage controls; harmonization-aware objectives with quantified residual batch effects; routine calibration and decision-curve analyses alongside discrimination; interpretability benchmarks emphasizing stability and circuit-level coherence; fairness audits with stratified performance; privacy-preserving federated training with auditable provenance; and prospective or quasi-prospective impact evaluations that tie predictions to concrete clinical actions.

Clinical Implications

Under leakage-controlled development and validation on patient-level, site-held-out cohorts, AI-enabled fMRI can credibly transition from methodological innovation to a clinically consequential platform for Alzheimer's disease. The appropriate evidentiary standard extends beyond cross-validated discrimination to include externally verified calibration across vendors and acquisition protocols, decision-curve net benefit against guideline-concordant comparators, and prespecified action thresholds that operationalize concrete clinical decisions (screening referral, trial enrichment, or allocation to disease-modifying therapy). Claims regarding the superiority of graph-based and spatiotemporal representations of dynamic functional connectivity must be established in head-to-head, out-of-site analyses with rigorous leakage controls, reproducible preprocessing pipelines, sensitivity analyses for acquisition heterogeneity, and transparent uncertainty quantification. Health-system feasibility requires concurrent

assessment of resource implications, including scanner time, competing biomarker pathways (PET/CSF/plasma), and incremental cost-effectiveness under realistic throughput constraints.

Methodological prerequisites for translational credibility are non-negotiable. Cross-site harmonization should be audited via comparative evaluations of state-of-the-art techniques (e.g., ComBat-family extensions versus domain-adaptive residual approaches), with residual batch structure explicitly quantified and its effect on clinical decisions reported. Multi-institutional learning ought to be implemented through privacy-preserving frameworks (e.g., federated training with auditable provenance) to enable statistically efficient model development without raw-data pooling. Interpretability constitutes a primary scientific endpoint: attribution and attention maps must demonstrate test–retest reliability, architectural stability, and neurobiological coherence with disease-relevant circuitry (e.g., hippocampal–default-mode interactions), supported by perturbation and counterfactual analyses rather than post hoc exemplars. Reporting should adhere to contemporary guidance (TRIPOD-AI for prediction, CONSORT-AI for interventional evaluation, CLAIM for imaging), with pre-registered analysis plans, patient-level leakage safeguards, complete code and parameter provenance, and deployable artifacts. Absent these elements, assertions of clinical impact remain, by definition, method development rather than translation.

Opportunities for Personalized Therapeutic Strategies

Personalization in Alzheimer's disease is plausible when fMRI-derived representations are integrated with orthogonal modalities (structural MRI, PET amyloid/tau, plasma/CSF biomarkers, neuropsychology, genomics) within multimodal frameworks that manage heterogeneity and quantify uncertainty. Properly specified models should yield auditable allocation rules for predictive enrichment, treatment selection, and monitoring—i.e., prespecified thresholds with calibrated risk and decision-curve net benefit against standard care. Spatiotemporal network indices (state occupancy, dwell-time asymmetries, transition structure within hippocampal–DMN/control circuits) are suited as mechanistically interpretable response and pharmacodynamic markers, updated sequentially as imaging and fluid data accrue. Translation requires patient-level, site-held-out generalization with harmonization audits of residual batch structure, alongside evaluation of operational endpoints (throughput, error costs) and incremental cost-effectiveness. Interpretability is a binding constraint: attributions must show test–retest stability, anatomical coherence, and cross-vendor robustness. Enablers include privacy-preserving federated training, standardized preprocessing/reporting with code and parameter provenance, and joint clinical–methodological governance. Under these conditions, AI-assisted fMRI progresses from descriptive connectomics to precision therapeutic instrumentation—assigning and adapting the right intervention for the right patient at the right time.

Conclusion

The integration of neuroimaging and AI heralds a new era of personalized medicine for Alzheimer's disease. By shifting clinical paradigms from reactive to proactive and from one-size-fits-all to individualized approaches, these innovations promise a transformative impact on early diagnosis, targeted interventions, and tailored patient care. Continued interdisciplinary collaboration among neuroscientists, clinicians, engineers, and policymakers will be pivotal in realizing this vision.

Acknowledgments

We wish to express our sincere appreciation to Dr. Seyed Amir Hossein Batouli for his substantial contributions to this work. We acknowledge the many researchers whose original studies made this synthesis possible. This review draws on their published work to summarize current evidence and highlight practical directions for the field.

Statements and Declarations

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Conception and study design were led by SM, with contributions from GhH and LSH. Literature search, screening, data curation, and extraction were performed by SM and FAA. Qualitative synthesis and analysis were conducted by SM with methodological and analytical contributions from BM. The first draft of the manuscript was written by SM, and critical revisions were provided by GhH, BM, FAA, and LSH. Supervision and project administration were provided by LSH (overall) and GhH. LSH serves as guarantor and corresponding author. All authors read and approved the final manuscript. All authors commented on manuscript revisions and approved the final version.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Use of AI-Assisting Tools in Manuscript Preparation

With full respect for the principles of academic integrity and transparency, the authors wish to disclose that AI-assisted tools were employed solely to enhance the clarity and quality of language during manuscript preparation. Their use was limited to grammar correction, sentence restructuring, and refinement of academic tone. At no stage were these tools used to generate original content, contribute to conceptual development, design methodology, analyze data, interpret findings, produce figures or tables, or generate references. The authors take full and sole responsibility for the scientific content, accuracy, and originality of the work presented. ■

Table 1. **Summary of representative AI–AD-fMRI studies**

Type	Subtype	rs-fMRI	No.		Main findings	Notes
			Org.	Rev.		
Convolutional Neural Network (CNN Family)	3D volumetric CNN	4D rs-fMRI volumes or temporally summarized 3D volumes	12	2	Learns spatio-temporal patterns directly from volumes; highlights DMN/MTL regions in saliency; works on ADNI-like cohorts.	Heavy compute; sensitive to preprocessing and scan length.
	CNN	Static FC or alternative connectivity (e.g., PSI) converted to 2D images	25	-	CNN can exploit texture-like patterns in FC matrices; combining topological features with CNN boosts accuracy over either alone.	Requires atlas choice; sensitive to thresholding.
Transformer & Attention-Based Models	Spatio-temporal graph transformer (STGTN)	ROI-wise time series → dynamic graphs	2	-	Jointly models spatial/topological and temporal dependencies; improved generalization on ADNI-like data.	Data-hungry; interpretability via attention maps is promising.
	Spatio-temporal transformer for dFC (MCI focus)	Sliding-window dFC sequence	2	-	Better capture of long-range temporal dependencies than RNN baselines; useful for prodromal (MCI) prediction.	Often framed as MCI prediction; still AD-relevant.
Graph Neural Networks (GNN)	Spatio-temporal GCN (ST-GCN) with CAM	Dynamic FC graphs	1	-	Identifies stable FC biomarkers across multi-site data; highlights topological mediators linking Aβ / metabolism to cognition.	Good biomarker localization; multi-site robustness emphasized.
	Local-to-global GNN	Static/dynamic graphs	1	-	Hierarchical message passing (local→global) improves classification of brain disorders including AD.	Often multi-disorder; AD subset shows consistent gains.
	BrainNet-GCN / feature-augmented GNN	rs-fMRI graph + clinical/cognitive features	54	1	Combining imaging graph with non-imaging features improves discrimination.	Requires harmonization of scales/modalities.
Autoencoders (AE)	Frequency-aware encoder / interpretable encoder	ROI time series → multi-band dFNC	1	-	Encoders that model frequency-specific dynamics reveal subcortical/sensorimotor/cerebellar disruptions; moderate AUCs with interpretability.	Emerging; emphasis on explanation over raw accuracy.
	Multiscale autoencoder	rs-fMRI + T1/DTI	-	-	AE used to fuse structural–functional features; yields small but consistent gains over single-modality baselines.	Includes fMRI but not fMRI-only.
Classical ML (Non-Deep)	SVM	Static FC → graph measures (efficiency, clustering, hubs)	28	1	Robust CN vs. AD; MCI vs. CN harder; PCC/precuneus hubs most disrupted; early evidence for MCI→AD prediction.	Sensitive to atlas & threshold; strong baselines even today.
	SVM / RF	Correlation, CorrTF, spectral features	9	-	Handcrafted connectivity or spectral features can reach high within-study accuracy; prone to overfitting without external validation.	Useful as transparent baselines.
Other AIs / Hybrids	CNN + topological features	PSI/FC images + graph features	5	-	Hybrid pipelines outperform either pure CNN or pure graph features; complementary information.	Practical, modular.
	GAN/Transformer fusion (multimodal)	rs-fMRI + DTI (fusion)	6	1	Cross-modal attention improves connectivity-centric biomarkers and multi-class staging.	Includes fMRI but relies on multimodal inputs.
	CNN-RNN hybrids (CRNN / CNN-LSTM)	Dynamic FC sequence (sliding-window FCs)	2	-	Modeling temporal order of FC states improves discrimination vs. static FC alone; gains particularly for MCI vs. CN.	Needs enough time-points; windowing hyper-parameters matter.

References

1. Frisoni GB, Fox NC, Jack Jr CR, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology. 2010;6(2):67–77.
2. By S, Kahl A, Cogswell PM. Alzheimer's disease clinical trials: what have we learned from magnetic resonance imaging. Journal of Magnetic Resonance Imaging. 2025;61(2):579–94.
3. Gu Y, Lin Y, Huang L, Ma J, Zhang J, Xiao Y, et al. Abnormal dynamic functional connectivity in Alzheimer's disease. CNS Neuroscience & Therapeutics. 2020;26(9):962–71.
4. Zhao C, Huang W-J, Feng F, Zhou B, Yao H-X, Guo Y-E, et al. Abnormal characterization of dynamic functional connectivity in Alzheimer's disease. Neural Regeneration Research. 2022;17(9):2014–21.
5. Matsui T, Yamashita K-i. Static and dynamic functional connectivity alterations in Alzheimer's disease and neuropsychiatric diseases. Brain Connectivity. 2023;13(5):307–14.
6. Alarjani M, Almarri B. fMRI-based Alzheimer's disease detection via functional connectivity analysis: a systematic review. PeerJ Computer Science. 2024;10:e2302.
7. Wang Y-F, Huang Y, Chang X-Y, Guo S-Y, Chen Y-Q, Wang M-Z, et al. Classification of neurodegenerative diseases using brain effective connectivity and machine learning techniques: a systematic review. Frontiers in Neurology. 2025;16:1581105.
8. Pini L, Lista S, Griffa A, Allali G, Imbimbo BP. Can brain network connectivity facilitate the clinical development of disease-modifying anti-Alzheimer drugs? Brain Communications. 2025;7(1):fcae460.
9. Huang C, Yan Y, Shi Y, Lin R, Li H. Impact of art-based intervention on brain functional connectivity in older adults along the alzheimer's continuum. Innovation in Aging. 2024;8(Supplement_1):940.
10. Chauhan N, Choi B-J. Comparison of functional connectivity analysis methods in alzheimer's disease. Applied Sciences. 2022;12(16):8096.
11. Ibrahim B, Suppiah S, Ibrahim N, Mohamad M, Hassan HA, Nasser NS, et al. Diagnostic power of resting-state fMRI for detection of network connectivity in Alzheimer's disease and mild cognitive impairment: A systematic review. Human Brain Mapping. 2021;42(9):2941–68.
12. Woo CW, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. Nat Neurosci. 2017;20(3): 365–77.
13. Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nature Protocols. 2017;12(3):506–18.
14. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage. 2012;59(3):2142–54.
15. Murphy K, Birn RM, Bandettini PA. Resting-state fMRI confounds and cleanup. Neuroimage. 2013;80:349–59.
16. Yamashita A, Yahata N, Itahashi T, Lisi G, Yamada T, Ichikawa N, et al. Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. PLoS Biology. 2019;17(4):e3000042.
17. Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, et al. Variability in the analysis of a single neuroimaging dataset by many teams. Nature. 2020;582(7810):84–8.
18. Fedorov A, Wu L, Sylvain T, Luck M, DeRamus TP, Bleklov D, et al., editors. On self-supervised multimodal representation learning: an application to Alzheimer's disease. 2021 IEEE 18th international symposium on biomedical imaging (ISBI); 2021: IEEE.
19. Li X, Gu Y, Dvornek N, Staib LH, Ventola P, Duncan JS. Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. Medical Image Analysis. 2020;65:101765.
20. Zhang Y, Xue L, Zhang S, Yang J, Zhang Q, Wang M, et al. A novel spatiotemporal graph convolutional network framework for functional connectivity biomarkers identification of Alzheimer's disease. Alzheimer's Research & Therapy. 2024;16(1):60.
21. Zhou X, Kedia S, Meng R, Gerstein M. Deep learning analysis of fMRI data for predicting Alzheimer's disease: A focus on convolutional neural networks and model interpretability. Plos One. 2024;19(12):e0312848.
22. Wei Y, Abrol A, Lah J, Qiu D, Calhoun VD, editors. A deep spatio-temporal attention model of dynamic functional network connectivity shows sensitivity to Alzheimer's in asymptomatic individuals. 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2024: IEEE.
23. Beheshti N, Johnsson L. Using CNNs for AD classification based on spatial correlation of BOLD signals during the observation. arXiv preprint arXiv:210410596. 2021.
24. Parmar H, Nutter B, Long R, Antani S, Mitra S. Spatiotemporal feature extraction and classification of Alzheimer's disease using deep learning 3D-CNN for fMRI data. J Med Imaging (Bellingham). 2020;7(5):056001.
25. Warren SL, Moustafa AA. Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: A systematic review. Journal of Neuroimaging. 2023;33(1):5–18.
26. Mittal A, Linderman S, Paisley J, Sajda P. Bayesian recurrent state space model for rs-fMRI. arXiv preprint arXiv:201107365. 2020.
27. Suk HI, Wee CY, Lee SW, Shen D. State-space model with deep learning for functional dynamics estimation in resting-state fMRI. Neuroimage. 2016;129:292–307.
28. Shappell H, Caffo BS, Pekar JJ, Lindquist MA. Improved state change estimation in dynamic functional connectivity using hidden semi-Markov models. Neuroimage. 2019;191:243–57.
29. Alharbi H, Juanatas RA, Al Hejaili A, Lim S-j. Spectral graph convolutional neural network for Alzheimer's disease diagnosis and multi-disease categorization from functional brain changes in magnetic resonance images. Frontiers in Neuroinformatics. 2024;18:1495571.
30. Mohammadi H, Karwowski W. Graph neural networks in brain connectivity studies: methods, challenges, and future directions. Brain Sci. 2024;15(1):17.
31. Teipel SJ, Wohlert A, Metzger C, Grimmer T, Sorg C, Ewers M, et al. Multicenter stability of resting state fMRI in the detection of Alzheimer's disease and amnestic MCI. Neuroimage Clin. 2017;14:183–94.
32. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. JAMA Psychiatry. 2020;77(5):534–40.
33. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika. 1915;10(4):507–21.
34. Dadi K, Rahim M, Abraham A, Chyzhyk D, Milham M, Thirion B, et al. Benchmarking functional connectome-based predictive models for resting-state fMRI. NeuroImage. 2019;192:115–34.
35. Yu-Feng Z, Yong H, Chao-Zhe Z, Qing-Jiu C, Man-Qiu S, Meng L, et al. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain and Development. 2007;29(2):83–91.
36. Tewarie P, van Dellen E, Hillebrand A, Stam CJ. The minimum spanning tree: an unbiased method for brain network analysis. Neuroimage. 2015;104:177-88.
37. Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature Methods. 2019;16(1):111–6.
38. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage. 2012;59(3):2142–54.
39. Jolliffe I. Principal component analysis. International encyclopedia of statistical science: Springer; 2011. p. 1094–6.
40. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Transactions on Medical Imaging. 2004;23(2):137–52.
41. Yu M, Linn KA, Cook PA, Phillips ML, McInnis M, Fava M, et al. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. Human Brain Mapping. 2018;39(11):4213–27.
42. Meszlényi RJ, Buza K, Vidnyánszky Z. Resting state fMRI functional connectivity-based classification using a convolutional neural network architecture. Frontiers in Neuroinformatics. 2017;11:61.
43. Qureshi MNI, Ryu S, Song J, Lee KH, Lee B. Evaluation of functional decline in Alzheimer's dementia using 3D deep learning and group ICA for rs-fMRI measurements. Frontiers in Aging Neuroscience. 2019;11:8.
44. Duc NT, Ryu S, Qureshi MNI, Choi M, Lee KH, Lee B. 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI. Neuroinformatics. 2020;18(1):71–86.
45. Ramzan F, Khan MUG, Rehmat A, Iqbal S, Saba T, Rehman A, et al. A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. Journal of Medical Systems. 2020;44(2):37.
46. Song B, Yoshida S, Initiative AsDN. Explainability of three-dimensional convolutional neural networks for functional magnetic resonance imaging of Alzheimer's disease classification based on gradient-weighted class activation mapping. Plos One. 2024;19(5):e0303278.
47. Ghafoori S, Shalbaf A. Predicting conversion from MCI to AD by integration of rs-fMRI and clinical information using 3D-convolutional neural network. International Journal of Computer Assisted Radiology and Surgery. 2022;17(7):1245–55.

48. Chelladurai A, Narayan DL, Divakarachari PB, Loganathan U. fMRI-based alzheimer's disease detection using the sas method with multi-layer perceptron network. Brain Sci. 2023;13(6):893.
49. Sheng J, Xin Y, Zhang Q, Yang Z, Wang L, Zhang Q, et al. Novel Alzheimer's disease subtypes based on functional brain connectivity in human connectome project. Scientific Reports. 2024;14(1):14821.
50. Zhang X, Maltbie EA, Keilholz SD. Spatiotemporal trajectories in resting-state FMRI revealed by convolutional variational autoencoder. NeuroImage. 2021;244:118588.
51. Li Y, Xu L, Zuo L, Chang Y, Ding Z, Anderson AW, et al. BrainVAE: Exploring the role of white matter BOLD in preclinical Alzheimer's disease classification. bioRxiv. 2025.
52. Warren SL, Moustafa AA. Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: A systematic review. J Neuroimaging. 2023;33(1):5–18.
53. Lin K, Jie B, Dong P, Ding X, Bian W, Liu M. Convolutional recurrent neural network for dynamic functional mri analysis and brain disease identification. Front Neurosci. 2022;16:933660.
54. Farhan S, Haq YU, Khaliq MA, Afza S, Ahmad F, Mahmood T, et al. Enhancing the early detection of Alzheimer's disease using an integrated CNN-LSTM framework: A robust approach for fMRI-based multi-stage classification. Plos One. 2025;20(8):e0317968.
55. Canal-Garcia A, Veréb D, Mijalkov M, Westman E, Volpe G, Pereira JB, et al. Dynamic multilayer functional connectivity detects preclinical and clinical Alzheimer's disease. Cerebral Cortex. 2024;34(2):bhad542.
56. Wang M, Zhu L, Li X, Pan Y, Li L. Dynamic functional connectivity analysis with temporal convolutional network for attention deficit/hyperactivity disorder identification. Front Neurosci. 2023;17:1322967.
57. Ali S, Piana M, Pardini M, Garbarino S. Graph Neural Networks in Alzheimer's Disease Diagnosis: A Review of Unimodal and Multimodal Advances. Frontiers in Neuroscience.19:1623141.
58. Zhou Z, Wang Q, An X, Chen S, Sun Y, Wang G, et al. A novel graph neural network method for Alzheimer's disease classification. Comput Biol Med. 2024;180(C):11.
59. He P, Shi Z, Cui Y, Wang R, Wu D. A spatiotemporal graph transformer approach for Alzheimer's disease diagnosis with rs-fMRI. Computers in Biology and Medicine. 2024;178:108762.
60. Liu L, Li Y, Yang K. Dynamically weighted graph neural network for detection of early mild cognitive impairment. Plos One. 2025;20(6):e0323894.
61. Gao J, Liu J, Xu Y, Peng D, Wang Z. Brain age prediction using the graph neural network based on resting-state functional MRI in Alzheimer's disease. Frontiers in Neuroscience. 2023;17:1222751.
62. Wang X, Chu Y, Wang Q, Cao L, Qiao L, Zhang L, et al. Unsupervised contrastive graph learning for resting-state functional MRI analysis and brain disorder detection. Hum Brain Mapp. 2023;44(17):5672–92.
63. Ma H, Xu Y, Tian L. RS-MAE: Region-State Masked Autoencoder for Neuropsychiatric Disorder Classifications Based on Resting-State fMRI. IEEE Trans Neural Netw Learn Syst. 2025;36(6):10707–20.
64. Huang S-C, Pareek A, Jensen M, Lungren MP, Yeung S, Chaudhari AS. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. NPJ Digital Medicine. 2023;6(1):74.
65. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage. 2003;19(4):1273–302.
66. Razi A, Kahan J, Rees G, Friston KJ. Construct validation of a DCM for resting state fMRI. Neuroimage. 2015;106:1–14.
67. Michalareas G, Schoffelen JM, Paterson G, Gross J. Investigating causality between interacting brain areas with multivariate autoregressive models of MEG sensor data. Human Brain Mapping. 2013;34(4):890–913.
68. Whittaker JR, Driver ID, Venzi M, Bright MG, Murphy K. Corrigendum: cerebral autoregulation evidenced by synchronized low frequency oscillations in blood pressure and resting-state fMRI. Frontiers in Neuroscience. 2020;14:544.
69. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020.
70. Parisot S, Ktena SI, Ferrante E, Lee M, Guerrero R, Glocker B, et al. Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. Medical image analysis. 2018;48:117–30.
71. Suk H-I, Lee S-W, Shen D, Initiative AsDN. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage. 2014;101:569–82.
72. Dinsdale NK, Jenkinson M, Namburete AIL. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. Neuroimage. 2021;228:117689.
73. Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017;30:5580–5590.
74. Fortin J-P, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. NeuroImage. 2018;167:104–20.
75. Chen T, Kornblith S, Norouzi M, Hinton G. A Simple Framework for Contrastive Learning of Visual Representations. In: Hal D, III, Aarti S, editors. Proceedings of the 37th International Conference on Machine Learning; Proceedings of Machine Learning Research: PMLR; 2020. p. 1597–607.
76. Hou Z, Liu X, Cen Y, Dong Y, Yang H, Wang C, et al. GraphMAE: Self-Supervised Masked Graph Autoencoders. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Washington DC, USA: Association for Computing Machinery; 2022. p. 594–604.
77. Atlason HE, Love A, Sigurdsson S, Gudnason V, Ellingsen LM. SegAE: Unsupervised white matter lesion segmentation from brain MRIs using a CNN autoencoder. NeuroImage: Clinical. 2019;24:102085.
78. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006;7(1):91.
79. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2006;8(1):118–27.
80. Long M, Cao Y, Wang J, Jordan M. Learning Transferable Features with Deep Adaptation Networks. In: Francis B, David B, editors. Proceedings of the 32nd International Conference on Machine Learning; Proceedings of Machine Learning Research: PMLR; 2015. p. 97–105.
81. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digital Medicine. 2020;3(1):119.
82. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security; Dallas, Texas, USA: Association for Computing Machinery; 2017. p. 1175–91.
83. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports. 2020;10(1):12598.
84. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, et al. Deep Learning with Differential Privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; Vienna, Austria: Association for Computing Machinery; 2016. p. 308–18.
85. Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:161005492. 2016.
86. Gal Y, Ghahramani Z, editors. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. international conference on machine learning; 2016: PMLR.
87. Guo C, Pleiss G, Sun Y, Weinberger KQ, editors. On calibration of modern neural networks. International conference on machine learning; 2017: PMLR.
88. Sundararajan M, Taly A, Yan Q, editors. Axiomatic attribution for deep networks. International conference on machine learning; 2017: PMLR.
89. Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proceedings of the National Academy of Sciences. 2004;101(13):4637–42.
90. Wen J, Thibeau-Sutre E, Diaz-Melo M, Samper-González J, Routier A, Bottani S, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. Medical Image Analysis. 2020;63:101694.
91. Pomponio R, Erus G, Habes M, Doshi J, Srinivasan D, Mamourian E, et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. NeuroImage. 2020;208:116450.